![WestEd logo]
WestEd.org

Evaluation of Learning by Making i3 Project: STEM Success for Rural Schools

Linlin Li
Rachel Tripathy
Katie Salguero
Betsy McCarthy

November 2018

WestEd is a nonpartisan, nonprofit research, development, and service agency that works with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont, and Georgia to Illinois, Arizona, and California, with headquarters in San Francisco.

WestEd

WestEd.org

Contents

Introduction

Overview of Learning by Making Project

The Learning by Making (LbyM) project is funded by the U.S. Department of Education's Investing in Innovation Fund (i3). As a five-year development project (2014-2018), Sonoma State University (SSU), in partnership with Mendocino County Office of Education (MCOE), is developing an innovative, integrated high school Science, Technology, Engineering, and Mathematics (STEM) curriculum. It addresses three absolute priorities for the FY 2013 i3 development competition: improving low-performing schools (absolute priority 2), improving Science, Technology, Engineering, and Mathematics (STEM) education (absolute priority 3), and serving rural communities (absolute priority 8). The LbyM curriculum focuses on computational thinking and scientific and engineering design practices. It consists of Disciplinary Core Ideas (DCIs) in Earth Science, Biology, Chemistry, and Physics as described in the Next Generation Science Standards (NGSS, 2013). The LbyM curriculum emphasizes the NGSS crosscutting concepts of Cause and Effect, Systems and Systems Models, and Stability and Change.

The NGSS recommend that students develop and use models; construct explanations and arguments from experimental evidence; and report and communicate their results to their peers and instructors. The Common Core State Standards in mathematics (CCSS-M) suggest that students "apply the mathematics they know to solve problems arising in everyday life, society, and the workplace." CCSS-M also recommend "the strategic use of technology to assist students in forming and testing conjectures, creating graph and data displays and determining and assessing lines of fit for data" as well as performing geometric constructions (CCSS-M Appendix A). The LbyM curriculum integrates the NGSS Scientific and Engineering Design Practices and CCSS-M Standards for Mathematical Practice into a modified project-based learning approach whereby the use of mathematics is integrated with the science content and technology as students do experiments related to real-world problems. Both NGSS and CCSS-M practices encourage students to think analytically about both the context around and the logic supporting STEM concepts. Ultimately, this requires teachers to teach these subjects differently, and leads students to engage in STEM in a conceptually deeper and more meaningful way than in previous years.

The curriculum developed by the LbyM project team includes an easy-to-use Logo network that supports data transfer from sensors used during student-designed investigations. The use of Logo is particularly appropriate for this curriculum since the Logo programming language is specifically designed for education, with the goal of fostering computational thinking, including modeling, simulations, and communicating graphical information. During LbyM investigations, students write Logo code to read the sensors and obtain data. Further, students build on basic coding to perform individualized experiments, create simulations, and explore models. They analyze and interpret data while using mathematics and engage in computational thinking without the difficulties usually encountered in learning complex programming languages.

Throughout the project, teachers are supported in using this project-based curriculum to inspire student learning through inquiry in a type of learning experience that is often referred to as "constructionism"

(Papert & Harel, 1991). To this end, teachers in the study are provided with ongoing professional development (PD) that emphasizes technology and engineering skills, such as computer coding and circuitry. PD training also covers instructional approaches centered around student learning through inquiry, data collection, and experimental design.

Logic Model

As mentioned previously, the LbyM curriculum was created by content specialists from SSU. The specialists have created a full experience for teachers. The SSU team supports teachers by providing a Summer Institute for PD, and follow-up training sessions throughout the school year to explore and learn the STEM concepts in the LbyM curriculum and implementation. Teachers then implement the LbyM units in their science classes, during which SSU staff is readily available when instructional or technological support is needed. For each year of the LbyM project, the LbyM courses have fulfilled the A-G course requirements for college track classes.

The LbyM outcomes are focused on teachers' instructional practices, student achievement in math and science, and their engagement, interest, and confidence in STEM. Many of these factors are tied to the computational thinking required to teach and engage in the LbyM STEM classes. Figure 1 shows the logic model of LbyM.
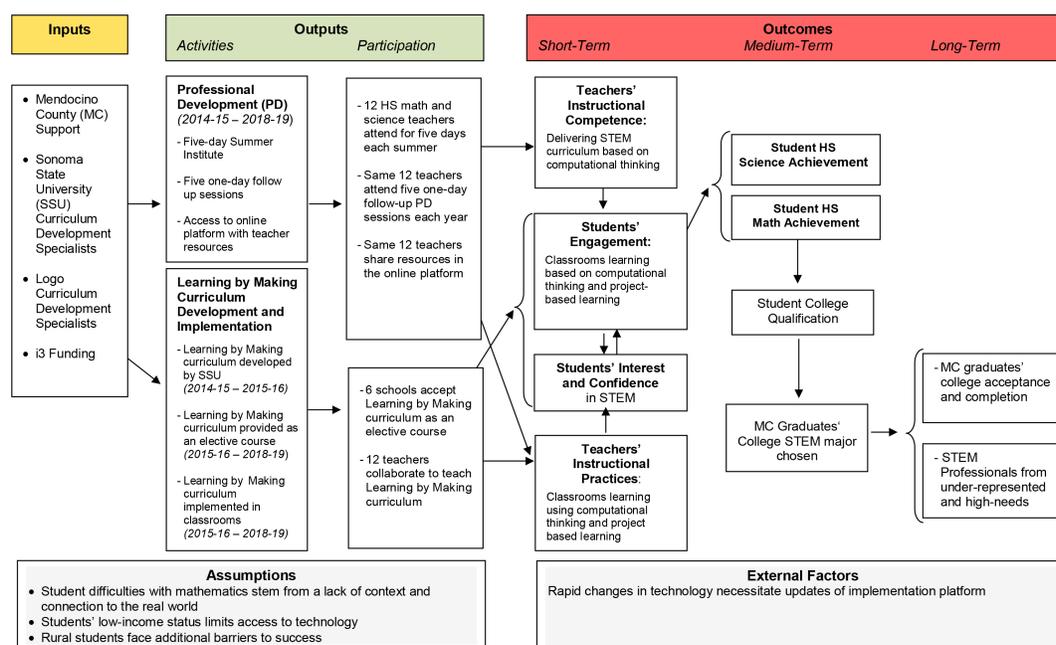
Figure 1. Learning by Making Curriculum Logic Model

Driven by the logic model, the study addresses the following research questions:

Confirmatory Research Questions:

RQ1: Does LbyM have an effect on mathematics performance of 9th to 12th grade students compared to mathematics performance of 9th to 12th grade students in the business-as-usual condition?

RQ2: Does LbyM have an effect on science performance of 9th to 12th grade students compared to science performance of 9th to 12th grade students in the business-as-usual condition?

Exploratory Research Questions:

RQ3: Does LbyM have an effect on high school mathematics and science teachers' competence in delivering computational thinking lab-based STEM curriculum?

RQ4: Does LbyM increase high school students' interests in STEM and a STEM career?

Study Design and Methodology

Background of the Implementation Sites

The study of the LbyM STEM project took place in the 2016-2017 academic year. A total of 278 high school students were recruited from six high-need rural high schools in Mendocino County, California. Among the 278 high school students, 137 received the LbyM STEM intervention, and 141 students who were enrolled in other high school science or math courses served as comparison students. These six high-need rural high schools in Mendocino County primarily serve families with high rates of family poverty and low rates of college readiness. As shown in Table 1, students in the participating schools are largely from low-income households, as indicated by the proxy measure of eligibility for reduced-price or free lunch.

Table 1. Participating School Demographics (2016-17 Academic Year)

	Anderson Valley High School	Layton-ville High School	Point Arena High School	Round Val-ley High School	Ukiah High School	Willits High School
Race						
Hispanic or Latino	74.2	10.8	51.7	27.0	47.7	32.1
African American	0.9	0.8	0.0	0.0	1.2	0.7
White	23.0	70.0	39.3	19.0	42.8	55.0
Asian/Pacific Islander	0.0	0.8	0.7	0.0	2.0	2.2
Native American	0.9	13.3	7.6	53.9	4.4	5.4
Two or More Races	0.9	4.2	0.0	0.0	1.7	1.7
Free/Reduced Price Meals	71.9	66.7	54.5	34.8	61.1	74.0
EL Students	18.9	0.0	15.2	20.2	14.7	8.5

Source: California Department of Education

Intervention and Comparison Conditions

The study of the LbyM STEM project used a quasi-experimental design. One hundred thirty-seven students were recruited to enroll in eight LbyM STEM classes in six participating high schools. Three of the participating high schools are small schools, and it is not possible to find comparison students from these small schools. Therefore, all comparison students were recruited from the larger schools, with a total of 141 comparison students from six classrooms in three out of six participating high schools. Figure 2 presents a flowchart showing the number of school, classroom, and student participants during the recruitment and the changes in the numbers of student participants.

Number of recruited classes:

N=14 classes from 6 participating schools

(total # of students: 278)

Number of classes in the treatment group:
N=8 classes from 6 participating schools

(total # of students: 137)

Number of classes in the comparison group:
N=6 classes from 3 schools out of 6 partici-pating schools

(total # of students: 141)

Number of classes remaining in the final analytic sample:

Student assessment:

- N=6 classes from 5 schools
- One class was dropped because they did not complete the post assessment.
- Two classes were merged because of the small sample size in one class.
- # of students: 98

Student attitude survey:

- N= 5 classes from 4 schools
- Two classes were merged because of the small sample size in one class.
- # of student: 75

Number of classes remaining in the final analytic sample:

Student assessment:

- N=6 classes from 3 schools
- # of student: 52
- Seven 8th graders were excluded be-cause they were not eligible for the study

Student attitude survey:

- N=4 classes from 2 schools
- Two classes were dropped because they did not complete the post sur-vey
- # of student: 26
- Seven 8th graders were excluded be-cause they were not eligible for the study

Figure 2. Learning by Making Curriculum Participant Flow Chart

The LbyM STEM curriculum comprises six units and focuses on the three-dimensional learning strategies of the NGSS, computer and computer programming language, and problem solving and troubleshooting. Based on the instructional design template provided by SSU, each unit includes 11 components: Objectives, Material List, Common Core Standards, NGSS, Background for the Teacher, Logo Vocabulary, Troubleshooting, Lesson (Procedure, Foundations, Going Further Extensions), Solutions, Teacher Materials, and Student Handout. Although the curriculum provides a suggested pacing guide for each lesson, it allows teachers to adapt to their own style.

Unit 1 is designed to introduce students to the class and to set the stage for the three dimensional learning strategies of the Next Generation Science Standards. It consists of two lessons, with four activities in Lesson 1 and two experiments in Lesson 2.

Unit 2 is designed to introduce students to the computer and computer programming language that they use throughout the course. It consists of seven lessons. Students begin Unit 2 by learning the structure and function of the HP Stream 11 and the Ubuntu operating system-Ubuntu, and they experience and practice TurtleLogo coding throughout the unit.

Unit 3 is designed to introduce basic electricity concepts in preparation for hardware troubleshooting in future experiments. Through three lessons in Unit 3, students learn about the underlying properties of power, voltage, current, and resistance. They also obtain practical skills through constructing functional circuits.

Unit 4 is designed to introduce how jLogo talks to uLogo, and how uLogo controls the outputs and inputs of analog devices. Through sixteen lessons with four experiments in Unit 4, students learn to define input, output, Analog-Digital-Conversion (ADC), and Analog to Digital Units (ADUs). They also practice three languages (TurtleLogo, jLogo, and uLogo) through the experiments to understand how analog sensor readings are converted to digital values using a mathematical relationship.

Unit 5 is designed to introduce heat transfer through conduction, convection, or radiation. Through nine lessons with three experiments in Unit 5, students explore different methods of heat transfer, convert sensor outputs in ADUs to meaningful quantities by deriving the linear relationships between ADU and degrees Celsius, and create plots in jLogo and explore mean, median, mode, and linear regressions.

Unit 6 provides students with the opportunity to create their first computational models of physical phenomenon. Through nine lessons in Unit 6, students design and conduct an experiment, build a MudWatt microbial fuel cell using locally collected soil, explore the random microscopic nature of heat transfer, create tactile models of heat transfer, and quantify the rate of heat transfer along different lengths of conductive wires.

SSU provides the LbyM curriculum, the equipment, and supporting materials. Students are expected to complete the six units in an academic year. Comparison students use business-as-usual mathematics and science curricula. These business-as-usual materials are not provided by SSU, but rather are educational materials that teachers normally have used with their students.

Instruments

Confirmatory Outcome Measures

Given that the NGSS-aligned science assessment is still under development in California and that there is no Smarter Balanced Summative Assessment in Math for 9th and 10th graders, to address the confirmatory research questions WestEd selected assessment items from the Certica Formative Assessment Item Bank. The Certica Formative Assessment Item Bank is a repository of mathematics items aligned to Common Core Standards in Mathematics and to NGSS. The selected assessment items focus on mathematics and science topic areas covered in the LbyM STEM curriculum. The science and

math assessments are not timed tests, and no precise time limits are required for students being tested. On average, the two assessments each take 40-50 minutes. All participating students (treatment and comparison) were asked to complete the math and science assessments in fall 2016 as the pre-test and in spring 2017 as the post-test.

Math Assessment. A total of eight math items were selected to measure two math topic areas: Linear and Non-linear Equations (5 items), and Research and Data Representation (3 items). The scores from the math assessment represented the percent of correctly answered items, which were calculated by summing the number of items correctly answered divided by the total number of items in the assessment (8 items). Therefore, the scores range from 0 to 100. The reliabilities on the pre-test and the post-test are 0.59 (KR-20) and 0.62 (KR-20), respectively.

Science Assessment. The science assessment includes 20 items that measure three science topic areas: Energy (8 items), Atmosphere and Weather (6 items), and Heat (6 items). The scores from the science assessment represented the percent of correctly answered items, which were calculated by summing the number of items correctly answered divided by the total number of items in the assessment (20 items). Therefore, the scores range from 0 to 100. The reliabilities on the pre-test and the post-test are 0.67 (KR-20) and 0.76 (KR-20), respectively.

Exploratory Outcome Measures

High School Student Attitudes Toward STEM Survey. The High School Student Attitudes Toward STEM Survey was developed by North Carolina State University (Faber, M., Unfried, A., Wiebe, E.N., Corn, J., & Collins, T.L., 2013). This survey includes 48 five-point Likert scale items and is designed to measure high school student attitudes toward science, mathematics, engineering and technology, and 21st Century skills. It also includes 16 items that measure student interest in STEM careers. Results from exploratory factor analysis indicated that the survey has good construct validity with four clear constructs measuring student attitudes toward science, math, engineering and technology, and 21st Century skills. The internal reliability coefficient was 0.83. All participating students (treatment and comparison) were asked to complete the survey in fall 2016 as the pre-survey and in spring 2017 as the post-survey.

Teacher Instructional Practice and Competency Survey. The teacher survey includes three sections. The first section collects teacher demographic data, including gender, ethnicity, academic and technology background, and teaching experience. The second section focuses on teacher instructional practices and includes 31 items adapted and adopted from Surveys of Enacted Curriculum (SEC) (Blank, Porter, & Smithson, 2001). The third section addresses teacher competencies and includes 34 items developed by WestEd and Sonoma State University in 2014.

The *teacher instructional practices* section includes three subscales of the SEC. All three subscales are five-point Likert-type scales ("1" = none, "5" = considerable). The first subscale addresses instructional practices in general and asks teachers to rate their instructional time spent on various activities in mathematics or science (14 items, α = .86) (e.g., How much of the instructional time do students use to engage in applying math/science concepts to "real-world" problems?). The second subscale addresses instructional practices supporting students to collect, organize, display and/or present data (7 items, α = . 94) (e.g., When students collect, organize, display and/or present data as part of the instruction, how

much of that time do they collect data by counting, measuring or observing?). The third subscale addresses instructional practices related to students' use of calculators, computers, and other educational technology (10 items, α = .82) (e.g., When students are engaged in activities that involve the use of calculators, computers, or other educational technology as part of the instruction, how much of that time do they practice skills and procedures?).

The *teacher competencies* section includes two subscales. The first subscale is a four-point Likert-type scale ("1" = need to learn this, "4" = I can teach other educators) that assesses teacher competencies in supporting students' critical thinking skills (12 items, α = .92) (e.g., I design learning activities that require students to apply existing knowledge to generate new ideas, products, or processes). The second subscale is a five-point Likert-type scale ("1" = strongly disagree, "5" = strongly agree) that measures teacher technology competencies (22 items, α = .93) (e.g., I know how to solve my own hardware problems).

Six of the LbyM teacher participants completed the survey before the Summer Institute in June 2016 and again before the Summer Institute in June 2017. For this report, pre to post changes in teacher instructional practices and competencies will be reported based on the six teachers' survey responses. Because of the small sample size, pre and post data were analyzed descriptively to provide a general picture of potential changes in teacher instructional practices and competencies through the LbyM PD and teaching experience. The results should be interpreted with caution and no statistical inferences should be made based on the results.

Implementation Measures

Sign-in. A daily sign-in sheet was used to track teachers' attendance during the Summer Institute and follow-up training sessions.

Teacher implementation logs. Teachers were asked to complete an online implementation log every other week throughout the 2016-17 academic year. The teacher logs included questions about how many LbyM units each class completed, which lessons were skipped, and which lessons were modified. Teachers were also asked to comment on the successes they had experienced with LbyM over the past two weeks, as well as the challenges they experienced. The overall response rate for the implementation logs was 55%.

Summer Institute observation. Evaluators observed the 4-day Summer Institute in June 2016. The observation focused on (1) facilitators' delivery of key content; and (2) participating teachers' engagement in the training.

Teacher focus groups. Evaluators conducted teacher focus groups at the end of the Summer Institute. All teachers were invited and eleven participated in the focus groups. The teacher focus groups captured teachers' overall reactions to Learning by Making content and practices, and ease or difficulty of engaging in the PD.

Teacher satisfaction survey. Participating teachers were asked to complete a satisfaction survey after the Summer Institute. The survey captured teachers' overall satisfaction with the training logistics and content.

Classroom observations. Researchers observed classrooms during the 2016-2017 academic year. The observations focused on the teachers' instructional practice. We adapted the classroom observation protocol from Horizon Classroom Observation Protocol (Weiss, Pasley, Smith, Banilower, & Heck, 2003). Dimensions observed included the richness of the content, student reasoning and meaning making, and the clarity and correctness of the content, which is consistent with the LbyM model.

Teacher interviews. Teachers were interviewed after each classroom observation. Topics discussed in the interviews included those directly related to the classroom observation, and broader topics addressing the LbyM curriculum and its implementation and effects more broadly.

Sample Characteristics

While it is not typical to have 8th grade students participating in courses at high schools, there were seven 8th grade students in one comparison classroom. The seven 8th grade students participated in a 9th grade algebra course since one was not offered in their middle school. Given that the seven 8th grade students did not participate in any other high school courses and were not official high school students, they were not eligible for a LbyM course. Therefore, these 8th grade students were excluded from the final analytic sample[1]. The final analytic sample for student math and science achievement includes a total of 150 students from 12 classrooms in five rural high schools with both non-missing pre and post assessment data. The final analytic sample for student attitudes toward STEM includes a total of 101 students from nine classrooms in four rural high schools with both non-missing pre and post attitude survey data.

Table 2 shows that the majority of the students in treatment group were in 9th grade, whereas the grade levels in comparison group ranged from 9th grade to 12th grade. The treatment and comparison groups were equivalent at baseline as measured by the LbyM science assessment (effect size <0.05). The difference between treatment and comparison on the baseline math assessment was adjustable (0.05< effect size <0.25; Table 3). In addition, the treatment and comparison groups were equivalent at baseline on their attitudes toward math (effect size <0.05). The baseline difference between treatment and comparison groups on their attitudes toward engineering was adjustable (0.05< effect size <0.25). However, treatment and comparison groups were not equivalent at baseline on their attitudes toward science and 21st Century skills (effect size > 0.25; Table 4). Tables A1 and A2 in Appendix A provide the unadjusted means and standard deviations of student pre assessments and pre attitude measures for the final analytic sample, respectively.

---

[1] Baseline equivalence tests, student math and science outcomes, and student attitudes toward STEM for the sample with 8th graders are presented in Tables B1-B7 of Appendix A.

Table 2. Participants' Demographic Information, by Experimental Condition

	Comparison		Treatment		
	Number	**Percent[a]**	**Number**	**Percent**	**p-value[b]**
Grade level					0.051
9th grade	24	46.15%	87	88.78%	--
10th grade	12	23.08%	8	8.16%	--
11th grade	11	21.15%	0	0%	--
12th grade	5	9.62%	3	3.6%	--

a. Computed based on valid (non-missing) data. Components may not sum to 100 because of rounding.

b. A test for equality of proportion between treatment and comparison students was conducted, and the corrected p-value, accounting for clustering effects (students were nested with schools), was reported here.

Table 3. Baseline Equivalence for Student Pre Assessments, by Experimental Condition

Achievement Measure	Comparison	Treatment	Difference[c]	p-value	SMD[d]
LbyM Science Assessment					
Mean	40.78	40.88	-0.10	0.98	0.005
Standard deviation	20.59	16.26	--	--	--
N	52	98	--	--	--
LbyM Math Assessment					
Mean	45.95	43.49	2.45	0.76	0.104
Standard deviation	23.55	24.03	--	--	--
N	52	98	--	--	--

c. Computed based on valid (non-missing) data. A multi-level regression model that accounted for clustering effects (students were nested with classrooms) was used to test whether students' math and science pretests at baseline were equivalent between treatment and comparison groups.

d. SMD refers to Standardized Mean Difference which was calculated by dividing treatment and comparison difference by the comparison group standard deviation of the pre measure variable.

Table 4. Baseline Equivalence for Student Pre Attitude Measures, by Experimental Condition

Attitude Measure	Comparison	Treatment	Difference[e]	p-value	SMD[f]
Student Attitude towards Math					
Mean	3.37	3.34	0.03	0.90	0.03
Standard deviation	1.00	0.78	--	--	--
N	26	75	--	--	--
Student Attitude towards Science					
Mean	3.45	3.17	0.28	0.14	0.38
Standard deviation	0.74	0.57	--	--	--
N	24	75	--	--	--
Student Attitude towards Technology and Engineering					
Mean	3.32	3.47	-0.15	0.33	-0.22
Standard deviation	0.69	0.69	--	--	--
N	26	74	--	--	--
Student Attitude towards 21st Century Skills					
Mean	4.20	3.90	0.30	0.02[*]	0.57
Standard deviation	0.53	0.52	--	--	--
N	26	73	--	--	--

e. Computed based on valid (non-missing) data. A multi-level regression model that accounted for clustering effects (students were nested with classrooms) was used to test whether students' attitudes towards math, science, technology and engineering, and 21st Century skills pretests at baseline were equivalent between treatment and comparison groups.

f. SMD refers to Standardized Mean Difference which was calculated by dividing treatment and comparison difference by the comparison group standard deviation of the pre measure variable.

Data Analysis Methods

Several approaches were utilized to address missing data. First, students who did not take both the pre- and post-content assessments were excluded from the analytic sample for that outcome measure. Students who did not take both the pre- and post-STEM attitudes surveys were excluded from the analytic sample for that outcome measure. On the content assessments, missing item responses were treated as incorrect responses. On the student attitudes toward STEM surveys, items missing responses were excluded from analysis; mean scores for each scale were calculated using the total number of questions answered for that scale, so as not to deflate the mean artificially. Researchers decided to exclude data from any students who skipped more than 10% of questions on the assessments or surveys, but no students met this criteria for exclusion.

The distribution of scores on the pre- and post-content assessments and surveys was analyzed by class period for excessive skewness, but none was found, indicating that there is no need to remove any students from the analytic sample due to outlying scores.

Student Content Assessment

Assignment to the intervention or comparison groups occurred at the classroom level, creating a "cluster" intervention design. Given this design, a hierarchical linear model was used to investigate the influence of the LbyM curriculum on student math and science performance (postMATH and postSCIENCE), structuring students nested within class periods. This model is preferred to more

precisely estimate the influence of the intervention when it is necessary to account for within-group and between-group effects (Raudenbush, 1997). To control for student performance at baseline, we included students' pretest scores on the content assessment (preMATH or preSCIENCE) as a covariate in the Level 1 model. Three school-level demographic information items were used as covariates in the outcome models estimating the treatment effects. These demographic data are publicly available from the California Department of Education.

- EL: Percent of English learners who have not yet been reclassified as English proficient

- MIN: Percent of non-White students, including Latino students

- FRL: Percent of socio-economically disadvantaged students, represented by the proxy measure of the percent of students who receive free or reduced price lunch.

The Level 2 model included the intervention variable (TREAT). A series of sensitivity analyses were conducted to confirm this modeling choice (Tables A3 and A4 in Appendix A). The main effect model was specified using the covariates listed above, and run using Stata 14 statistical analysis software. The models were specified as follows:

$$\text{Level 1: postMATH}_{ij} = \pi_{0j} + \pi_{1j}(\text{preMATH})_{ij} + e_{ij}$$
$$\text{Level 2: } \pi_{0j} = \beta_{00} + \beta_{01}(\text{TREAT})_j + \beta_{02}(\text{EL})_j + \beta_{03}(\text{MIN})_j + \beta_{04}(\text{FRL})_j + r_{0j}$$
$$\pi_{1j} = \beta_{10}$$

$$\text{Level 1: postSCIENCE}_{ij} = \pi_{0j} + \pi_{1j}(\text{preSCIENCE})_{ij} + e_{ij}$$
$$\text{Level 2: } \pi_{0j} = \beta_{00} + \beta_{01}(\text{TREAT})_j + \beta_{02}(\text{EL})_j + \beta_{03}(\text{MIN})_j + \beta_{04}(\text{FRL})_j + r_{0j}$$
$$\pi_{1j} = \beta_{10}$$

This model allowed researchers to compare the intervention group's post-intervention outcomes with those of the comparison group, after adjusting for difference in baseline scores and demographic factors. Adjusted intervention and comparison group means were reported for each outcome variable.

Teacher Instructional Practice and Competency Survey

To explore changes in teacher instructional practices and technological competencies, each of the survey subscales was examined independently. Six of the LbyM teacher participants completed the survey before the Summer Institute in June 2016 and again before the Summer Institute in June 2017. For this report, pre to post changes on teacher instructional practices and competencies are reported based on the six teachers' survey responses. Because of the small sample size, pre and post data were analyzed descriptively to provide a general picture of potential changes on teacher instructional practices and competencies through the LbyM PD and teaching experience. The results should be interpreted with caution and no statistical inferences should be made based on the results.

Student Attitudes Towards STEM Survey

To explore the effect of the intervention on student attitudes toward STEM subjects and careers, separate analyses were conducted on outcomes from two sections of the STEM attitudes survey. First, per the preferred reporting method of the survey developers (Faber, M. et al., 2013), mean scores were calculated for each of the four content area scales (Math, Science, Engineering and Technology, and 21st Century Skills) by treatment group. Each scale contains between eight and 11 items, using a five-point scale, with 1 representing "Strongly Disagree" and 5 representing "Strongly Agree." Four negatively-worded items were reverse coded, so that a high value indicates the same tone of response as for a positively-worded item. A hierarchical linear model was used to examine the effect of the treatment on student attitudes, as measured by these four area scales. The model employed mirrored the model used to analyze the content assessment data, as shown below, where SCALE represents one area scale.

$$\text{Level 1: postSCALE}_{ij} = \pi_{0j} + \pi_{1j}(\text{preSCALE})_{ij} + e_{ij}$$
$$\text{Level 2: } \pi_{0j} = \beta_{00} + \beta_{01}(\text{TREAT})_j + \beta_{02}(\text{EL})_j + \beta_{03}(\text{MIN})_j + \beta_{04}(\text{FRL})_j + r_{0j}$$
$$\pi_{1j} = \beta_{10}$$

This model allowed researchers to compare the intervention group's post-intervention attitudinal outcomes with those of the comparison group, after adjusting for difference in baseline responses and demographic factors. Adjusted intervention and comparison group means were reported for each scale in the survey.

Second, chi-square analyses were used to determine if students' interest in a variety of STEM careers changed significantly when comparing baseline scores to those at the end of the intervention. Chi-square analyses are the preferred method for comparing nominal categories, such as "interested" and "not interested," that often occur in survey data (Lynch, 2013). The STEM attitudes survey included 12 items asking if students were "Not interested at all," "Not so interested," "Interested," or "Very interested" in a series of STEM careers including: Physics, Environmental work, Biology and Zoology, Veterinary work, Mathematics, Medicine, Earth Science, Computer Science, Chemistry, Energy, and Engineering. Responses of "Not interested at all" or "Not so interested" were collapsed to "Not interested," while responses of "Interested" or "Very interested" were collapsed to "Interested." Chi-square statistics were calculated for treatment group differences at baseline, and then again after the intervention. Researchers then compared chi-square statistics between pre- and post-surveys to look for changes in group differences after the intervention.

Learning by Making Implementation

Teacher Professional Development

Attendance. Teacher attendance for the professional development trainings was consistently high. For the 2016-2017 Summer Institute, there was a 92% teacher attendance rate on four out five days of the week. On day three, there was an 83% teacher attendance rate. For the follow-up professional development training sessions in October, December, and March, there was 100% teacher attendance. In January, 92% of teachers attended the follow-up training, and in April 83% attended. Appendix B shows LbyM fidelity matrix and the fidelity of implementation results in the 2016-2017 academic year.

Fidelity of Implementation at Intervention Sites

Curriculum dosage and fidelity. Overall curriculum dosage and fidelity to the curriculum were measured using bi-monthly online teacher logs. All teachers in the study finished the 2016-17 school year with a moderate fidelity rating. In total, there were six units included in the 2016-17 LbyM curriculum. Four classes completed the LbyM lessons nearly through Unit 4—up to Lesson 13, 14, or 16. Two classes completed the lessons through the beginning of Unit 6. While classes completed most of the LbyM lessons in the order outlined in the curriculum, there were instances where teachers skipped lessons or completed lessons out of order. In certain cases, teachers skipped lessons due to time constraints, feeling that it was more important to move forward in the curriculum than to teach every lesson in a unit. In other cases, teachers felt that an activity was not particularly relevant or engaging for their students. Teachers also explained that they skipped certain lessons to begin Turtle Logo coding—as students expressed enthusiasm about using Logo—but came back to those lessons later. One teacher skipped Unit 5 altogether because the year was ending and she wanted her students to experience Unit 6: Mud Watts.

Most teachers reported modifying LbyM lessons when necessitated by their classroom context. Two primary reasons emerged from the analysis that explain teachers' reasons for modifying the lessons: 1) to provide content background for students in preparation for upcoming LbyM lessons, and 2) to provide opportunities for students to practice the skills that they learn through LbyM. Teachers reported that many of their students enter high school with low abilities, especially in math. Since the LbyM lessons require some background knowledge in science and algebra, some teachers felt the need to scaffold the LbyM curriculum with content-specific supplementary lessons that they developed themselves. One teacher, for example, explained, "I'm noticing that when they [SSU] give an activity, there's a lot of prep work I have to do in order to do the activity. Like for this [referring to the bacterial growth lab], I had to teach them graphing. They don't know how to graph. I put up really simple data— number of tadpoles, temperature, or something—and I got all kinds of shapes [from the students]. They had no idea how to graph. So, it took me a few days to teach graphing before I could even jump into this." Another teacher shared a similar situation he encountered, saying, "I did more frontloading of atoms and electrons and how elements work. Because if you're going to sit there and talk about conductivity, you really want to talk about why metals are special. I wanted them to have a little bit of that background."

Another way in which teachers reported supplementing the LbyM curriculum was to create opportunities for practice. Because the computer skills were completely new to many students, one teacher explained, they needed a lot of time to practice, especially once they reached the lessons involving Logo. One teacher, for example, created a lesson where students collaboratively wrote a story by printing one line of text at a time in the Logo console. The goal of this lesson was to build comfort using basic Logo commands to print text. The students began the lesson at a slower pace, often referring to the Logo documentation or asking each other for help with coding syntax. With each round of printing a new line of text, however, students worked faster; the repetition appeared to reinforce their confidence and their abilities to complete the coding task at hand. While this lesson was related to LbyM learning objectives, the teacher created it in response to her perception that students needed applied practice with this topic before moving on in the unit.

Challenges for implementation. While implementation challenges varied by school site, findings revealed that the most common barriers to implementation were student absenteeism and the time-consuming nature of teaching coding. Student absenteeism or inconsistent attendance was a problem faced by several teachers in the study, especially in the smaller schools. The most commonly reported reason for student absenteeism was a repeating conflict with athletics, but other reasons included skipping school during fishing season and leaving town for families' seasonal work. One teacher plainly stated, "Student absenteeism is, I would say, my greatest challenge." She went on to elaborate: "This is a class where the skills build on the last skill....I don't think it would be possible to really achieve in the area of computer programming if you missed too much, because you have to keep building your skills." Teachers explained that the academic consequences of absenteeism were more acute in LbyM than other classes, because there were no structures or materials in place for students to make up the work.

According to the teachers in the study, the primary reason that they did not complete the curriculum with their class—making it only to Unit 4 or 6—was that Units 3 and 4 were too long, consuming many more class periods than they had anticipated. Learning the many layers of Logo syntax was challenging for students, many of whom had never worked with computer code before. Eventually, students grew tired of coding on the computers, and teachers sensed that they needed other activities to provide more variety. Most classrooms never reached the phase of the curriculum where students would have begun to use Logo to run their own experiments.

Results

Student Mathematics and Science Performance

Results show that the intervention was associated with significant gains in science content knowledge (p < 0.01; effect size = 0.34), as indicated in Table 5. The students in the treatment group, on average, scored over seven points more than those in the comparison group, which represents a medium effect size of 0.34. Although there was no significant difference between the student scores for treatment and comparison groups on math content knowledge, the students in the treatment group scored over four points more than those in the comparison group with an effect size of 0.17. These assessment data were analyzed using the eligible analytic sample of 9th through 12th grade students who completed both the pre- and post-assessments (N = 150). Analysis including the seven 8th grade students who participated in the control group can be found in Appendix A.

Table 5. Student Math and Science Outcomes (N=150)

Outcome Measure	Covariate-Adjusted Mean (SD)		Covariate-Adjusted Mean Difference[g]	p-Value	95% Confidence Interval	Effect Size[h]
	Treatment	Comparison				
postMATH	50.73 (24.77)	46.21 (26.08)	4.52	0.25	-3.10 – 12.14	0.17
N	98	52				
postSCI	49.36 (19.22)	41.49 (22.95)	7.87	0.01	1.85 – 13.88	0.34
N	98	52				

g. Data were regression-adjusted using multi-level regression models to account for differences in baseline characteristics.

h. Effect size was calculated by dividing impact estimates by the comparison group standard deviation of the outcome variable.

Analysis of the data generated in teacher interviews, focus groups, and implementation logs supports the observed intervention group growth in math performance. Several teachers, for example, noted student improvement in math, particularly for their lower-achieving students. One teacher stated,

> "One of the things that I noticed last year but I feel like it's really happening [now] is that the kids are learning a lot of math. I feel like there's such a connection between what we do in this class…So I just think that the math that's embedded in the curriculum is really incredible. It's accessible to a real spectrum of kids with different math skills."

Another teacher echoed this sentiment, explaining that the curriculum is a great vessel for learning about graphing, with which many of her students—even the seniors—struggle. Given the significant treatment effect for science performance, presented in Table 5, it is surprising to note that teachers felt there was not enough science content covered in the LbyM curriculum. While teachers agreed that LbyM integrates science with math, technology, and engineering, they felt that including additional science content would serve to benefit student science performance even more. One teacher described

how she felt the technology components were over-emphasized at the expense of science content: "It's still too much technology. Too much curriculum to get through the year. It's all valuable and we see where it fits in…but it has to be cut back in order to get through everything that we need to cover, in order to get to the experiments, in order to cover into the MudWatts [microbes unit] and the biology."

Teacher Instructional Practices and Technological Competency

Findings from the teacher instructional practices and technological competency survey, along with classroom observations and teacher interviews and focus groups, show that teachers' approaches to instruction have changed in some areas more than others since the start of the study. The most observable changes between pre- and post-survey responses from six teachers in five schools occurred in the following two areas: 1) teachers' instructional practices in supporting students to collect, organize, display and/or present data, and 2) teachers' own technological competencies.

With regard to the amount of instructional time during which students worked with data, pre- to post-survey changes were observed across several activities. The percentage of teachers who reported spending at least 25% of instructional time dedicated to *collecting data by counting, measuring, or observing* rose from 40% at pre-test to 83% at post-test. The percentage of teachers who reported spending at least 25% of instructional time dedicated to *displaying and analyzing data* rose from 20% to 83%. The percentage of teachers who reported spending at least 50% of instructional time on *using technology to solve problems* rose from 20% to 67%. The percentage of teachers who reported spending at least 50% of instructional time on *organizing, outlining, or summarizing information* rose from 60% to 100%.

Findings from teacher interviews and focus groups indicate that they view data collection and analysis as aligned with the practice- and process-based NGSS goals. One teacher, for example, reported that the experience of teaching LbyM, combined with training from SSU, has helped him transition to a more NGSS-aligned approach of project-based learning. He explained,

> "The idea of giving [students] a little bit of information about a thing, letting them practice it, seeing how to build the skill, and then let them come up with their means of going through a project. I felt like if I didn't have [LbyM]…I would have a hard time making the transition to the whole project-based learning thing."

He added, "I've become more student centered. It may have been part of this program, too, because this program is very student centered." Other teachers echo this sentiment, describing how they feel increasingly comfortable letting the students direct their own inquiry, and work through their own problems. One teacher shared, "I love when I step back, because really I'm steering the ship, but they're really the ones that are making it go. And I like that." These teacher perspectives help to explain the increases in instructional time dedicated to data collection, analysis, and presentation.

Noticeable changes from pre- to post-survey were also observed in teachers' reported technological competencies. The percentage of teachers who responded that they either "strongly agree" or "agree" with the following statements at pre- and post-survey are displayed in Table 6. These changes suggest a

substantial increase in teachers' level of comfort with technology and perceived technological capabilities.

Table 6. Percentage of Teachers Who Either "Strongly Agree" or "Agree" with Statements of Technological Competency at Pre- and Post-survey (N = 6)

Survey item	Pre-survey	Post-survey
I know how to solve my own hardware problems.	40%	100%
I know how to solve my own software problems.	40%	80%
I can solve basic electronic problems.	40%	100%
I can use electronics tools to measure voltage, current, and resistance.	40%	100%
I can build and debug electronic circuits that will read data from sensors.	40%	100%
I know about technologies that I can use for understanding and doing mathematics.	40%	83%
I know about technologies that I can use for understanding and doing science.	60%	100%

After working with LbyM for two years and receiving consistent professional development from SSU, teachers explained that they felt more confident teaching the course to their students, given the technological requirements of the curriculum. As one teacher shared, "I'm like a different teacher than I was last year—I can feel it!" She reported feeling more knowledgeable now about the process of coding, which allowed her not only to debug more technical issues that came up, but also to manage her classroom better. For example, she knew at which points during coding exercises she needed to instruct students to test their code to check for errors. Being able to foresee coding challenges, she explained, allowed her to plan her lessons better.

Student STEM Attitudes

Results indicate no significant changes in student attitudes toward math, science, engineering, or 21st Century skills as measured by the area scales in the STEM attitudes survey that can be attributed to the LbyM curriculum (Table 7). While moderate effect sizes are observed for the science (d = 0.29) and 21st Century skills (d = 0.30) area scales, these effects are not statistically reliable. These models were run using the eligible analytic sample of 9th through 12th grade students who completed both the pre- and post-assessments (N = 101). Results including the seven 8th grade students who participated in the control group can be found in Appendix A.

Table 7. Adjusted Means of Student Responses by Scale (N = 101)

Scale	Covariate-Adjusted Mean (SD)		Covariate-Adjusted Mean Difference[i]	p-Value	95% Confidence Interval	Effect Size[j]
	Treatment	Comparison				
Math	3.26 (0.87)	3.30 (1.07)	-0.05	0.83	-0.47 – 0.37	-0.05
N	75	26				
Science	3.17 (0.70)	2.95 (0.76)	0.22	0.40	-0.29 – 0.74	0.29
N	75	24				
Engineering	3.26 (0.85)	3.28 (0.78)	-0.02	0.94	-0.58 – 0.53	-0.03
N	74	26				
21st Century	3.97 (0.63)	3.81 (0.54)	0.16	0.49	-0.29 – 0.61	0.30
N	73	26				

i. Data were regression-adjusted using multi-level regression models to account for differences in baseline characteristics.

j. Effect size was calculated by dividing impact estimates by the comparison group standard deviation of the outcome variable.

To understand whether the LbyM curriculum influenced students' interest in pursuing a career across several STEM fields, researchers conducted chi-square analyses by treatment group for both the pre- and post-surveys. The percentages of each group who reported being "Interested" and "Not Interested" in each of twelve STEM careers, and the associated chi-square statistic p-values, are presented in Table 8. For the pre-survey, differences between treatment group and comparison group responses were found to be statistically significant at α = 0.05 for the careers of Chemistry (X^2 (1, N = 101) = 3.93, p = 0.05) and Engineering (X^2 (1, N = 101) = 7.79, p = 0.01). For the post-survey, no differences between treatment group and comparison group responses were found to be statistically significant, suggesting that LbyM did not influence student attitudes toward STEM careers as measured by the student STEM attitude survey.

Table 8. Chi-squared Test of Student Interest in STEM Careers (N = 101)

Field	Pre-Intervention					Post-Intervention				
	Intervention		Comparison		P-value	Intervention		Comparison		P-value
	N	%	N	%		N	%	N	%	
Physics					0.26					0.44
Not interested	47	62.67	13	50.00		55	73.33	17	65.38	
Interested	28	37.33	13	50.00		20	26.67	9	34.62	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Env. Science					0.11					0.18
Not interested	48	64.00	21	80.77		59	78.67	17	65.38	
Interested	27	36.00	5	19.23		16	21.33	9	34.62	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Biology					0.37					0.93
Not interested	45	60.00	13	50.00		44	58.67	15	57.69	
Interested	30	40.00	13	50.00		31	41.33	11	42.31	
Total	*75*	*100.00*	*58*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Vet. Medicine					0.63					0.19
Not interested	45	60.00	17	65.38		47	62.67	20	76.92	
Interested	30	40.00	9	34.62		28	37.33	6	23.08	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Mathematics					0.47					0.22
Not interested	46	61.33	18	69.23		53	70.67	15	57.69	
Interested	29	38.67	8	30.77		22	29.33	11	42.31	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Medicine					0.46					0.34
Not interested	38	50.67	11	42.31		37	49.33	10	38.46	
Interested	37	49.33	15	57.69		38	50.67	16	61.54	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Earth Science					0.72					0.97
Not interested	52	69.33	19	73.08		58	77.33	20	76.92	
Interested	23	30.67	7	26.92		17	22.67	6	23.08	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Comp. Science					0.62					0.24
Not interested	42	56.00	16	61.54		50	66.67	14	53.85	
Interested	33	44.00	10	38.46		25	33.33	12	46.15	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Med. Science					0.37					0.67
Not interested	45	60.00	13	50.00		44	58.67	14	53.85	
Interested	30	40.00	13	50.00		31	41.33	12	46.15	
Total	*75*	*100.00*	*26*	*100.00*		*45*	*100.00*	*26*	*100.00*	
Chemistry					**0.05***					0.15
Not interested	51	68.00	12	46.15		52	69.33	14	53.85	
Interested	24	32.00	14	53.85		23	30.67	12	46.15	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Energy					0.40					0.53
Not interested	45	60.00	18	69.23		54	72.00	17	65.38	
Interested	30	40.00	8	30.77		21	28.00	9	34.62	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	
Engineering					**0.01****					0.16
Not interested	31	41.33	19	73.08		40	53.33	18	69.23	
Interested	44	58.67	7	26.92		35	46.67	8	30.77	
Total	*75*	*100.00*	*26*	*100.00*		*75*	*100.00*	*26*	*100.00*	

* Significant at $\alpha = 0.05$
** Significant at $\alpha = 0.01$

Despite the lack of evidence from the STEM attitudes survey that student attitudes toward STEM subjects improved as a result of the LbyM experience, an analysis of data from teacher interviews, focus groups, and implementation logs indicates that students were highly engaged by the LbyM curriculum, and benefited from the non-traditional, inquiry-driven structure of the course. Findings related to student engagement emerged around student enthusiasm for the LbyM course, increased problem-solving stamina, and inclusion of students with different academic abilities.

Overall, teachers report high levels of student engagement, and general student satisfaction with the LbyM curriculum. Students are especially excited by the opportunity to conduct high-tech experiments. For example, students were eager to use the experimental set-up for the Heat and Sand experiment because it is exciting and unknown to them. In describing her students' enthusiasm for computer coding, one of the teachers who was observed this fall recalled, "Kids would come in a few minutes before school started. They would get their laptop out, and they would start working, and they would just be here. And I've never had that in a first period class, where the kids came, and they just started working without any instruction from me." Another teacher recounted how a group of girls she had recruited to join the class, though hesitant at first, became fully immersed in learning how to code. She noted with great pride, "By the end of the year, they were like, 'I want to be a computer engineer!'" One of the teachers noted that to keep the students engaged in any kind of learning activity, "You have to make it fun." She said the LbyM curriculum lends itself that that kind of engaging, fun learning.

Over the course of the year, students developed the persistence and positive attitude needed to solve problems they encountered. "I think they learn problem solving techniques," one teacher shared. "I think they learn to feel good about themselves when they accomplished something." Acquiring this skill was not without a struggle. Teachers explained that their students were not accustomed to a curriculum like LbyM, where the answers are not spelled out in a textbook. Initially, they became frustrated when they could not figure out key concepts right away. After a while, however, students adapted to the notion that struggling is part of LbyM, as one teacher described:

> "After they were frustrated and finally worked through it and had success a couple times, then it was no problem, then it became a game of, 'How do I figure out how this fits together?' . . . and they'd start looking over here and that person would be like, 'Don't tell me yet,' . . . and it's like, 'I want to figure it out myself.'"

One teacher commented on how the challenges of the curriculum actually engaged students more in the long-run, reflecting, "The difficulty is just getting them to put the energy in. It does take more work on their part, but they also have more fun."

One feature that teachers found to be unique to the LbyM curriculum was its ability to engage students who typically struggle in school. Findings from teacher interviews and focus groups revealed that some individual students who typically struggle to participate exhibited higher levels of participation in LbyM, and even demonstrated leadership in certain instances. As one teacher commented, "What I like about it is that different people are shining," referring to her own observations of typically disengaged students volunteering to lead certain tasks. Another teacher shared, "I think that kids that maybe haven't always felt strong come into this class, and they're like, 'Oh wow, I am strong. I'm doing really well.' Or, 'I get it!'" Two separate teachers reported that their students with special needs, while still

requiring extra attention, remained engaged in the curriculum, and were even quicker to complete certain activities than the other students. "There's some lessons where she's the only one who gets it," one teacher explained, referring to a student with special needs. "When we were learning the Tower of Hanoi, all of the other kids were experimenting, and she just got it. She went through the entire process perfectly with no errors." Similarly, another teacher described an English-learning student who, though typically shy and quiet, had begun to consistently take the lead in computer coding activities. Teachers proposed that the alternative structure of the curriculum could be responsible for the increased participation of students who often disengage.

Discussion

The current study describes the implementation and impact of a highly innovative high school integrated science curriculum that prompts students and teachers to use computational thinking and technological tools and resources to solve meaningful, real-world problems. Building on traditional project-based learning curricula, the LbyM STEM curriculum allows students and teachers to focus on CCSS-M and NGSS, including the Scientific and Engineering Practices, in a collaborative environment where teachers and students work together to conduct investigations and solve meaningful and complex problems related to science topics. The structure of the curriculum allows teachers to learn alongside their students as they engage in purposeful research. Findings from the study contribute to our knowledge of how an innovative, technology and problem-based curriculum can change the way science and mathematics instruction is delivered in rural high schools. They also provide insights around the value of a curriculum based on computational thinking; how technology-rich, inquiry- and PBL-based learning environments support learning and engagement in NGSS and CCSS-M; and best practices in teacher professional development related to this innovative curriculum.

Contributing to the Research Base in Rural Science Education

Research in rural education suggests that student engagement, motivation, and learning is enhanced when instructional strategies focus on valuing and using local knowledge in science education (Avery, 2013), and when instruction supports students' interests and is connected to their lives (Hardre & Reeve, 2003). In addition, inquiry-based science instruction has been shown to be effective in rural secondary science education (Marshall & Allston, 2014). Findings from the current study support results from this previous research. Findings from analysis of teacher focus groups and interviews suggest that students responded positively to the inquiry-based activities in the curriculum, and recognized the relevance of the tasks and investigations to their everyday lives.

Using Computational Thinking as a Foundation for Learning

The study also provides evidence around the topic of computational thinking as a foundation for instruction in STEM fields (National Research Council, 2010, 2011; Yadav, Hong, & Stephenson, 2016), and the growing research base suggesting that STEM curricula based on computational thinking can promote learning in in rural schools (Cham, 2017; Repenning, et al., 2015). The LbyM intervention is, in part, based on students' and teachers' use of Logo programming activities to solve problems in STEM subject areas. Logo was specifically designed to foster computational thinking in education (Papert, 1999). Despite struggles during the school year around the appropriate levels of complexity of the Logo syntax used in the intervention, the computational thinking aspects of the LbyM curriculum were enthusiastically taken up and expanded upon by students and teachers alike. Teachers reported that students were motivated by the coding aspects of the curriculum, and often mastered coding tasks more quickly than teachers. These findings are aligned with findings from other research studies focusing on STEM curricula based on computational thinking. Future study of LbyM should specifically focus on the role of computational thinking with regard to learning outcomes.

Building Teachers' Capacity to Deliver a Technology-Rich Inquiry-Based Curriculum

The intervention in the current study included a curriculum that incorporates activities and investigations adapted from project-based learning (PBL) models to include the use of technological tools and resources, including electronic sensors, breadboards, and programming resources. The intervention included intensive professional development and support to teachers to build their capacity to effectively deliver this new model of instruction. The findings of the current study contribute to a growing research base around best practices in supporting teachers to deliver instruction in technology-rich, PBL-based learning environments that support learning in NGSS and CCSS-M. Two strands of research address professional development strategies to support this type of learning in STEM classrooms. One strand focuses on a professional development framework for the development of teachers' Technological Pedagogical and Science Content Knowledge, or TPASK (Angeli & Valanides, 2009). Adapted from Mishra and Koehler's (2007) TPACK model (Technological Pedagogical Content Knowledge), TPASK professional development models include capacity-building in the areas of the use of technology, science content knowledge, and pedagogies for effective science instruction using technology. The TPASK framework has been found to be a useful model for teachers to use in lesson design (Angeli & Valanides, 2009) and has been found to support instruction that promotes learning in NGSS core ideas (Koh, Chai, & Lim, 2017) and practices (Harris, 2016; Koh, Chai, & Lim, 2017).

A second strand of research that addresses professional development strategies to support technology-rich, PBL-based learning environments in STEM classrooms involves instructional strategies that promote "deeper learning," what researchers call a collection of student competencies, including: mastering rigorous academic content; learning how to think critically and solve problems; working collaboratively; communicating effectively; directing one's own learning; and developing an academic mindset and a belief in one's ability to grow (Grover & Pea, 2016; Heller, Wolfe, & Steinberg, 2017; Hewlett Foundation, 2017; Pellegrino & Hilton, 2013). Instructional strategies that promote deeper learning include student-led and collaborative PBL, use of technological computational thinking practices and resources, and strategies that promote higher-order thinking (Grover & Pea, 2016; Heller, Wolfe, & Steinberg, 2017; Hewlett Foundation, 2017; Nogera, 2017).

The LbyM teacher professional development provided by SSU during the current study's intervention included many aspects included in TPASK framework, including supporting science content knowledge and science pedagogical content knowledge, as well as knowledge about technologies. The professional development was successful in building teachers' capacity to effectively use technologies to support science learning. The intervention also included many of the strategies suggested in deeper learning professional development models, including student-led and collaborative PBL; computational thinking as an integral part of the curriculum; and strategies that promote higher-order thinking, such as complex problem-solving, critical thinking, and effective communication. Findings from the current study contribute to recent research related to building teacher capacity to deliver a technology-rich inquiry-based curriculum to support students' mastery of NGSS and CCSS-M.

Conclusion

The 2016-17 evaluation of LbyM was designed to test whether the curriculum impacts student performance in math and science, to determine the extent to which teacher instructional practices and technological competencies are influenced by the curriculum, and to gauge whether the course impacts student attitudes toward STEM. Student outcome data revealed that of the 9th through 12th grade students in the sample, LbyM students scored statistically higher on the science content assessment following the intervention than students in the comparison group, but indicated no difference between groups on the math content assessment. Findings from teacher input indicate that students benefit academically from LbyM, so it would be beneficial to continue to monitor indicators of student academic growth during LbyM course implementation. Findings from the teacher survey on instructional practices and technological competency reveal that teachers are dedicating more class time to specific data collection, analysis, and display, and that their own comfort with technology has increased since teaching LbyM. While quantitative evidence of changes in student attitudes toward STEM is inconclusive, analysis of data from observations, teacher interviews, teacher focus groups, and implementation logs shows high levels of student engagement and an increase in students' stamina for problem solving and overcoming unfamiliar challenges in the classroom. Teachers agreed that LbyM is especially beneficial for many students who do not excel in traditional high school classes.

States and districts across the country continue to integrate NGSS and CCSS-M into their high school curricula. However, studies on integrated curricula—and specifically on the effectiveness of technology- and computer science-based programs on student math and science performance—are lacking, highlighting the importance for continued work on such curricula (Pang & Good, 2000). Evidence from this study suggest that LbyM is associated with increases in student math and science performance, although further research with a larger student sample would be beneficial to examine these increases with more precision.

References

Angeli, C. & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). Computers & Education, 52 (1), 154–168.

Avery, L. (2013) Rural science education: Valuing local knowledge. Theory Into Practice 52 (1), New York: Taylor and Francis.

Grover, S. & Pea, R. (2016). Designing a blended, middle school computer science course for deeper learning: A design-based research approach. Proceedings of the 12th International Conference on Learning Science. V2. Singapore. *Retrieved* from www.isls.org/icls/2016/docs/ICLS2016_Volume_2.pdf

Heller, R., Wolfe, R.E., Steinberg, A. (2017). Rethinking readiness *deeper learning for college, work, and life*. Cambridge: Harvard University Press.

Hewlett Foundation (2017). For teachers, deeper learning is about letting go. *Retrieved* from https://www.hewlett.org/strategy/deeper-learning/

Koh, J.H.L., Woo, H.L., & Lim, W.Y. (2013). Understanding the relationship between Singapore pre-service teachers' ICT course experiences and technological pedagogical content knowledge (TPACK) through ICT course evaluation. Educational Assessment, Evaluation and Accountability, 25, 321-229.

Lynch, S.M. (2013). Statistical Approaches for Nominal Data: Chi-Square Tests. In *Using Statistics in Social Research* (pp. 107-116). New York, NY: Springer.

Marshall, J.C. & Alston, D.M. (2014) Effective, Sustained Inquiry-Based Instruction Promotes Higher Science Proficiency Among All Groups: A 5-Year Analysis. J Science Teacher Education 25, p 807. https://doi.org/10.1007/s10972-014-9401-4

Mishra P. & Koehler M.J. (2006) Technological pedagogical content knowledge: a framework for teacher knowledge. Teachers College Record. 108(6):1017–1054

Pang, J. & Good, R. (2000). A Review of the Integration of Science and Mathematics: Implications for Further Research. *School Science and Mathematics, 100*(2), 73-82.

Papert, S. & Harel, I. (1991). Situating Constructionism. In S. Papert & I. Harel (Eds.), *Constructionism: research reports and essays*. New York: Ablex Publishing Corporation.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173-185.

Weiss, I.R., Pasley, J.D., Smith, P.S., Banilower, E.R., & Heck, D.J. (2003). Looking inside the classroom: A study of K-12 mathematics and science education in the United States. *Retrieved from* http://www.horizon-research.com/insidetheclassroom/

Repenning, A., Webb, D.C., Koh, K.H., Nickerson, H., Miller, S.B., Brand, C., Her Many Horses, I., Basawapatna, A., Gluck, F., Grover, R., & Gutierrez, K. (2015). Scalable game design: A strategy to bring systemic computer science education to schools through game design and simulation creation. *Transactions on Computing Education* (15), 2, New York, ACM DOI=http://dx.doi.org/10.1145/2700517

Yadav, A., Hong, H., & Stephenson, C. (2016), Computational thinking for all: Pedagogical approaches to embedding 21st Century Problem Solving in K-12 classrooms, TechTrends, (60), p. 565. https://doi.org/10.1007/s11528-016-0087-7

Nogera, P. (2017). Equal Opportunity for Deeper Learning. *Retrieved from* https://edpolicy.stanford.edu/publications/pubs/1383

Appendix A

Table A1. Means and Standard Deviations of Student Pre Assessments for the Final Analytic Sample, by Experimental Condition

Achievement Measure	Comparison	Treatment	Difference[a]
LbyM Science Assessment			
Mean	40.77	40.10	0.67
Standard deviation	20.59	16.26	--
N	52	98	--
LbyM Math Assessment			
Mean	46.15	43.75	2.40
Standard deviation	23.56	24.03	--
N	52	98	--

a. Computed based on valid (non-missing) data. Student level descriptive analysis without taking consideration of clustering effects (students were nested with classrooms).

Table A2. Means and Standard Deviations of Student Pre Attitude Measures for the Final Analytic Sample, by Experimental Condition

Attitude Measure	Comparison	Treatment	Difference[b]
Student Attitude towards Math			
Mean	3.38	3.35	.03
Standard deviation	1.00	.78	--
N	26	75	--
Student Attitude towards Science			
Mean	3.48	3.14	.34
Standard deviation	.74	.57	--
N	24	75	--
Student Attitude towards Technology and Engineering			
Mean	3.32	3.49	-.17
Standard deviation	.69	.69	--
N	26	75	--
Student Attitude towards 21st Century Skills			
Mean	4.21	3.91	.30
Standard deviation	.53	.52	--
N	36	74	--

b. Computed based on valid (non-missing) data. Student level descriptive analysis without taking consideration of clustering effects (students were nested with classrooms).

Table A3. Additional Regression Analysis of Predictors of Mathematics End-line Scores (percent correct) for the Final Analytic Sample (N = 150)

Predictor variables	GLM Model 1	HLM Model 1	HLM Model 2	HLM Model 3
Baseline math percent correct	0.60**	0.60**	0.52**	0.48**
Treatment	2.74	2.74	4.52	3.45
School-level demographics				
Percent ELs	--	--	−1.40	−0.96
Percent free/reduced lunch eligible	--	--	0.61	0.39
Percent minority ethnicity (non-white)	--	--	0.58	0.45
Cluster (base: Block 0)				
Block 1	--	--	--	5.11
Block 2	--	--	--	−8.23
Block 3	--	--	--	−2.58
Block 4	--	--	--	−1.41
Block 5	--	--	--	−5.03

[+] Significant at α = 0.10
* Significant at α = 0.05
** Significant at α = 0.01

Table A4. Additional Regression Analysis of Predictors of Science End-line Scores (Percent Correct) for the Final Analytic Sample (N = 150)

Predictor variables	GLM Model 1	HLM Model 1	HLM Model 2	HLM Model 3
Baseline science percent correct	0.68**	0.6**	0.61**	0.58**
Treatment	6.34*	5.96	7.87**	5.74
School-level demographics				
Percent ELs	--	--	−3.88**	−3.22
Percent free/reduced lunch eligible	--	--	0.84*	0.47
Percent minority ethnicity (non-white)	--	--	2.07*	1.57
Cluster (base: Block 0)				
Block 1	--	--	--	1.17
Block 2	--	--	--	−3.50
Block 3	--	--	--	1.29
Block 4	--	--	--	−2.02
Block 5	--	--	--	−7.82

[+] Significant at α = 0.10
* Significant at α = 0.05
** Significant at α = 0.01

Table B1. Participants' Demographic Information for the Sample with 8th Graders, by Experimental Condition

	Comparison		Treatment		
	Number	**Percent[c]**	**Number**	**Percent**	**p-value[d]**
Grade level					0.06
8th grade	7	11.86%	0	0%	--
9th grade	24	40.58%	87	88.78%	--
10th grade	12	20.34%	8	8.16%	--
11th grade	11	18.64%	0	0&	--
12th grade	5	8.47%	3	30.6%	--

c. Computed based on valid (non-missing) data. Components may not sum to 100 because of rounding.

d. A test for equality of proportion between treatment and comparison students was conducted, and the corrected p-value, accounting for clustering effects (students were nested with schools), was reported here.

Table B2. Baseline Equivalence of Student Pre Assessments for the Sample with 8th Graders, by Experimental Condition

Achievement Measure	Comparison	Treatment	Difference[e]	p-value	Effect Size[f]
LbyM Science Assessment					
Mean	40.82	40.88	-0.06	0.99	
Standard deviation	20.04	16.26	--	--	
N	59	98	--	--	
LbyM Math Assessment					
Mean	45.46	43.49	1.97	0.81	
Standard deviation	23.36	24.03	--	--	
N	59	98	--	--	

e. Computed based on valid (non-missing) data. A multi-level regression model that accounted for clustering effects (students were nested with classrooms) was used to test whether students' math and science pretests at baseline were equivalent between treatment and comparison groups.

f. Effect size was calculated by dividing treatment and comparison difference by the comparison group standard deviation of the pre measure variable.

Table B3. Baseline Equivalence for Student Pre Attitude Measures for the Sample with 8th Graders, by Experimental Condition

Attitude Measure	Comparison	Treatment	Difference[g]	*p*-value	Effect Size[h]
Student Attitude towards Math					
Mean	3.48	3.35	.13	.45	.14
Standard deviation	.96	.78	--	--	
N	33	75	--	--	
Student Attitude towards Science					
Mean	3.57	3.14	.43	.001**	.61
Standard deviation	.71	.57	--	--	
N	31	75	--	--	
Student Attitude towards Technology and Engineering					
Mean	3.34	3.49	-.14	.31	.21
Standard deviation	.66	.69	--	--	
N	33	75	--	--	
Student Attitude towards 21st Century Skills					
Mean	4.17	3.91	.26	.052	.46
Standard deviation	.57	.52	--	--	
N	33	74	--	--	

g. Computed based on valid (non-missing) data. A multi-level regression model that accounted for clustering effects (students were nested with classrooms) was used to test whether students' attitudes towards math, science, technology and engineering, and 21st century skills pretests at baseline were equivalent between treatment and comparison groups.

h. Effect size was calculated by dividing treatment and comparison difference by the comparison group standard deviation of the pre measure variable.

Table B4. Student Math and Science Outcomes for the Sample with 8th Graders (N=157)

Outcome Measure	Covariate-Adjusted Mean		Covariate-Adjusted Mean Difference	*p*-Value	95% Confidence Interval	Effect Size[i]
	Treatment (N = 98)	Comparison (N = 59)				
postMATH	49.89 (2.09)	48.07 (2.76)	1.82 (3.61)	0.61	-5.25 − 8.90	0.07
postSCI	48.87 (2.15)	43.81 (2.59)	5.06 (3.52)	0.15	-1.84 − 11.95	0.22

i. Effect size was calculated by dividing impact estimate by the comparison group standard deviation of the outcome variable (Math SD = 24.79; Science SD = 22.50).

Table B5. Additional Regression Analysis of Predictors of Mathematics End-line Scores (Percent Correct) for the Sample, with 8th Graders (N = 157)

Predictor variables	GLM Model 1	HLM Model 1	HLM Model 2	HLM Model 3
Baseline math percent correct	0.58**	0.58**	0.52**	0.47**
Treatment	1.41	1.41	1.82	0.72
School-level demographics				
Percent ELs	--	--	-1.09	-0.71
Percent free/reduced lunch eligible	--	--	0.38	0.07
Percent minority ethnicity (non-white)	--	--	0.35	0.18
Cluster (base: Block 0)				
Block 1	--	--	--	2.41
Block 2	--	--	--	-12.01
Block 3	--	--	--	-4.10
Block 4	--	--	--	-1.76
Block 5	--	--	--	-8.87

[+] Significant at α = 0.10
* Significant at α = 0.05
** Significant at α = 0.01

Table B6. Additional Regression Analysis of Predictors of Science End-line Scores (Percent Correct) for the Sample, with 8th Graders (N = 157)

Predictor variables	GLM Model 1	HLM Model 1	HLM Model 2	HLM Model 3
Baseline math percent correct	0.68**	0.65**	0.62**	0.58**
Treatment	4.80[+]	4.65	5.06	2.19
School-level demographics				
Percent ELs	--	--	−2.99[+]	−2.86
Percent free/reduced lunch eligible	--	--	0.58	0.04
Percent minority ethnicity (non-white)	--	--	1.53	1.19
Cluster (base: Block 0)				
Block 1	--	--	--	−2.27
Block 2	--	--	--	−8.00
Block 3	--	--	--	−0.45
Block 4	--	--	--	−2.35
Block 5	--	--	--	−12.54

[+] Significant at α = 0.10
* Significant at α = 0.05
** Significant at α = 0.01

Table B7. Student Attitudes Toward Math, Science, Engineering, and 21st Century Skills, with 8th Graders (N = 108)

Scale	Covariate-Adjusted Mean		Covariate-Adjusted Mean Difference[a]	p-Value	95% Confidence Interval	Effect Size[j]
	Treatment (N = 75)	Comparison (N = 33)				
Math	3.29 (0.06)	3.31 (0.11)	-0.03 (0.14)	0.85	-0.30 – 0.25	-0.03
Science	3.16 (0.07)	3.16 (0.13)	0.00 (0.16)	1.00	-0.32 – 0.32	0.00
Engineering	3.28 (0.08)	3.20 (0.14)	0.08 (0.18)	0.64	-0.27 – 0.44	0.11
21st Century	3.93 (0.07)	3.93 (0.11)	0.00 (0.15)	0.98	-0.29 – 0.29	0.00

j. Effect size was calculated by dividing impact estimates by the comparison group standard deviation of the outcome variable (Math SD = 1.00; Science SD = 0.77; Engineering SD = 0.76; 21st Century SD = 0.55)

Table B8. Chi-squared Test of Student Interest in STEM Careers for the Sample with 8th Graders (N = 108)

Field	Pre-Intervention					Post Intervention				
	Intervention		Comparison		P-value	Intervention		Comparison		P-value
	N	%	N	%		N	%	N	%	
Physics					0.62					0.70
Not interested	47	62.67	19	57.58		55	73.33	23	69.70	
Interested	28	37.33	14	42.42		20	26.67	10	30.30	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Env. Science					0.57					0.05*
Not interested	48	64.00	23	69.70		59	78.67	20	60.61	
Interested	27	36.00	10	30.30		16	21.33	13	39.39	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Biology					0.09+					0.33
Not interested	45	60.00	14	42.42		44	58.67	16	48.48	
Interested	30	40.00	19	57.58		31	41.33	17	51.52	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Vet. Medicine					0.60					0.69
Not interested	45	60.00	18	54.55		47	62.67	22	66.67	
Interested	30	40.00	15	45.45		28	37.33	11	33.33	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Mathematics					0.60					0.18
Not interested	46	61.33	22	66.67		53	70.67	19	57.58	
Interested	29	38.67	11	33.33		22	29.33	14	42.42	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Medicine					0.43					0.07+
Not interested	38	50.67	14	42.42		37	49.33	10	30.30	
Interested	37	49.33	19	57.58		38	50.67	23	69.70	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Earth Science					0.78					0.40
Not interested	52	69.33	22	66.67		58	77.33	23	69.70	
Interested	23	30.67	11	33.33		17	22.67	10	30.30	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Comp. Science					0.88					0.07+
Not interested	42	56.00	19	57.58		50	66.67	16	48.48	
Interested	33	44.00	14	42.42		25	33.33	17	51.52	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Med. Science					0.16					0.12
Not interested	45	60.00	15	45.45		44	58.67	14	42.42	
Interested	30	40.00	18	54.55		31	41.33	19	57.58	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Chemistry					0.02*					0.14
Not interested	51	68.00	15	45.45		52	69.33	18	54.55	
Interested	24	32.00	18	54.55		23	30.67	15	45.45	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Energy					0.21					0.81
Not interested	45	60.00	24	72.73		54	72.00	23	69.70	
Interested	30	40.00	9	27.27		21	28.00	10	30.30	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	
Engineering					0.00**					0.06+
Not interested	31	41.33	25	75.76		40	53.33	24	72.73	
Interested	44	58.67	8	24.24		35	46.67	9	27.27	
Total	*75*	*100.00*	*33*	*100.00*		*75*	*100.00*	*33*	*100.00*	

+ Significant at α = 0.10
* Significant at α = 0.05
** Significant at α = 0.01

Appendix B

Table 1. 2016-2017 Learning by Making Fidelity Matrix
Key Component 1: Learning by Making Professional Development

Indicators	Definition	Unit of implementation	Data Source(s)	Score for levels of implementation at unit level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure
Key Component 1: Professional Development						
1a A five-day summer institute provided once per year	Teachers attend 5-day summer institutes	Teachers	Learning by Making records: attendance rosters	0-3 points 0=None (0 days) 1=Low (1-2 days) 2=Moderate (3-4 days) 3=High (5 days)		Teachers delivering the intervention during the 2016-17 and 2017-18 school years.
1b Five one-day follow-up PD sessions provided throughout the year	Teachers attend 5 follow-up PD sessions throughout the year	Teachers	Learning by Making records: attendance rosters	0-3 points 0=None (0 sessions) 1=Low (1-2 sessions) 2=Moderate (3-4 sessions) 3=High (5 sessions)		Teachers delivering the intervention during the 2016-17 and 2017-18 school years.
1c Online platform provided for teacher resources sharing	Teachers share resources in the online platform	Teachers	Learning by Making online platform records	0-1 points 0=None (No) 1=High (Yes)		Teachers delivering the intervention during the 2016-17 and 2017-18 school years.
All indicators				Sum across all 3 indicators (1a, 1b, and 1c) 1=Low (score of 0-2) 2=Moderate (score of 3-5) 3=High (score of 6-7)	% of 12 teachers with 3= high implementation 1=Low (0-50%) 2=Moderate (51%-75%) 3=High (At least 76%)	

Table 1. 2016-2017 Learning by Making Fidelity Matrix (continue)
Key Component 2: Learning by Making Curriculum Development and Implementation

Indicators	Definition	Unit of implementation	Data Source(s)	Score for levels of implementation at unit implementation at sample level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure
Key Component 2: Learning by Making Curriculum Development and Implementation						
2a Learning by Making curriculum developed	SSU develops two hardware and one software curriculum modules for each semester for two years	Project	Learning by Making records	1-3 points 1=Low (1-2 modules) 2=Moderate (3-5 modules) 3=High (6 modules)	1-3 points 1=Low (1-2 modules) 2=Moderate (3-5 modules) 3=High (6 modules)	Project director participating during the 2016-17 school year
2b Learning by Making curriculum provided as an elective course	Learning by Making curriculum is offered in each school	School	Learning by Making records	0-1 points 0=None (No) 1=High (Yes)	% of 6 schools with 1=High implementation 1=Low (0-50%) 2=Moderate (51%-82%) 3=High (at least 83%)	Schools participating during the 2016-17 and 2017-18 school years
2c Learning by Making curriculum implemented in classrooms	Teachers teach Learning by Making curriculum	Teachers	Learning by Making records	0-3 points 0=None (0 modules) 1=Low (1-2 modules) 2=Moderate (3-5 modules) 3=High (6 modules)	% of 12 teachers with 3=High implementation 1=Low (0-50%) 2=Moderate (51%-75%) 3=High (at least 76%)	Teachers participating during the 2016-17 and 2017-18 school years
All indicators				sum across all 3 indicators (2a, 2b, and 2c) 1=Low (score of 1-2) 2=Moderate (score of 3-5) 3=High (score of 6-9)	% of 12 teachers with 3= high implementation 1=Low (0-50%) 2=Moderate (51%-75%) 3=High (At least 76%)	

Table 2. Fidelity of Implementation Results for Implementation Year 3 (2016-2017) and Implementation Year 4 (2017-2018)

Table 2a. Findings from Learning by Making Study of Implementation: IMPLEMENTATION YEAR 3 2016-2017

Intervention Component	Implementation measure (total number of measurable indicators representing each component etc)	Sample Size at the Sample Level (# of schools, districts, etc)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Planned Intervention Activities [i.e., key components]							
Professional Development	3 measures	12 schools		Sum across all 3 indicators 1=Low (score of 0-2) 2=Moderate (score of 3-5) 3=High (score of 6-7)	76 – 100% of teachers have "high implementation"	91.6% with "high implementation"	Yes
Learning by Making Curriculum Implementation	3 measures	12 teachers from 6 schools	No longer required	Sum across all 3 indicators (2a, 2b, and 2c) 1=Low (score of 1-2) 2=Moderate (score of 3-5) 3=High (score of 6-9)	76 – 100% teachers have "high implementation"	100% with "high implementation"	Yes

Table 2b. Findings from Learning by Making Study of Implementation: IMPLEMENTATION YEAR 4 2017-2018

Intervention Component	Implementation measure (total number of measurable indicators representing each component)	Sample Size at the Sample Level (# of schools, districts, etc)	Representativeness of sample: Measured on All (A), Some (S), or None (N) of the units representing the intervention group in the impact analyses	Component Level Threshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Professional Development	3 measures	8 teachers* from 6 schools		Sum across all 3 indicators 1=Low (score of 0-2) 2=Moderate (score of 3-5) 3=High (score of 6-7)	76 – 100% of teachers have "high implementation"	100% with "high implementation"	Yes
Planned Intervention Activities [i.e., key components]							
Learning by Making Curriculum Implementation	2 measures	8 teachers* from 6 schools	No longer required	Sum across all 3 indicators (2a, 2b, and 2c) 1=Low (score of 1-2) 2=Moderate (score of 3-5) 3=High (score of 6-9)	76 – 100% of teachers have "high implementation"	100% with "high implementation"	Yes

Note: The component level fidelity scores for the entire sample are based on the sample size of 8 teachers from 6 schools. In year 4 of the study, 2017-18, the study is still implemented in 6 schools. However, the instructional structure is changed from co-teaching by two teachers per school to either co-teaching by two teachers or teaching by one teacher only. The instructional structure is determined by the schools. Therefore, only 8 teachers from 6 schools participate in year 4 implementation.