CPR2 Year Three Report

Summer 2022

Prepared by:
SRI International
Andrea Beesley
Jared Boyce
Patrik Lundh
Carol Tate
Mindy Hsiao
Elise Levin-Guracar

SRI Education™
A DIVISION OF SRI INTERNATIONAL

SRI International • 333 Ravenswood Avenue • Menlo Park, CA 94025-3493 • 650.859.2000 • www.sri.com

This material is based upon work supported by a grant from the National Science Foundation (#1933678). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Contents

Executive Summary

This study set out to answer research questions about the design of CPR2, the implementation of CPR2 in middle school classrooms, and the potential of CPR2 to improve both students' and teachers' abilities to engage in mathematical generalization. Toward these ends, we conducted a 3-year study spanning 2019-22, divided into three phases:

1. A co-design year with the designers of CPR2 and mentor teachers with experience delivering CPR2 content in their classrooms, aimed at refining CPR2 instructional materials and training,
2. A pilot year to test and further refine CPR2 instructional materials and training, which was particularly important given the unexpected COVID-19 pandemic, and
3. An initial efficacy year to measure the extent to which CPR2 activities improved teachers' and students' mathematical generalization skills and attitudes toward programming.

In the co-design year, we found the CPR2 designers from the University of North Alabama (UNA) and the mentor teachers engaged in a focused discussion about a shared problem of practice—teaching generalization in mathematics—and reflected on how CPR2 could help address it. They shared ideas about integrating CPR2 into the classroom context, anticipated student reactions to the activities, examined related state standards, and suggested refinements to the program. The CPR2 designers revised the materials before the pilot year.

We found that CPR2 was largely implemented as intended in the piloting year. The CPR2 professional development activities provided teachers with the programming skills and preparation needed to deliver CPR2 lessons, teachers consistently implemented three of the four steps of the CPR2 Instructional Model (essential mathematical concepts, writing mini-programs, writing general expressions), and students largely completed the required activities. We seldom observed teachers leading students in the fourth step of the CPR2 Instructional Model, making conjectures and writing convincing arguments. Observers found CPR2 instruction to be mostly teacher-led, with little time for the exploration and discussion that the instructional model calls for.

In Year 3, observation data suggested that students without prior programming experience were able to successfully write simple Python programs as part of the CPR2 lessons. In interviews, teachers reported students improved their programming skills as a result of CPR2. Unfortunately, extensive attrition across the school year (over 80% of control students and 90% of treatment students did not complete outcome data collection) prevented the impact analysis from producing reliable findings regarding impact on either teachers' or students' mathematics generalization skills.

This study was conducted during the COVID-19 pandemic, which likely contributed to attrition, general teacher stress, and limitations on class time devoted to CPR2. Further study will be needed to determine what conditions and lesson dosage would be required for teachers to adopt the intended pedagogy of CPR2 and realize its intended benefits. Additionally, more sensitive measures of generalization ability would strengthen a second impact study.

Teacher Sample Information

This chapter describes the teacher sample for this study. First, we describe the sample size from initial randomization in May 2021, the sample of teachers retained into the fall, and finally the sample of teachers retained to the end of the school year in June 2022. Second, we describe the results of the Teacher Background Survey to better describe our teacher sample.

Study Sample Tracking

In spring 2021, UNA recruited 49 teachers to participate in the 2021-22 pilot efficacy year for implementing CPR2 in middle school classrooms. These teachers were recruited across seven states (see Appendix A). After randomization, there were 25 treatment teachers assigned to participate in the CPR2 Summer Institute and implement CPR2 lessons in their classrooms and 24 control teachers who would receive delayed CPR2 training after the study.

These 49 teachers needed to complete the following activities to be retained into the fall: (1) complete the Teacher Background Survey, (2) complete the Learning Mathematics for Teaching (LMT) pretest, (3) administer the student assessment and student survey pretests in at least one of their classrooms, and (4) participate in the CPR2 Summer Institute (treatment only). Teachers who did not complete these activities, whether due to scheduling issues, lack of district approval or other reasons, were considered to no longer be participating in (attritted from) the study. Additionally, teachers left the study due to being promoted into non-teaching roles or being reassigned to subjects or grade levels outside the scope of the study. A total of 24 teachers (14 treatment teachers, 10 control teachers) were retained into the fall.

These 24 teachers needed to complete the following activities to be retained for our spring impact analyses: (1) complete the spring Implementation Questionnaire (to assess treatment-control contrast), (2) complete the LMT posttest (for teacher impacts), and (3) administer the student assessment and student survey (for student impacts). Teachers who did not complete these activities were attritted from the study. Additionally, we removed teachers' data from the student impact analyses when we were unable to match students' pretest and posttest data, which occurred when teachers changed classes between fall and spring semesters. We included 20 teachers (13 treatment teachers, seven control teachers) of the 24 retained teachers in our teacher impact analysis. We included only five teachers (three treatment teachers, two control teachers) in the student impact analyses.

Table 1. Teacher sample after randomization

	Treatment	Control
Randomization, May 2021	25	24
Retained into the fall	14	10
Retained for teacher impacts	13	7
Retained for student impacts	3	2

Teacher Background Survey

Background: SRI Education conducted a Teacher Background Survey with treatment and control teachers before the 2021 Summer Institute. The purpose of the survey was to collect information about the participants' teaching backgrounds and previous professional development experiences related to math, computer science, and generalization.

Design: The Teacher Background Survey was the same survey as 2020's Pre-Summer Institute Questionnaire given to the 10 pilot teachers. The SRI, UNA, and Horizon Research, Inc. (HRI) teams reviewed the Teacher Background Survey prior to summer 2021 and did not see any need to adjust the questions asked.

Data Collection & Analysis: SRI staff emailed Qualtrics links for the Teacher Background Survey to treatment and control participants after they completed the LMT and before the 2021 Summer Institute. The UNA team and research associates at SRI provided email reminders to complete the survey 1 week after sending the survey to participants and completed two other rounds of follow up with participants. By July 19, 2021, 37 of the 40 non-attritted teachers in the sample completed the background survey, leading to a 92.5% response rate.

Findings: In this section, we highlight the overall findings from the Teacher Background Survey. Please see Appendix B for specific values from the survey. Findings are based on the 37 responses (20 treatment, 17 control) we received from teachers.

1. **On average, the teachers have 11 years of teaching experience.** Respondents in both treatment and control conditions have similar years of teaching experience. More teachers have experience teaching math than computer science or programming. On average, teachers have 10 years teaching math and less than 1 year teaching computer science. Teachers on average have taught at their current school for about 6 years.
2. **At least half of the respondents have a master's degree or higher.** Teachers in treatment and control conditions have similar educational attainment. About 40% of teachers who responded to the survey described their education level as a bachelor's degree or some courses past a bachelor's. About 50% of teachers who responded to the survey have a master's degree and about 5% have a doctorate.
3. **Most of the respondents have a degree in education, and about half have a degree in mathematics.** There were 12 teachers with both mathematics and education degrees. There is a slight difference in treatment and control teachers' fields of

study. More treatment teachers received a math degree than control teachers; however, control teachers had a wider variety of degrees including those in statistics and computer science.

4. **Of the 37 respondents, 19 teachers received math professional development and three received computer science professional development in the last 12 months.** For those who received math professional development, respondents reported the most common formats were a professional development/workshop or online course/webinar. On average, respondents who received professional development spent 21 hours in math professional development in the last 12 months. For those who received computer science professional development, respondents reported the most common formats were a professional development/workshop or online course/webinar.
5. **Half of the respondents said they signed up for CPR2 to learn how to incorporate math with computer science.** Other specific reasons included a desire to learn: ways to increase student engagement in computer science and/or math how to improve their own mathematical pedogeological practices, how to program, and methods for teaching generalization. Many respondents also provided broad answers such as a desire to find better ways to equip their students and learn new ways to teach in general.
6. **Most treatment teachers anticipate using individual Chromebooks to teach CPR2 lessons.** There were 17 treatment teachers who planned to have students use Chromebooks, six others intended to use laptops, five proposed to use tablets, and two responded that students would use desktops. There were 11 treatment teachers with either a 1:1 technology program at the school or had computers in their classroom readily available for each student to use. Four treatment teachers had a shared computer cart, and only one teacher was unsure of the technology access in their classroom or school.

We re-analyzed the Teacher Background Survey data at the end of the 2022–23 school year specifically focused on the final analytic sample of 20 teachers (13 treatment, 7 control) who completed at least one outcome data collection activity. Overwhelmingly, we found that the backgrounds of the final analytic sample mirrored those of the initial recruitment sample. Any differences were minor. For example, the final analytic sample on average had 1 year more of teaching experience in general (12 years vs. 11 years) and 1 year more of mathematics teaching experience in particular (11 years vs. 10 years) than the initial recruitment sample.

2021 CPR2 Summer Institute Findings

A complete writeup of our 2021 CPR2 Summer Institute findings can be found in the Year 2 Research Performance Progress Report (RPPR). We provide a summary of the key findings here as framing for the research completed over the past year.

Summer Institute Observations

Following last year's Summer Institute model, the 2021 CPR2 Summer Institute was held online via Zoom. We observed the sessions virtually. There were usually three to four observers per day over the course of eight professional development days from June 1 through June 29. After each session, one observer stayed to observe and take notes on the mentor and UNA leader debrief and planning session, which usually lasted approximately 30 minutes. We were unable to observe individual mentor and participant sessions, in which mentors and participants worked together on prework assignments and planning for their practice teaching assignments. The findings below summarize the themes and variations we saw across all Summer Institute observations.

1. **As in 2020, sessions were still consistently implemented as designed with respect to content, timing, and roles**. Mentor teachers continued to be instrumental to the participants' experiences through communication, support, and learning that occurred during and outside of Institute sessions.
2. **Participants had more opportunity to practice teaching than in 2020.** Most teachers were able to deliver lessons to their "students" in breakout sessions, using questioning techniques and providing feedback. Some teachers were able to receive descriptive feedback from their colleagues about their lessons.
3. **Participants were more likely to be actively engaged during breakout sessions, which constituted about 45 minutes of each session.** By "actively engaged," we mean that participants were problem-solving or otherwise acting on their own initiative, rather than primarily watching and listening. This was an increase from about 30 minutes of each session in the prior year.
4. **Activity leaders and participants primarily used a teacher-driven instructional style.** Activity leaders emphasized that participants should facilitate student-focused discussions, student-driven tinkering, and exploration, but there was limited opportunity for participants to practice how they could do this in their classrooms.
5. **Activity leaders provided evaluative feedback and some descriptive feedback.** Participants mostly received evaluative feedback (praise, in this case) after their practice teaching. Less feedback was descriptive, pointing out specific features of instruction.
6. **The Summer Institute did not focus on assessing student learning.** Participants generally did not practice interpreting CPR2 work to understand what students were learning, discussing formative or summative assessment of student work, or measuring whether the CPR2 lessons had an impact on students' understanding of generalization.

Daily Participant Feedback Survey

The SRI team emailed Qualtrics links for the daily survey feedback to teachers during the closing activities for each of Days 1–7 of the Summer Institute 2021. Overall response rates were high with a minimum of 18 teachers (90+% of 20 treatment teachers) responding to the daily survey on all days other than Day 4, when 15 teachers (75%) responded, likely due to the different structure of the day and the lack of a formal closing routine. The findings below summarize the themes and variations we saw across all daily feedback surveys.

1. **Participants were overwhelmingly positive about the content and facilitation of the Summer Institute 2021.** Participants overwhelmingly responded with 5's and 6's on the 1-6 Likert-style questions (6 being the most positive option) on each of Days 1-7. These responses were supported by participants' responses to the open-ended questions, which included many appreciative comments about CPR2 content and the institute's instruction.
2. **Based on participant feedback, Summer Institute 2021 was significantly improved from Summer Institute 2020.** Participants in 2020 initially reported mixed reactions to the prework assignments in terms of both content and length. During 2021, respondents were consistently positive about the prework assignments throughout the entire institute. Participants in 2020 noted a lack of opportunity for them to practice teaching CPR2 content themselves during the institute. In 2021, participants appreciated that teaching practice time was included in the institute's program.
3. **Teachers reported being appropriately challenged by the material.** For each of Days 1-7, most teachers reported that the math and/or programming content was new to them and reported feeling challenged by the day's content. The reported challenges appeared to change over time, suggesting that teachers grew comfortable with content as the institute progressed; there were also fewer overall reported challenges over time, suggesting a general increased level of confidence.
4. **Teachers identified several opportunities for additional resources to be created to support teaching CPR2 in their classrooms.** The most common resource requests were to provide the following: a coding "cheat sheet" or one-pager with the most common code samples/programming terms, a proof "outline" or similar resource to support students in learning how to write a proof, and a compilation of participants' example proofs from the Summer Institute that they could refer to during the year.

Summative Participant Feedback Survey

The SRI team emailed Qualtrics links for the summative survey feedback to teachers during the closing activities for Day 8 of the Summer Institute 2021. The response rate was reasonable with

15 teachers (75% of 20 treatment teachers) responding to the summative survey. The findings below summarize the themes and variations we saw in the survey.

1. **Participants thought the Summer Institute 2021 had positive impacts on their knowledge and teaching.** Participants responded very positively about how the Summer Institute increased their knowledge of computer programming, knowledge of mathematical generalization, and confidence with teaching computer programming.
2. **Participants overwhelmingly reported feeling confident and prepared to teach CPR2 content in their classrooms.**
3. **Participants expected few challenges with implementing CPR2 in their classrooms.** Respondents reported the two greatest anticipated challenges were "Lack of time to implement the activities" and "Challenges with debugging code." Still, most teachers reported they did not anticipate any significant/major challenges.
4. **Participants overwhelmingly reported that the Summer Institute 2021 met or surpassed their expectations.** The expectations teachers reported having for the institute were well-aligned to CPR2 content, including learning mathematical generalization, programming and how to teach it, and how to bridge the gap between computer science and mathematical thinking.
5. **The most common request for additional support was for ongoing contact/access with CPR2 instructors, mentor teachers, and fellow participants.**

Teacher Focus Groups

The SRI researchers interviewed 14 of the 20 teachers who participated in the 2021 Summer Institute in five small groups via videoconference. These findings are summarized below.

- **For the most part, teachers held a common conception of mathematical generalization.** Teachers largely defined mathematical generalization in terms of the ability to see, communicate, and use patterns when engaged in math problem-solving activities.
- **Teachers said they felt that mathematical generalization is an important skill worthy of classroom time.** All the teachers we spoke with cited at least one benefit of developing generalization skills. A few pointed out that mathematical generalization is part of their math curriculum.
- **Teachers cited students' aversion to struggle and their lack of foundational knowledge as the primary challenges to developing mathematical generalization skills.** Teachers' own lack of deep mathematical knowledge was also seen by some as a hindrance to fostering mathematical generalization.

- **Teachers were positive in their descriptions of the Summer Institute.** Teachers described the Summer Institute as 'beneficial,' 'intense,' 'challenging,' 'engaging' and 'a constant barrage.' For the most part, they enjoyed the experience.
- **While the Summer Institute experience was intellectually invigorating for the teachers, some expressed concerns about translating CPR2 for their students.**
- **Teachers anticipated that proofs would be one of the biggest challenges, a concern that had also been voiced in 2020.**
- **About half the teachers said they left the Summer Institute prepared to teach CPR2. Others felt the pacing and online format limited their preparation.** One said that the pacing of the Summer Institute was much faster than what they would do with their students. Another teacher said that the online format of the Summer Institute made it hard to see how students would react. (Pre-COVID, in-person Summer Institutes had a practical component of delivering CPR2 instruction to students.)

Fall Implementation Findings

This chapter describes the implementation findings for fall 2021. First, we describe our findings from observing CPR2 lessons taught by teachers in the treatment condition. Second, we describe the results of our Implementation Questionnaire to confirm that treatment teachers implemented CPR2 in the fall and spring semesters as planned and control teachers did not teach similar content.

Fall CPR2 Classroom Lesson Observations

Background: The purpose of the fall lesson observations was to describe how teachers who attended the 2021 Summer Institute implemented CPR2. We analyzed in what ways they implemented the CPR2 lessons and supported student engagement and learning, and whether students participated in CPR2 lessons in ways that supported CPR2 learning objectives.

Design: We used the same observation protocol as in fall 2020 and spring 2021, which consisted of two parts: 1) time-stamped running notes to document activities, teacher and student talk, and notes about the learning environment and issues relevant to understanding the lesson; and 2) a debrief organized by descriptive categories aligned with the project's constructs table. The debrief categories were based on the CPR2 instructional model and on other aspects of instruction that we believe support the CPR2 instructional model, including facilitating rich classroom discussions that allow for student questions and reasoning, checking for student understanding, and addressing student misconceptions. Observers took running notes on individual lessons and then wrote summaries for each of the debrief categories.

Data Collection & Analysis: Due to COVID-19 restrictions, we conducted lesson observations virtually. Observers were able to see either the front of the classroom (usually the screen and the teacher, although at times we were unable to see the board/screen) or the teacher's desktop. Importantly, observers did not see students and often the audio quality significantly limited the student talk observers were able to hear. Our observation findings therefore do not fully capture aspects of student engagement, teacher-student interactions outside of front-of-class teacher-led activities, or peer interactions.

We observed 10 teachers implementing between one and five lessons for a total of 34 observed lessons. Most of the observed teachers taught the Intro to Python and What is Even? lessons. Some of the observed teachers did not teach the What is Odd? and What is Zero? lessons.

To analyze the data, we created a summary debrief for all observations for each teacher. One SRI researcher reviewed all debrief categories across all teacher summaries and described themes and/or variations for each debrief category (e.g., what kinds of questions did teachers ask students, or to what degree did teachers provide student opportunities to write general expressions to represent the mathematical relationships they discovered?). The findings below summarize the themes and variations we saw across all teacher observations.

Findings

Overall, with some exceptions, teachers provided students opportunities to program. Teachers generally seemed to feel confident in the CPR2 content. Teachers tended to follow the "script," with an emphasis on students following directions correctly. Teachers' questions and reasoning dominated instruction and whole-class discussions. There was little support for checking and adapting to student understanding, or for exploring mathematical concepts in depth or generating and discussing mathematical conjectures.

Presence of CPR2 Instructional Model: Most teachers provided opportunities for students to write mini-programs and to write general expressions. There was an emphasis on following instructions, with less opportunity for students to explore mathematical concepts through the programming or write and explore mathematical conjectures based on what they were learning through the programming. Teachers generally did not identify and explain generalization. One teacher did have students work on debugging programs as well as writing their own programs (to add another column of even numbers) from scratch. Four teachers provided time for students to program, but students copied code rather than exploring on their own. One teacher gave students time to write general expressions to represent even and odd numbers, and to discuss other ways to write expressions for even and odd numbers.

Teacher Capacity: Most of the teachers appeared to be confident in, and have a good grasp of, the CPR2 content. For example, one teacher knew the general expressions, knew what to ask when students were not sure about general expressions for even or odd; this teacher also guided them to 2n and 2n + -1 and knew how to debug student work. Another teacher taught lessons that were not rushed; students were engaged and participating and were all finished with each activity before moving on. Other teachers occasionally struggled. One teacher seemed to know the programming, but did not know the compiler very well, at one point asking the observer for help. Another teacher appeared confident when talking about content, while they frequently struggled with the debugging, such as indenting in the loop, or realizing that n<5 would not print a column of 1-5 but rather 1-4.

Instructional Practices: Most teachers did not clearly frame the lessons or connect activities to prior learning. Several teachers emphasized compliance and following teacher directions. Some teachers encouraged students to speak up and to ask questions. One teacher, who generally seemed more flexible and responsive to students, gave students ample time to explore their programs and come to an understanding of them. Some teachers asked questions and gave students time to think and answer. But overall, teachers rather than students did most or all of the cognitive work in the classrooms observed. This meant the teacher was primarily the one to ask questions, initiate discussions, verbalize their reasoning, and do the explaining, with varying degrees of student input. For example, one teacher drove the reasoning with some student input; this teacher did most of the problem solving, explaining, and reasoning. When debugging, the teacher did very little to clarify what they were doing and why. Instead, they simply debugged the program and moved onto the next student. Given the virtual observation format, it was difficult for observers to gauge student engagement. We did note that students appeared engaged or interested in about half the classes, and more passive and mostly following directions in the other half.

Whole-class Discussion and Teacher Questions: The common instructional approach of teachers owning the reasoning and students following directions was reflected during whole class discussion as well, which can otherwise be an opportunity to surface student questions, struggles, and ideas. During whole class discussions, about half the teachers tried to respond to and build on students' ideas. The discussions were primarily led and informed by the teachers' questions. One teacher built on student answers in the style of "What else are we missing?" The other half of classes were dominated by teacher reasoning. For example, one teacher drove the conversation toward answers they were already looking for. Students answered correctly, and the teacher explained the reasoning rather than having the students explain. In all but two classrooms, teachers primarily asked fill-in-the-blank and funneling questions (those that point to one specific answer) rather than open-ended questions to prompt students' own reasoning. In a couple of cases, observers noted that teachers would quickly answer their own questions without giving students enough time to think and respond. Two teachers who relied primarily on fill-in-the-blank questions also included more open-ended ones as well. One of them had students explain what they noticed when they printed arithmetic operators. The teacher did the same for equal to/less than/etc. This teacher also had students explain why they thought each operator did what it did in the code.

Teacher Support for Student Understanding: Teachers' practices for checking student understanding primarily took the form of call-and-response or walking the room and checking student work. Across all classrooms (except one in which student questions were not audible), students either asked no questions or only asked questions about procedure. No students were observed asking conceptual questions. Teachers provided little descriptive feedback to students. There were some instances of praise and encouragement. One teacher, while not providing feedback *per se*, did continue to build on students' responses, asking follow-up questions to explore their reasoning. Given that teachers did not do much checking for understanding, there were few adaptations to instruction in response to student thinking. Most teachers appeared focused on going through the slides with little or no change.

Fall Implementation Questionnaire

Background: The purposes of the Implementation Questionnaire were to collect information about the instructional practices in both groups and to understand CPR2 implementation in the treatment group.

Design: In fall 2021, teachers completed a short Implementation Questionnaire covering classroom practices related to CPR2 activities. Both groups answered questions about their teaching assignments, mathematical generalization instruction, and Python programming in the classroom. The treatment teachers answered additional questions about which CPR2 lessons they taught, any modifications to the lessons, challenges they faced, and how students responded to the lessons.

Data Collection: All treatment (14) and control (10) teachers retained from the summer into the fall completed the Implementation Questionnaire.

Findings: We did not see evidence of control teachers engaging in programming instruction or mathematical generalization instruction similar to CPR2. The treatment group overall implemented CPR2 content as intended for fall 2021, which was further confirmed by our fall CPR2 lesson observations.

Specific findings for the treatment group include:

- Eleven treatment teachers taught all portions of the fall CPR2 lessons; three skipped some portions.
- One treatment teacher said they taught generalization (generalizing sequences of various types) in their implementation class during the fall semester outside of the CPR2 lessons.
- No teachers reported teaching proof writing to their students in the implementation class during the fall semester outside of the CPR2 lessons.
- Four treatment teachers said they used Python with their students in the implementation class during the fall semester outside of the CPR2 lesson.

- Two teachers reported providing “starter code” for students to copy down: “I put the code in Google Classroom so that some students could copy and paste in the code if they had difficulty typing it.”
- One teacher reported needing to provide upfront support around using the computer: “I had to get them up to par with the computer because many of my students are not very knowledgeable of computers.”
- Teachers did not report making other substantial adaptations. Minor adaptations included using task cards for the questions, using alternative compilers, and putting questions into Google Classroom or Schoology.
- Four teachers reported no challenges, three teachers reported low engagement as a challenge, and three teachers reported COVID-19-related challenges (low attendance/ quarantined students).
- Most teachers reported their students were very engaged. One computer science teacher reported low engagement and attributed it to the fact that students “are used to developing programs for games and apps.” Two reported mixed levels of engagement.
- One treatment teacher expressed it was difficult to simultaneously teach programming and the CPR2 mathematical generalization content: “Students may respond better with more background in practicing programming in Python. Several students seemed to get so hung up on the programming that they couldn't see the math/generalization portion of the lessons.”

Specific findings for the control group include:

- Four of the 10 control teachers reported teaching mathematical generalization or pattern finding to their students during the fall semester.
    1. “We introduced linear relationships, specifically focusing on nonproportional relationships, which is an essential standard for 8th grade. We start by reviewing proportional relationships, which they learned in 7th grade, so we review that first and then build up to nonproportional relationships. So we focus on real-world problems and finding patterns that way with money, bank accounts, cell phone plans, and more.”
    2. “Finding patterns and writing equations for tile diagrams.”
    3. “Decimal Operations such as if-then statements…Also in ratio tables. In statistics, the shape of the data and what that tells you.”
    4. “Ratios, rates”
- Four of the 10 control teachers reported teaching proof-writing or communicating with mathematical language to their students in the fall semester.
    1. "We have formalized much math language including center and spread. We discussed equivalence in detail, as well as language associated with the mathematical

properties. Students also invented their own methods of finding the target number in a data set (which we later formalized into mean median and mode) and precision number in a data set (which we are currently formalizing into IQR and MAD)."

2. "Explain how to graph inequalities"
3. "Student had to justify the Pythagorean Theorem as well as justify steps to solve equations"
4. "When learning a new math concept, we first define new mathematical vocabulary and make sure to use it during instruction. Every lesson, we define and utilize mathematical language."

- None of the control teachers reported teaching programming.

Spring Implementation Findings

This chapter describes the implementation findings for spring 2022. First, we describe our findings from observing CPR2 lessons taught by teachers in the treatment condition. Second, we describe the results of our Implementation Questionnaire to confirm that treatment teachers implemented CPR2 in the fall semester as planned and control teachers did not teach similar content. Third, we describe the results of our spring interviews with treatment teachers to understand their experiences teaching CPR2 content in their classrooms across 2021–22.

Spring CPR2 Classroom Lesson Observations

Background: The purpose of the spring lesson observations was to describe how teachers who attended the 2021 Summer Institute implemented CPR2 during the 2022 spring semester. We analyzed how they implemented the CPR2 lessons and supported student engagement and learning.

Design: We modified the observation protocol used during the last three data collection rounds (fall 2020, spring 2021, fall 2021) that originally consisted of two parts: 1) time-stamped running notes to document activities, teacher and student talk, and notes about the learning environment and issues relevant to understanding the lesson; and 2) a debrief organized by descriptive categories aligned with the project's constructs table. We kept the debrief categories based on the CPR2 instructional model and on other aspects of instruction that we believe support the CPR2 instructional model, including facilitating rich classroom discussions that allow for student questions and reasoning, checking for student understanding, and addressing student misconceptions. We removed the category on student engagement and student learning because it was difficult to observe these categories when conducting virtual observations. Observers took running notes on individual lessons and then wrote summaries for each of the debrief categories.

Data Collection & Analysis: Due to COVID-19 restrictions, we continued to conduct lesson observations virtually. Observers were able to see either the front of the classroom (usually the screen and the teacher, although at times we were unable to see the board/screen) or the teacher's desktop. Importantly, observers still could not see students and the audio quality remained significantly limited for observers to hear student talk. Even with modifications to our protocol, our observation findings still do not fully capture aspects of teacher-student interactions outside of front-of-class teacher-led activities.

We observed eight out of 14 treatment teachers implementing 1-4 lessons for a total of 25 observed lessons. Most of the observed teachers reviewed the fall lessons and taught content from the What is Even+Odd? and What is Even+Even? Lessons. Most observed teachers did not complete the full lesson plans during the observed lessons.

To analyze the data, we created a summary debrief for all observations for each teacher. Three SRI researchers reviewed all debrief categories across all teacher summaries and described themes and/or variations for each debrief category (e.g., what kinds of questions did teachers ask students, or to what degree did teachers provide student opportunities to write general expressions to represent the mathematical relationships they discovered?). Then, the three researchers compared the themes and variations with the themes and variations from fall 2020 and fall 2021 to see what changes, if any, occurred in implementation over the duration of the study. The findings below summarize the themes and variations we saw across all teacher observations.

Findings

Overall, with some exceptions, teachers provided students opportunities to program. Teachers generally seemed to feel confident in the CPR2 content. In the lessons we observed, there was little support for exploring mathematical concepts in depth or generating and discussing mathematical conjectures. Teachers tended to follow the "script," with an emphasis on students correctly following directions. Teachers' questions and reasoning dominated instruction and whole-class discussions. There was little support for checking and adapting to student understanding.

Presence of CPR2 Instructional Model: We observed similar student opportunities and instructional patterns related to the CPR2 instructional model in spring 2022 compared to prior implementations. Students had limited opportunity to explore mathematical concepts through the programming or to write and explore mathematical conjectures based on what they were learning through the programming. All students did have opportunities to write mini-programs. In four of the classes, students were copying some semi-completed Python code from the teacher and filling in blanks. In the other four classes, students had opportunities to write programs independently and do some exploration of mathematical concepts. However, this often took the form of confirmation, rather than exploration, of patterns. For example, one teacher asked students to complete the sentence "evens plus evens is always going to be _____," and students respond in unison, "evens." All teachers mentioned or identified at least one math concept, including generalization, but only three teachers were observed defining or explaining a mathematical concept. There was one exception of students being more actively involved, when a teacher asked students to define evens and odds through questions such as "How do you know a number is even?" or "How do you know a number is odd?" While students had opportunities to write general expressions in all but one of the eight classes, in six of these classes the writing was primarily directed by the teacher, meaning students did not generate their own expressions. Instances of students writing expressions were also not based on exploring or understanding mathematical relationships. In four of the classes, students did not have opportunities to make conjectures; rather, the teacher showed them a conjecture or students copied the teacher's conjecture. In three classes students had some form of opportunity to explain their reasoning or generate conjectures as a whole class. Only one instance in which students had the opportunity to write arguments and generate their own conjectures was observed.

Teacher Capacity: Unlike the fall, where there was a range of teacher confidence in the CPR2 content, most of the teachers appeared to be confident in and have a good grasp of the CPR2 content in the spring. Seven of the eight teachers appeared confident in the math and the computer programming. One teacher was less confident with the mathematical content (was not a math teacher) and leveraged their discomfort with the math concepts as a way to teach productive struggle to students. While some teachers in the fall 2021 implementation faced challenges such as struggles with debugging, observers noted that teachers in the spring did not face any challenges.

Instructional Practices: None of the teachers, when observed by SRI researchers, articulated learning objectives or lesson goals to frame the lessons, nor did they speak to the value of the work beyond the classroom or connections with other content. This was similar to prior implementations, except that three teachers provided some framing of their lessons in fall 2021. Teachers prioritized following the lesson script while allowing for some student input, again similar to prior implementations. In six of the classrooms, observers described the classroom norms as being focused on students providing correct answers, following the teacher's direction, or otherwise deferring to teacher's initiative and reasoning. For example, in one classroom the implicit norm appeared to tend toward correct answers, in the sense that there was very little discussion and elaboration of student contributions. Once a student had stated a correct answer, the teacher would move on. In two classrooms, the teacher invited student ideas and questions and encouraged student reasoning and explanations. For example, in one classroom student voice was encouraged and all students participated in sharing their code, with a focus on explaining rather than providing correct answers. Overall, these patterns in social norms were similar across implementations. Observers noted the teacher did most of the cognitive work in all classrooms, as in other implementations. Six teachers appeared to pace their lessons well and did not appear to feel rushed. At the same time, observers noted that some teachers did not always spend time to pause and address student ideas or otherwise go into more depth with the content. Nevertheless, this was different compared to the first implementation, when most teachers appeared to feel rushed about covering the content.

Whole-class Discussion and Teacher Questions: We observed similar styles of whole-class discussion and teacher questions in spring 2022 compared to prior implementations. Whole-class discussions were limited, and student contributions during discussions came largely from a few consistent volunteers. As in the fall, teachers performed most of the cognitive work, with seven out of eight teachers dominating discussions with their own questions, ideas, and elaborations on their reasoning. One teacher held a whole class discussion that involved multiple students building onto their classmates' ideas. In all the classrooms, teachers usually asked “fill-in-the-blank" and funneling questions. In one instance, the teacher did some call-and-response questioning, but the rest of the lesson consisted of warm-calling (calling on student volunteers) with funneling questions. In another instance, the teacher asked students to explain their reasoning for why they thought the number zero was even, odd, both or neither; however, because the teacher had them write their response in a private chat to the teacher, they were unable to generate a discussion from this activity.

Teacher Support for Student Understanding: We observed similar patterns for teacher support regarding student understanding in spring 2022 compared to prior implementations. The teachers’ practices for checking student understanding continued to be circulating the room and warm-calling on students. In one instance, a teacher used informal polling of students to check how students were doing. As in the fall, across all classrooms, students rarely asked questions. Student questions most often related to debugging syntax. Teachers mostly provided feedback in the form of affirming correct answers. In one instance, a teacher praised a student for understanding the mathematical theory and for how they wrote their code. Observers noted that half of the teachers adapted their lessons based on their checks for understanding. Teachers who adapted their lessons gave students more time to share their code and revisited the distributive property and how to prove by example.

Spring Implementation Questionnaire

As in the fall, SRI researchers administered the Implementation Questionnaire in the spring to gain insight on the CPR2 implementation of the treatment group and to understand treatment-control contrast. The spring Implementation Questionnaire was adapted from the fall version to account for the differences in the CPR2 content planned for the spring semester. We updated the section in which treatment teachers selected the CPR2 lessons they taught based on the differences in planned CPR2 content across the two semesters. All teachers retained for teacher impact analyses (13 treatment teachers, seven control teachers) and student impact analyses (three treatment teachers, two control teachers) completed the Implementation Questionnaire.

Based on the evidence from the questionnaire, we believe we have appropriate contrast between treatment and control groups. We did not see significant evidence of control teachers engaging in programming instruction or mathematical generalization instruction similar to CPR2. To the extent that mathematical generalization or proof-writing was taught, it was aligned to common junior high mathematics standards (i.e., identifying the next item in a sequence, proving the Pythagorean theorem). We also have evidence the treatment group overall implemented CPR2 content as intended for spring 2022, which was further confirmed by our spring CPR2 lesson observations.

Specific findings for the treatment group include:

- Twelve treatment teachers taught all or almost all of the fall review content to remind students of key programming concepts from the fall; one did minimal review.
- All 13 treatment teachers taught the programmatic portions of the CPR2 spring content, such as modifying programs to general different sums of even and odd numbers.
- Ten treatment teachers taught the conjecturing and argumentative portions of the CPR2 spring content, such as asking students to describe what they think will happen when adding two even numbers and then writing a convincing argument.
- Two treatment teachers reported teaching mathematical generalization in their classes outside of CPR2 lessons. One teacher specifically mentioned teaching generalization in sequences and graphs; one teacher generally reported teaching generalization without specific examples.
- One treatment teacher reported teaching proof outside of CPR2 lessons, explaining that their class's Algebra 1 content incorporates proof-writing.
- One treatment teacher reported students using Python outside of CPR2 for simple coding lessons.

  Two teachers reported providing their students with more structured note-taking supports: "I also created guided, fill in the blank notes to go along with the proofs for Even+Even and Even+Odd."
- Treatment teachers did not report making other substantial adaptations. Minor adaptations included spending more time on the fall review than planned, providing extra debugging practice, and making some modifications to the UNA-provided PowerPoint slides.

  Two treatment teachers reported no challenges, four teachers reported challenges with time and scheduling, three teachers reported challenges with proof-writing, one teacher reported internet connectivity challenges, and one teacher reported students' learning loss from the fall Python lessons: "It was a bit of a challenge to get the students to understand the proof writing and why proof by example is not a valid form of proving a mathematical truth."
- Ten treatment teachers reported high levels of student engagement. Three teachers reported mixed engagement, such as students being split between "engaged and bored."

Specific findings for the control group include:

- Three of the seven control teachers reported teaching mathematical generalization or pattern-finding to their students in the spring:
    1. "We used patterns as an intro to our expressions and equations unit."
    2. "Quadratic equations"
    3. "Students worked with sequences to identify the next item in the pattern. They did this with shapes and numbers. Additionally, they did very basic recursive functions."
- Four of the seven control teachers reported teaching proof-writing or communicating with mathematical language to their students in the spring:
    1. "Pythagorean theorem"
    2. "Mathematical language"
    3. "When students had to prove the Pythagorean theorem we discussed proof writing. Most problems in math students were required to justify answers with mathematical language."
    4. "Communicating with mathematical language enhances the math concept retention; line = linear; x-axis and y-axis intersect = perpendicular"
- None of the control teachers reported teaching programming.

Spring Interviews of CPR2 Teachers

Background: The purpose of the teacher interviews was to get teachers' perspectives on their preparedness to teach CPR2 and their views on their own experience and that of their students. Another related aim was to learn what teachers see as the benefits and challenges of implementing CPR2. Interviews also offered an opportunity to triangulate among data sources.

Design: The teacher interviews followed a semi-structured protocol organized into three sections: preparing to teach CPR2 lessons, teaching CPR2 lessons, and reflecting on potential impacts of CPR2 content on student learning. Interview topics included CPR2 professional development, lesson preparation and modifications to lessons, perceived benefits of CPR2, challenges related to implementation, and teachers' impressions of student engagement and learning.

Data Collection & Analysis: SRI contacted all 14 of the treatment teachers who had been retained into the spring to request an interview. We were able to schedule interviews with six teachers. Interviews took place via videoconference and lasted approximately 40 minutes.

Findings

All six teachers reported feeling well-prepared by the CPR2 training and materials. This was consistent with the SRI team's observations that teachers appeared comfortable teaching the lessons. Teachers said they were confident going into the lessons, except for one computer science teacher who reported being nervous about making math teaching mistakes. All the teachers used the materials provided in the Summer Institute, in some cases modifying them to suit their needs. Examples of these modifications to the materials include animating the slide deck to present the information in a step-by-step manner and creating an additional practice worksheet (modeled on one of the teacher worksheets from the institute) for students. One teacher made modifications to make the materials work better in her virtual classroom.

The pacing of the lessons generally went as expected, though a bit faster for some teachers. Teachers planned between 2 and 5 days for teaching the planned CPR2 content in each of fall and spring. Four teachers found they needed less time than planned. Two teachers said they spent only one class period on the spring material. In these two cases, it was evident the teachers did not complete the lesson sequence as designed because they did not get to the conjecture and argumentation portion of the lesson.

Four teachers described students as engaged; two said students participated but were not enthusiastic. Teachers' comments on student engagement ranged from "They kept asking me when we were going to do more" to "I think they caught onto it, but the excitement wasn't there." The second comment was about a computer class that had been working with robots and designing games, indicating that perhaps engagement was higher among students for whom coding was a new experience.

Five of the six teachers interviewed said the proofs were too hard for their students. The teacher of accelerated students said the proofs weren't too challenging for her class. One teacher said the lesson would be better for pre-calculus students rather than his middle school students. The belief that the proofs are too challenging may account for observation findings that teachers tended to do most of the cognitive work for their students and dominate discussions with their own reasoning.

Five of the six teachers interviewed said they taught CPR2 in a similar way to their regular classes. Most teachers we spoke with reported that they usually employ group work, they are "nontraditional," and they typically encourage student exploration. This is somewhat at odds with observation findings that teachers tended to direct classroom activities closely rather than allow time for exploration of ideas. One teacher reported that he used more direct instruction with the CPR2 lessons because students were having trouble following them: "Normally I let them work together more and lecture less."

Teachers offered a range of advice for others who may want to implement CPR2. One suggested having a pictorial way to describe the proofs would make it easier to teach and understand. Another offered, "So just don't be afraid to fail and let the kids struggle with it there." Another suggestion was to teach it all at once instead of separating the lessons across semesters.

Teachers varied in their estimates of how much their students learned. Four teachers reported that most of their students got the material, while two said some students were not picking it up. One teacher estimated that 60-70 percent of her class understood the coding part, but only about a quarter were understanding the proofs. She said, "I think that was probably the part where I probably lost more students than the coding." Three teachers reported having to do a lot of troubleshooting during the lessons.

Five of the six teachers we interviewed named coding experience as CPR2's primary benefit to students. When asked what the benefits of CPR2 are, one teacher said, "I think the first thing I was hoping was that they would learn how to code, because I think that's a useful skill." Half the teachers (3/6) mentioned mathematical generalization as a benefit, but this was cited as the most important benefit by only one teacher (who teaches accelerated students). This may explain why no explicit emphasis on mathematical generalization was noted by observers. No teacher specifically mentioned experience with proof-writing as a benefit.

Impact Analyses

This chapter describes the impact analyses for the 2021–22 efficacy pilot. To summarize the overall finding, severe attrition during the efficacy year of the study prevents any reliable quantitative claims to be made about the extent to which CPR2 impacts teachers' or students' mathematical generalization skills or students' attitudes around computer science.

In this chapter, we first describe the impacts of CPR2 on students' performance on problems involving mathematical generalization. Second, we describe the impacts of CPR2 on students' comfort and interest in programming activities as measured through a survey. Third, we describe the impacts of CPR2 on teachers' understanding of mathematical pedagogical and content knowledge related to teaching mathematical generalization.

Student Assessment Analysis

Background: One of the goals of CPR2 is to increase student performance in problems involving mathematical generalization. We created a student assessment specifically designed to measure this skill.

Design: In fall 2020, we piloted a student assessment instrument based on several validated instruments (see Year 2 RPPR for details). The fall 2020 instrument did not perform as desired in two ways. First, students performed very poorly on the open-ended questions (less than 20% correct) such that there was little meaningful variation in either the pretest or posttest. Second, several team members questioned the appropriateness of some of the multiple-choice questions on further review given the mathematical content of the questions, doubting that CPR2 would influence performance on these items. In the spring 2021 iteration of the student assessment, we retained two of the Mathematics Assessment Resources Services (MARS) items from the fall 2020 assessment that piloted well and added seven released items from the National Assessment of Educational Progress (NAEP) 4th and 8th grade math assessment. We searched for items under "number properties and operations" and "algebra" in the content areas and, where available, "conceptual understanding" and "problem solving" for ability. The UNA and SRI teams reviewed the items to ensure the questions were a reasonable match with the CPR2 program. The NAEP assessment provides performance data for their released questions, and we used the data to determine these questions were at an appropriate difficulty level for the sample. Additionally, we piloted these items in spring 2021 and found them to perform well for our purposes. We used this nine-item version of our student assessment for the 2021-22 efficacy pilot.

Data Collection: Students in classes taught by both treatment and control teachers took the student assessment as a pretest. Students completed the assessment prior to any teacher implementation of CPR2 in classes taught by treatment teachers. In classes taught by control teachers, students completed the assessment by September 30, 2021 to minimize the likelihood of business-as-usual mathematics instruction influencing the pretest. We obtained student assessment data from 574 treatment students and 472 control students who completed the instrument and consented to have their data used for research purposes.

We then asked the 24 teachers who were retained into the fall (14 treatment teachers, 10 control teachers) to administer the student assessment again as a posttest in the spring. All teachers were asked to administer the assessment after April 15, 2022 to ensure that most of the mathematics instruction during 2021-22 had occurred prior to the posttest being administered. Treatment teachers were also required to complete their spring CPR2 instruction prior to administering the student assessment. Due to COVID-19-related complications and scheduling pressures, we received student assessment data from only 67 students in four treatment teachers' classrooms and 111 students in three control teachers' classrooms. However, due to issues with matching students' pretests and posttests (e.g., teachers changed classes during the semester break), we were only able to match 37 students from two treatment teachers' classrooms and 87 students from three control teachers' classrooms. Accordingly, the student-level attrition for the student assessment is 94% for students of treatment teachers and 82% for students of control teachers.

Findings

We conducted three analyses on the student pretest data: 1) baseline equivalence testing in the style of What Works Clearinghouse (WWC), 2) attrition *t*-tests for the initial and final samples, and 3) an impact analysis of the extent to which students receiving CPR2 instruction improved their performance on problems involving mathematical generalization. The findings—particularly those concerning impact—should be viewed as preliminary, as severe attrition precludes robust and accurate statistical modeling. In particular, our impact findings must be understood in the context of severe attrition that prevents accurate statistical modeling.

The mean pretest score of the initial student sample in treatment teachers' classrooms was 2.68 out of 9 with a standard deviation of 1.73, while the initial mean score of students in control teachers' classrooms was 3.60 with a standard deviation of 2.26 (see Table 2). The initial treatment and control student samples did not achieve baseline equivalence on the student assessment pretest based on WWC standards, meaning students' pretest survey scores across the two conditions were not similar enough to support a meaningful comparison of CPR2's effectiveness at improving students' mathematical generalization skills. Baseline equivalence tests are reported as effect sizes, and the effect size found for this pretest was -0.46, which is outside of the +/- 0.25 guidelines set by WWC. Control students outperformed treatment students on their overall scores and on seven specific test items.

The mean pretest score of the final analytic student sample in treatment teachers' classrooms was 2.59 out of 9 with a standard deviation of 2.13, while the final mean score of students in control teachers' classrooms was 2.68 with a standard deviation of 2.13 (see Table 2). The final treatment and control student samples did achieve baseline equivalence on the student assessment pre-test based on WWC standards, with an effect size of -0.05, which is within the +/- 0.25 guidelines set by WWC.

Our attrition *t*-tests for the treatment and control student samples showed the initial and final student samples for treatment students were not significantly different as measured by students' mean scores on the student assessments ($p = 0.76$). However, our attrition tests identified the initial and final student samples for control students as being significantly different ($p < 0.01$). This means the control students in our final analytic sample are measurably different from those in our initial sample at the start of the school year.

Table 2. Student assessment baseline equivalence tests and pretest attrition t-tests

		Initial Student Sample	Final Student Sample	Attrition *p*-value
Treatment	Mean	2.68	2.59	
	SD	1.73	2.13	0.76
	n	551	37	
Control	Mean	3.60	2.68	
	SD	2.26	2.13	<0.01
	n	460	87	
Baseline Equivalence		-0.46	-0.05	

Note: The student assessment is scored out of 9 points.

The combination of three factors creates significant challenges for finding any valid student-level impacts. First, since the two student groups were not equivalent at initial baseline, any potential impacts of CPR2 on student learning will need to be interpreted conservatively because from the outset we were working with groups that were demonstrably different from one another. Second, our final student sample for control students is measurably different from our initial sample of control students, meaning there may be confounding factors affecting student outcomes as represented by teachers of higher-performing control students being more likely to attrit. Third, the very high student-level attrition rates (94% and 82%) and small number of teachers (two treatment teachers, three control teachers) create severe methodological challenges for estimating student impacts.

Our original design for the student-level impact analyses was to use hierarchical linear modeling (HLM). Our study randomized participants at the teacher level and students were clustered within teachers' classrooms; HLM appropriately accounts for the clustered nature of the data. Unfortunately, HLM cannot be reliably used with a sample size of five at level-two (teacher-level) as there are too few clusters to accurately estimate teacher-level effects on student outcomes. An investigation of relevant literature[1] did not reveal any best practices for appropriately modeling clustered data with as few as five clusters. We conducted an impact analysis as best we could using students' treatment condition and pretests as independent variables, with cluster-robust standard errors to account for the clustered nature of the data.

We found a statistically significant positive effect of CPR2 on student performance on the student assessment, with treatment students on average answering 0.91 more correct student assessment questions than otherwise similar control students ($p = 0.04$). Given the challenges with baseline nonequivalence, high levels of attrition, and measurable differences between initial and final samples, we do not have confidence that this finding appropriately estimates the impact of CPR2 on students' mathematical generalization skills. We interpret this finding as suggesting that CPR2 may have had a statistically significant positive impact on students that was not reliably detectable due to high attrition.

Student Survey Analysis

Background: One of the research goals of the CPR2 study is to increase "the extent to which students feel comfortable with the programming activities and with the associated mathematics, [and] the extent to which they would be interested in similar activities in the future." We intended to measure students' comfort and interest in programming activities during the efficacy study through a student survey.

---

[1] Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.

Design: We based our student survey on the Student Computer Science Attitude Survey[2]. This survey was tested and validated in 2010–2016 for grade 8+ students for measuring five attitudinal constructs related to computer science:

- Students' **confidence** in their ability to learn CS skills and solve CS problems
- Students' **interest** in learning computer science and solving problems
- Students' perceptions of **belonging** in computer science
- Students' beliefs in the **usefulness** of learning computer science
- Students' perceptions of being **encouraged** to study computer science

In fall 2020, in partnership with UNA and HRI, we reviewed the survey items and concluded they were appropriate for measuring the study objectives. During the 2020-21 school year, we piloted the student survey. We found the instrument to have acceptable internal consistency reliability and found students responded meaningfully to the items. Four of the survey factors had good reliability with our student population: Confidence, Interest, Usefulness, Encouragement. Our piloting data indicated that one factor, Belongingness, did not perform as well with middle school students. We decided to keep the survey instrument intact (i.e., in the same form used in prior validation research by its creator) rather than remove the Belongingness survey items. See Appendix B for full reliability testing information.

Data Collection: Students in both treatment and control classes took the student survey as a part of our single data collection instrument for students. Students completed the survey immediately after taking the pre-assessment. We obtained responses from 445 treatment students and 398 control students who completed the survey and consented to have their data used for research purposes.

We then asked the 24 teachers who were retained into the fall (14 treatment teachers, 10 control teachers) to administer the student survey again as a posttest in the spring. All teachers were asked to administer the survey after April 15, 2022 to ensure that most of the mathematics instruction in the school year had occurred prior to the posttest being administered. Treatment teachers were also required to complete their spring CPR2 instruction prior to administering the student survey. Due to COVID-19-related complications and scheduling pressures as well as issues with matching students' pretests and posttests, we were only able to match 25 students from two treatment teachers' classrooms and 69 students from three control teachers' classrooms. Accordingly, the student-level attrition for the student survey is 94% for students of treatment teachers and 83% for students of control teachers.

---

[2] Haynie, K.C. and Packman, S. (2017). *AP CS Principles Phase II: Broadening Participation in Computer Science Final Evaluation Report*. Prepared for The College Board and the National Science Foundation, February 12, 2017. Skillman, NJ.

Findings

We conducted three analyses of the student survey pretest data: baseline equivalence testing in the style of WWC, attrition *t*-tests for the initial and final samples, and an impact analysis of the extent to which students receiving CPR2 instruction increases their confidence, interest, belongingness, usefulness, and encouragement attitudes toward computer science. Again, these impact findings must be viewed with caution as the attrition from the sample prevents accurate statistical modeling.

Each survey item was measured on a 1-4 Likert-style scale (1 = strongly disagree, 4 = strongly agree). Each construct consisted of five items that were summed for an overall factor score range of 5-20 (12.5 midpoint).

For the pretests of the initial student sample, students overall reported either modest disagreement or mixed agreement across all five factors. All factors had an average score between 10 and 13, which equates to an average item score between 2.0 (disagree) and 2.6 (in between disagree and agree). The treatment and control samples achieved baseline equivalence on four factors (Confidence, Interest, Usefulness, Encouragement) based on WWC standards. One factor, Belongingness, did not achieve baseline equivalence. Belongingness has a baseline difference of 0.26, which is above the +/- 0.25 guidelines set by WWC. See Table 3 below for full review of the initial student sample.

Table 3. Initial student samples pretest survey results by construct, treatment, and control

		Confidence	Interest	Belongingness	Usefulness	Encouragement
Treatment	Mean	12.1	11.6	12.3	12.9	10.6
	SD	2.88	3.41	2.70	3.03	3.12
	n	445	442	445	441	443
Control	Mean	11.5	11.3	11.6	12.3	10.2
	SD	3.02	3.34	2.63	3.16	3.06
	n	398	398	398	398	398
Baseline Difference		0.20	0.09	0.26	0.19	0.13

Note: Factors were only assessed for a student if they replied to all five items for a factor, which is why the *n*'s may vary across factors.

For the pretests of the final analytic student sample, students overall reported either modest disagreement or mixed agreement across all five factors. All factors had an average score between 10 and 13.5, which equates to an average item score between 2.0 (disagree) and 2.7 (in between disagree and agree). The treatment and control samples failed to achieve baseline equivalence on four factors (Confidence, Interest, Belongingness, Usefulness) based on WWC standards with greater baseline differences than we found with the initial student samples. See Table 4 below for full review of the final student sample.

Table 4. Final student samples pretest survey results by construct, treatment, and control

		Confidence	Interest	Belongingness	Usefulness	Encouragement
Treatment	Mean	12.8	12.3	12.9	13.3	10.6
	SD	2.10	2.23	2.05	1.57	2.25
	n	25	25	25	25	25
Control	Mean	11.5	11.2	11.9	11.9	10.3
	SD	2.74	3.22	2.49	3.20	3.00
	n	69	69	69	69	69
Baseline Difference		0.52	0.35	0.42	0.49	0.11

Our attrition *t*-tests for the treatment and control student samples did not find any statistically significant differences between the initial and final samples on any of the survey constructs. However, this may in part be due to the overall low number of treatment students in the final sample ($n = 25$) resulting in underpowered statistical tests.

Table 5. Attrition t-tests of the student survey pretest

	Confidence	Interest	Belongingness	Usefulness	Encouragement
Treatment *p*-values	0.23	0.31	0.27	0.51	1.00
Control *p*-values	1.00	0.81	0.38	0.33	0.80

The combination of two factors creates significant challenges for finding any valid student-level impacts. First, since the two student groups were not equivalent in most factors in the final analytic sample, any potential impacts of CPR2 on students' attitudes toward computer science will need to be interpreted conservatively as we are investigating students who were demonstrably different from one another in our analysis sample. Second, the very high student-level attrition rates (94% and 83%) and small number of teachers (two treatment teachers, three control teachers) create severe methodological challenges for estimating student impacts. We describe our original design for student-level impact analyses above in the student assessment section. For the student survey, we conducted an impact analysis as well as we could for each construct using students' treatment condition and pretest factor scores as independent variables. We did not find a statistically significant impact of CPR2 on the attitudinal measures.

Learning Mathematics for Teaching (LMT)

Background: The LMT assessment was developed by the University of Michigan to measure teachers' understanding of the mathematical pedagogical and content knowledge teachers need to teach mathematics well. We selected items from the Middle School Patterns, Functions, and Algebra content area to measure whether CPR2 improves middle school mathematics teachers' understanding of generalizability and patterns.

Design & Analysis: During the 2019-20 pilot year we developed and piloted a pre- and post-test using items from the Middle School Patterns, Functions, and Algebra content area and balanced the two tests on difficulty, total number of items, content, and types of questions. We focused our item selection on questions related to mathematical generalization that were well-aligned to CPR2 content in general without being overly aligned to the specific CPR2 learning activities.

We ran a two-parameter item response theory (IRT) analysis using the difficulty and discrimination values for each item that were provided by the test developers based on prior data. The IRT analysis estimates both item parameters and ability estimates. For the IRT model we used the parameters item difficulty and item discrimination. The item parameters set the scale and are used to estimate ability. This type of model accounts for how challenging items are versus more traditional methods that give equal weight to each item. The analysis returned estimates of teachers' results expressed as standard scores that can be readily compared to one another. Our piloting process indicated the items we selected and the form overall worked well for our purposes. Thus, we used our piloted forms for the 2021–22 efficacy trial. We consulted with both UNA and Horizon Research team members in item selection, item balancing across tests, and interpretation of the piloting results.

Data Collection: We administered our piloted version of the LMT to both treatment and control teachers as both a pretest at the end of the 2020–21 school year and a post-test at the end of the 2021-22 school year. Treatment teachers were required to complete the LMT before the first day of the CPR2 Summer Institute to avoid CPR2 content affecting their pretest scores.

From the initial recruitment sample, 22 treatment teachers (out of 25) and 17 control teachers (out of 24) completed the LMT. Of those teachers, 13 treatment teachers and seven control teachers completed the LMT as a posttest at the end of the 2021–22 school year. Relative to the initial recruitment sample, we experienced 48% attrition from the teacher treatment group and 71% attrition from the teacher control group for our teacher impact analysis. Given the significant attrition, we must interpret the findings with caution.

Findings

We conducted three analyses of the LMT data: baseline equivalence tests for the final analytic sample in the style of WWC, pretest attrition *t*-tests, and impact analyses using single-level regression. In particular, the impact findings were affected by high attrition, and therefore the findings should be viewed as preliminary.

We calculated the mean standardized teacher LMT pretest scores both for the initial recruitment sample (all of the teachers who took the LMT prior to attrition) and for the final analytic sample (the teachers who completed an LMT posttest). The mean standardized teacher LMT score for the original sample (n = 22) was 0.15, and the control mean for the original sample (n = 17) was 0.04. With the small sample sizes for each condition, the standard deviations (0.82 for the initial treatment sample and 0.75 for the initial control sample) were quite high relative to the mean difference. The means of the final analytic samples were similarly different from one another (0.25 for the final treatment sample, 0.13 for the final control sample) with similar standard deviations (0.86 for the final treatment sample, 0.70 for the final control sample).

The treatment and control samples achieved baseline equivalence based on WWC standards, meaning teachers' LMT pretest scores across the two conditions were similar enough to support a meaningful comparison of CPR2's effectiveness at increasing teachers' mathematical generalization knowledge. Baseline equivalence tests are reported as effect sizes, with the initial samples having a baseline difference of 0.14 and the current samples having a baseline difference of -0.15. These values are within the +/- 0.25 guidelines set by WWC. Even though baseline equivalence was achieved, we must be cautious when interpreting our results given the overall low sample sizes and significant teacher attrition.

The attrition *t*-tests did not identify any statistically significant differences between the initial and final treatment samples (p = 0.73) or the initial and final control samples (p = 0.79). This suggests teacher attrition did not significantly change the mean values of the LMT pretest for either condition. However, this could be more reflective of the small sample sizes than a lack of change in the mean scores due to attrition, with the *t*-tests likely being underpowered to detect such a change. The significant attrition rates could have impacted the teacher samples in unmeasured ways.

Table 6. LMT baseline equivalence tests and pretest attrition t-tests

		Initial Recruitment sample	Final Analytic sample	Attrition *p*-value
Treatment	Mean	0.15	0.25	0.73
	SD	0.82	0.86	
	n	22	13	
Control	Mean	0.04	0.13	0.79
	SD	0.75	0.70	
	n	17	7	
Baseline Difference		0.14	0.15	

We then conducted LMT impact analyses to examine the extent to which participating in CPR2 professional development activities and delivering CPR2 instruction had any impacts on teachers' understanding of how to teach mathematical as measured by the LMT. We used single-level regression to estimate the impacts with teachers' treatment condition and LMT pretest as independent variables (see Table 7). Our analyses did not find evidence of a statistically significant impact of CPR2 experience on teachers' knowledge on how to teach mathematical generalization.

Table 7. LMT impact analysis

Variable	β	SE	*p*-value
Treatment (1 = yes)	-0.13	0.24	0.59
LMT Pretest	0.64	0.15	<0.01

Note: Treatment *n* = 13, control *n* = 7, R-squared = 0.53.

Conclusion

This study was designed to answer research questions about the initial design and implementation of CPR2, CPR2 implementation in middle school classrooms, and the potential of CPR2 to improve mathematical generalization skills of both students and teachers. Below we answer our original seven research questions using 3 years of accumulated findings.

> *To what extent are teachers able to implement CPR2 in 7th and 8th grade math classrooms? What are the challenges and what works well?*

Teachers who experienced the CPR2 Summer Institute were able to deliver the CPR2 lessons largely as modeled during the Summer Institute and as the CPR2 materials directed. In CPR2 practice prior to the current project, the activities had always been led by visiting UNA faculty. This project demonstrated that teachers could, with professional development and support, lead the CPR2 activities themselves. Regarding the four-step CPR2 Instructional Model, teachers were able to articulate essential mathematical concepts (Step 1), lead students to write mini-programs to explore those concepts (Step 2), and in writing general expressions (Step 3). The research team saw teachers leading students in following along as they showed conjectures based on the Python programs. However, we did not observe students independently making conjectures and writing convincing arguments for those conjectures (Step 4).

The CPR2 instruction we observed was predominantly teacher-centered, with students following directions and sometimes copying code from teachers, projected slides, or fellow students. At times teachers were challenged by a lack of computers or reliable internet connections, difficulty setting up the coding environments, and uncertainty around debugging during classroom instruction. Teachers were generally not observed providing students with open time and space for tinkering, exploration, and thinking through programming or mathematical generalization challenges on their own as the instructional model calls for.

> *To what extent does CPR2 focus on teacher and student needs in learning programming for generalization?*
>
> *To what extent does CPR2 professional development support teachers in learning to use programming as a tool to develop mathematics generalization skills?*

The design-based implementation sessions in Years 1 and 2 were successful in involving stakeholders who knew student and teacher needs well, and whose feedback was incorporated into CPR2 materials. The CPR2 Summer Institute and instructional materials were successful in supporting teachers in developing their Python programming skills for the specific purpose of using programming as a tool for teaching mathematical generalization. Over the course of the 3-year study, the CPR2 professional development and instructional materials became more successful in supporting teachers in developing their Python programming skills. In the final year, teachers reported very high levels of preparedness and confidence in their Python skills for teaching CPR2. While some teachers struggled at times with programming when delivering CPR2 content, teachers generally were able to resolve their programming challenges.

Observation and interview data suggest that teachers increased their computer programming knowledge and skill at delivering the programming content required for CPR2 lessons. Our impact analysis of the LMT assessment data did not find a statistically significant impact of receiving CPR2 professional development and delivering CPR2 instruction on teachers' mathematics generalization skills.

During the initial implementation and pilot studies, students were able to follow along with the CPR2 lessons and complete the coding and mathematical activities. Assessing with greater detail whether CPR2 instruction met students' needs in learning programming for generalization was difficult due to three factors: 1) virtual observations necessitated by COVID-19 made it harder to see student work and impossible to do the interviews we had planned, 2) the common practice of students' copying down teacher-provided code allowed them to go through the lessons without generating their own code, and 3) our study did not incorporate a programming performance task, as we anticipated having greater access to student classroom work products than we were afforded by virtual observations.

> *How can teachers support students in their engagement and learning through the designed activities?*
>
> *How can students engage in CPR2 as intended in order to achieve the targeted outcomes?*

Teachers delivered CPR2 lessons as modeled during the CPR2 Summer Institute, which enabled students to follow along with the lesson activities as intended. During the CPR2 Summer Institute, however, UNA faculty described several important types of engagement, including tinkering, exploration, student-driven experimentation, and student-led whole-class discussions. While some teachers engaged students with substantive discussions about programming or mathematical generalization, the delivery and facilitation of the CPR2 lessons was mostly teacher-led with little room for students to grapple independently or collaboratively with the CPR2 content. Similarly, while teachers successfully delivered CRP2 content aligned to the first three steps of the CPR2 instructional model, they did not in most cases engage students in the fourth step, conjecture and argumentation.

> *To what extent do students feel more comfortable with programming, and using programming as a tool for math?*

Classroom observations suggested most students who were new to Python were able to successfully write simple Python programs with reasonable supports and instruction for the purpose of CPR2 lessons. As many students did not previously have any programming experience, this means that many students became more comfortable with programming because of CPR2 instruction. In interviews, teachers reported their students improved their programming skills as a result of CPR2.

Due to COVID-19-related challenges with collecting student data, our study had severe student attrition from pretest to posttest. Thus we were not able to systematically assess whether improvement in attitudes toward programming occurred due to CPR2. Moreover, the final analytic sample failed to achieve baseline equivalence for four of the five attitudinal constructs our survey measured.

> *What effect does CPR2 have on student performance in problems involving generalization?*

We found a statistically significant impact of CPR2 on student performance on our student assessment. However, we do not believe this result is reliable due to the severe student attrition and lack of baseline equivalence. Additionally, we found that our initial recruitment and final analytic student samples differed, confirming attrition substantively changed the makeup of our student samples.

Future Work

We see two main avenues for future work. First, we believe another efficacy trial of CPR2 is needed to more appropriately assess the potential impacts of CPR2 on students' mathematical generalization skills and attitudes toward computer science. This second trial could also explore whether additional time allotted to CPR2 lessons would allow for a more student-centered pedagogy, as the instructional model calls for. Second, we believe the field of mathematics instruction would benefit from a new mathematics assessment aimed at assessing mathematical generalization skills. One of the challenges for our study was developing a grade level-appropriate assessment of students' mathematical generalization ability. We discovered that not only was there little assessment of mathematical generalization for middle school grades, but there were also few standardized and/or validated assessments of mathematical generalization at all. While we believe the assessment we used was the best option available, there is an opportunity to develop an assessment of students' mathematical generalization skills similar to the University of Michigan's LMT assessment.

Appendices

Appendix A: Map of Recruited Teachers

[Removed from public reporting to protect participant privacy]

Appendix B: Teacher Background Survey Responses

In our initial recruitment sample, we received teacher background survey responses from 20 treatment teachers and 17 control teachers. In our final analytic sample, 13 treatment teachers and seven control teachers completed at least one outcome data collection activity.

One of the treatment teachers did not respond to all background survey questions, which is why some answers have only *n* = 19 or *n* = 12 treatment teachers for some questions. For one question, one control teacher did not respond, which is why *n* = 16 for one question. For this question, the control teacher did not complete an outcome data collection activity and was not included in the final analytic sample.

Responses for Question #1: "Number of years spent teaching…"

Variable		Initial Recruitment Sample			Final Analytic Sample		
		Overall	Treatment	Control	Overall	Treatment	Control
Any grade level and any subject	Max	37	37	32	32	27	32
	Mean	11.4	12.0	11.4	11.9	11.8	12.2
	Min	1	4	1	1	4	1
At current school	Max	30	30	30	30	15	30
	Mean	5.9	5.9	7.5	6.1	4.3	9.2
	Min	1	1	1	1	1	1
Math	Max	37	37	30	30	27	30
	Mean	10.4	10.8	10.8	10.9	10.3	11.9
	Min	1	1	1	1	1	1
Programming and/or computer science	Max	4	4	3	2	2	0
	Mean	0.4	0.5	0.2	0.3	0.5	0
	Min	0	0	0	0	0	0

Note: In the initial recruitment sample, overall responses *n* = 36, treatment *n* = 19, and control *n* = 17. In the final analytic sample, overall responses *n* = 19, treatment *n* = 12, and control *n* = 7.

Responses for Question #2: "What is the highest level of education you've attained?"

Degree	Initial Recruitment Sample			Final Analytic Sample		
	Overall	Treat.	Control	Overall	Treat.	Control
Bachelor's degree	25%	26.3%	23.5%	26.3%	33.3%	14.3%
Some courses past bachelor's degree	13.9%	10.5%	17.6%	21.1%	16.7%	28.6%
Master's degree	52.8%	57.9%	52.9%	36.8%	33.3%	42.9%
Ph.D. or other doctorate degree	5.6%	10.5%	0%	10.5%	16.7%	0%
Other degree	2.8%	0%	5.9%	5.3%	0%	14.3%

Note: In the initial recruitment sample, overall responses *n* = 36, treatment *n* = 19, and control *n* = 17. In the final analytic sample, overall responses *n* = 19, treatment *n* = 12, and control *n* = 7. One respondent wrote "Educational Specialist" for "other."

Responses for Question #3: “What, if any, of the following additional certifications do you have?”

Certification	Initial Recruitment Sample			Final Analytic Sample		
	Overall	Treatment	Control	Overall	Treatment	Control
National Board	5	0	5	1	0	1
STEM Certificate	0	0	0	0	0	0
Other	9	5	4	5	3	2

Note: In the initial recruitment sample, overall responses *n* = 14, treatment *n* = 6, and control *n* = 8. Other responses include: AMSTI, Computer science thru code.org, Online teaching, Google, math, currently working on NCTB, and ESL endorsement. In the final analytic sample, overall responses *n* = 6, treatment *n* = 3, and control *n* = 2. Other responses included: currently working on NCTB, online teaching, and ESL endorsement.

Responses for Question #4: “Have you been awarded one or more bachelor’s and/or graduate degrees in the following fields? (With regard to bachelor’s degrees, count only areas in which you majored. Do not count endorsements or certificates.)”

Field	Initial Recruitment Sample			Final Analytic Sample		
	Overall	Treat.	Control	Overall	Treat.	Control
Education (general or subject specific)	29	15	6	16	10	6
Mathematics	18	10	8	9	6	3
Statistics	1	0	1	0	0	0
Computer Science	1	0	1	0	0	0
Engineering	2	1	1	2	1	1
Other	5	3	2	2	1	1

Note: In the initial recruitment sample, overall responses *n* = 37, treatment *n* = 20, and control *n* = 17. Other responses include: business administration - finance/accounting, economics and business, human development and family studies, accounting, and economic development. In the final recruitment sample, overall responses *n* = 20, treatment *n* = 13, control *n* = 7. Other responses include: business administration - finance/accounting and human development and family studies.

Responses for Question #5: “If “Education” selected, what type of education degree do you have?”

Type of Education Degree	Initial Recruitment Sample			Final Analytic Sample		
	Overall	Treat.	Control	Overall	Treat.	Control
Elementary Education	8	5	3	5	3	2
Secondary Mathematics	16	9	7	9	7	2
Secondary Science	0	0	0	0	0	0
Other	5	1	4	2	0	2

Note: In the initial recruitment sample, overall responses *n* = 37, treatment *n* = 20, and control *n* = 17. Other responses include: interdisciplinary studies, middle level education, middle grades math/science, and continuing ed. In the final analytic sample, overall responses *n* = 20, treatment *n* = 13, control *n* = 7. Other responses include: middle level education.

Responses to Question #6: "Have you received any type of professional development apart from this summer institute related to math during the last 12 months?"

	Initial Recruitment Sample			Final Analytic Sample		
Response	Overall	Treatment	Control	Overall	Treatment	Control
Yes	19	12	7	13	8	5
No	18	8	10	7	5	2

Note: In the initial recruitment sample, overall responses *n* = 37, treatment *n* = 20, control *n* = 17. In the final analytic sample, overall responses *n* = 20, treatment *n* = 13, control *n* = 7.

Responses to Question #7: "If yes, please select the format(s) of the professional development apart from this summer institute related to math during the last 12 months?"

	Initial Recruitment Sample			Final Analytic Sample		
Professional Development Format	Overall	Treat.	Control	Overall	Treat.	Control
I attended a professional development program/workshop.	16	11	5	10	7	3
I attended a national, state, or regional mathematics teacher association meeting.	2	2	0	2	1	1
I completed an online course/webinar.	17	12	5	10	7	3
I participated in a professional learning community/lesson study/teacher study group	3	3	0	3	3	0
I received assistance or feedback from a formally designated coach/mentor.	5	2	3	3	1	2
I took a formal course for college credit.	3	0	3	2	0	2
Other	1	1	0	0	0	0

Note: In the initial recruitment sample, overall responses *n* = 19, treatment *n* = 12, control *n* = 7. Other response: Leadership. In the final analytic sample, overall responses *n* = 13, treatment *n* = 8, control *n* = 5.

Responses to Question #8: "What is the total amount of time you've spent on professional development related to mathematics or mathematics teaching in the last 12 months?"

	Initial Recruitment Sample			Final Analytic Sample		
	Overall	Treatment	Control	Overall	Treatment	Control
Max	100	100	30	100	100	30
Mean	20.7	21.8	16.6	21.4	24.4	16.6
Min	3	6	3	5	6	5

Note: In the initial recruitment sample, overall responses *n* = 19, treatment *n* = 12, control *n* = 7. In the final analytic sample, overall responses *n* = 13, treatment *n* = 8, control *n* = 5.

Responses for Question #9: "Have you received any type of professional development apart from this summer institute related to math during the last 12 months?"

Response	Initial Recruitment Sample Overall	Initial Recruitment Sample Treatment	Initial Recruitment Sample Control	Final Analytic Sample Overall	Final Analytic Sample Treatment	Final Analytic Sample Control
Yes	3	2	1	2	1	1
No	34	18	16	18	12	6

Note: In the initial recruitment sample, overall responses *n* = 37, treatment *n* = 20, control *n* = 17. In the final analytic sample, overall responses *n* = 20, treatment *n* = 13, control *n* = 7.

Responses for Question #10: "If yes, please select the format(s) of the professional development apart from this summer institute related to computer science during the last 12 months?"

Professional Development Format	Initial Recruitment Sample Overall	Initial Recruitment Sample Treat.	Initial Recruitment Sample Control	Final Analytic Sample Overall	Final Analytic Sample Treat.	Final Analytic Sample Control
I attended a professional development program/workshop.	2	1	1	1	0	1
I attended a national, state, or regional computer science teacher association meeting.	1	1	0	1	1	0
I completed an online course/webinar.	2	1	1	1	1	0
I participated in a professional learning community/lesson study/teacher study group	1	1	0	0	0	0
I received assistance or feedback from a formally designated coach/mentor.	0	0	0	0	0	0
I took a formal course for college credit.	0	0	0	0	0	0
Other	0	0	0	0	0	0

Note: In the initial recruitment sample, overall responses *n* = 3, treatment *n* = 2, control *n* = 1. In the final analytic sample, overall responses *n* = 2, treatment *n* = 1, control *n* = 1.

Responses for Question #11: "What is the total amount of time you have spent on professional development related to computer science or computer science teaching in the last 12 months?"

	Initial Recruitment Sample Overall	Initial Recruitment Sample Treatment	Initial Recruitment Sample Control	Final Analytic Sample Overall	Final Analytic Sample Treatment	Final Analytic Sample Control
Max	90	90	5	5	5	5
Mean	31.7	47.5	5	5	5	5
Min	5	5	5	5	5	5

Note: In the initial recruitment sample, overall responses *n* = 3, treatment *n* = 2, control *n* = 1. In the final analytic sample, overall responses *n* = 2, treatment *n* = 1, control *n* = 1.

Responses for Question #12 (open-ended): "What did you hope to learn or achieve by signing up for CPR2?**"**

Reason	Initial Recruitment Sample Overall	Initial Recruitment Sample Treat.	Initial Recruitment Sample Control	Final Analytic Sample Overall	Final Analytic Sample Treat.	Final Analytic Sample Control
Methods for teaching generalization	1	1	0	1	1	0
How to incorporate math with computer science	13	7	6	9	5	4
Ways to increase student engagement in computer science and/or math	5	3	2	1	1	0
How to improve their own mathematical pedagogical practices	4	3	1	1	1	0
Programming	2	1	1	0	0	0
Other (general)	11	5	6	7	5	2

Note: In the initial recruitment sample, overall responses *n* = 36, treatment *n* = 20, control *n* = 16. In the final analytic sample, overall responses *n* = 19, treatment *n* = 13, control *n* = 6. Other (general) responses include: better ways to equip my students, I enjoy professional development, and new ways to teach.

Responses for Question #13: "What technology do you plan to have students use for CPR2 lessons?"

Technology	Initial Recruitment Sample Treatment	Final Analytic Sample Treatment
Chromebook	17	11
Laptop	6	5
Tablet	5	4
Desktop	2	1

Note: Only treatment teachers were asked this question. In the initial recruitment sample, treatment *n* = 20. In the final analytic sample, treatment *n* = 13.

Responses for Question #14: "How do you plan for your students to access technology to engage in CPR2 lessons?"

Type of access	Initial Recruitment Sample Treatment	Final Analytic Sample Treatment
With computers in my classroom that each student can use	6	2
With our school's 1:1 technology program, in which each student has his/her own school-supplied device	9	6
A computer cart that I share with other teachers	4	3
Unsure	1	2

Note: In the initial recruitment sample, all 20 treatment teachers responded to this question. In the final analytic sample, all 13 teachers responded to this question**.**

Responses for Question #15: "After completing your undergraduate degree and prior to becoming a teacher, did you have a full-time job in a mathematics-related field (for example: accounting, engineering, computer programming)? If yes, please describe."

	Initial Recruitment Sample			Final Analytic Sample		
Response	Overall	Treatment	Control	Overall	Treatment	Control
Yes	4	1	3	1	0	1
No	33	19	14	19	13	6

Note: In the initial recruitment sample, overall responses *n* = 37, treatment *n* = 20, control *n* = 17. Previous mathematics-related field careers include: accountants and production engineer. In the final analytic sample, overall responses *n* = 20, treatment *n* = 13, control *n* = 7. Previous mathematics-related field careers include production engineer.

Responses for Question #16: "Do you participate in any informal STEM training or STEM activities (e.g. hobbies, citizen science, volunteer work)? If yes, please describe."

	Initial Recruitment Sample			Final Analytic Sample		
Response	Overall	Treatment	Control	Overall	Treatment	Control
Yes	23	13	10	12	7	5
No	14	7	7	8	6	2

Note: In the initial recruitment sample, overall responses *n* = 37, treatment *n* = 20, control *n* = 17. Some of these informal STEM training or STEM activities include: AMSTI, tutoring, TI Innovator workshop, teaching at STEM camps, and TECHFIT (a program where students learn to compute programs to create exercise games). In the final analytic sample, overall responses *n* = 20, treatment *n* = 13, control *n* = 7. Some of these informal STEM training or STEM activities include: AMSTI, TI Innovator workshop, and teaching at STEM camps.

Appendix C: Implementation Questionnaires

Treatment Fall Implementation Questionnaire

Introduction/Background:

We'd like you to complete a separate questionnaire for each class in which you gave the student assessment and student survey earlier in the semester. For the following questions, please respond for one class in which you taught CPR2 only.

1. For which class period are you filling out this questionnaire?
2. What grade level(s) were the students in the class you taught your CPR2 lessons to? (Select all that apply)
   - ☐ 7
   - ☐ 8
   - ☐ Other: ___________________
3. In what course were the students in the class you taught your CPR2 lessons to?
   - ☐ 7th Grade Math
   - ☐ 8th Grade Math
   - ☐ Pre-Algebra
   - ☐ Algebra
   - ☐ Computer Science
   - ☐ Other: ____________________
4. Over how many days did you teach the CPR2 lesson to that class?
   - ☐ 1
   - ☐ 2
   - ☐ 3
   - ☐ 4
   - ☐ 5
   - ☐ 6+
5. Select the portions of the CPR2 lessons you taught to that class: (select all that apply)

   Intro to Programming:

   - ☐ Programming and Computing (4 major concepts in computing and 3 major constructs in programming)
   - ☐ Introduction to Compilers/Interpreters
   - ☐ Operators and Strings in Python
   - ☐ Modeling with Flow Charts
   - ☐ Iteration in Python ("While" loops)
   - ☐ Multiple Columns and Separators

   EOC Even:

   - ☐ Response 1 – write out a program modification to add a column of even numbers
   - ☐ Response 2 – write a general expression for even numbers
   - ☐ Response 3 – do you think 2n will always produce an even number? Why or why not?

EOC Odd:

- ☐ Response 4 – write out a program modification to add a column of odd numbers
- ☐ Response 5 – write a general expression for odd numbers
- ☐ Response 6 – is 0 and even, odd, neither, or both?

Open-Ended Questions

6. Did you teach generalization to your students in that class this semester outside of the CPR2 lessons?
   - ☐ Yes
   - ☐ No

   [if Yes] Please briefly describe the lesson(s) you taught that addressed generalization.
7. Did you teach proof-writing to your students in that class this semester outside of the CPR2 lessons?
   - ☐ Yes
   - ☐ No

   [if Yes] Please briefly describe the lesson(s) you taught that addressed proof-writing.
8. Did you use Python with your students in that class this semester outside of the CPR2 lessons?
   - ☐ Yes
   - ☐ No

   [if Yes] Please briefly describe the lesson(s) you taught that addressed Python.
9. What adaptations, modifications, or additional supplemental resources did you provide to your students in that class?
10. What challenges did you experience while implementing the CPR2 lessons in that class, if any?
11. How would you describe your students' participation in the CPR2 lessons in that class (e.g., engaged, enthusiastic, bored)?
12. What do you think your students learned from the CPR2 lessons in that class? What do you see as evidence of this learning?
13. Did anything out of the ordinary occur during your CPR2 lessons in that class? (e.g., weather-related class disruptions, unexpected assemblies, hybrid schooling models, students pulled out for testing, etc.)

Treatment Spring Implementation Questionnaire

Introduction/Background:

We'd like you to complete a separate questionnaire for each class in which you gave the student assessment and student survey earlier in the semester. For the following questions, please respond for one class in which you taught CPR2 only.

1. For which class period are you filling out this questionnaire?
2. What grade level(s) were the students in the class you taught your CPR2 lessons to? (Select all that apply)
   - ☐ 7
   - ☐ 8
   - ☐ Other: ___________________

3. In what course were the students in the class you taught your CPR2 lessons to?
   - ☐ 7th Grade Math
   - ☐ 8th Grade Math
   - ☐ Pre-Algebra
   - ☐ Algebra
   - ☐ Computer Science
   - ☐ Other: ____________________
4. Over how many days did you teach the CPR2 lesson to that class?
   - ☐ 1
   - ☐ 2
   - ☐ 3
   - ☐ 4
   - ☐ 5
   - ☐ 6+
5. Select the portions of the CPR2 lessons you taught (select all that apply):

   Review Programming and General Expressions

- ☐ Basic Operators
- ☐ Basic Loop in Python
- ☐ Create a Table with 2 Columns
- ☐ General Expression for Even
- ☐ General Expression for Odd
- ☐ What Is Zero?

   EOC Even + Even

- ☐ Response 7 – Write out a program modification to generate the sum of two even numbers.
- ☐ Response 8 – What do you think happens when you add an even number to an even number?
- ☐ Response 9 – Write a convincing argument for your conjecture in Response 8.

   EOC Even + Odd

- ☐ Response 10 – Write out a program modification to generate the sum of an even number and an odd number.
- ☐ Response 11 – What do you think happens when you add an even number to an odd number?
- ☐ Response 12 – Write a convincing argument for your conjecture in Response 11.

Open-Ended Questions

6. Did you teach generalization to your students in that class this semester outside of the CPR2 lessons?
   - ☐ Yes
   - ☐ No

   [if Yes] Please briefly describe the lesson(s) you taught that addressed generalization.
7. Did you teach proof-writing to your students in that class this semester outside of the CPR2 lessons?

- ☐ Yes
- ☐ No

[if Yes] Please briefly describe the lesson(s) you taught that addressed proof-writing.

8. Did you use Python with your students in that class this semester outside of the CPR2 lessons?
   - ☐ Yes
   - ☐ No

   [if Yes] Please briefly describe the lesson(s) you taught that addressed Python.
9. What adaptations, modifications, or additional supplemental resources did you provide to your students in that class?
10. What challenges did you experience while implementing the CPR2 lessons in that class, if any?
11. How would you describe your students' participation in the CPR2 lessons in that class (e.g., engaged, enthusiastic, bored)?
12. What do you think your students learned from the CPR2 lessons in that class? What do you see as evidence of this learning?
13. Did anything out of the ordinary occur during your CPR2 lessons in that class? (e.g., weather-related class disruptions, unexpected assemblies, hybrid schooling models, students pulled out for testing, etc.)

Control Fall & Spring Questionnaire

Introduction/Background:

1. What grade level(s) are you teaching this semester? (Select all that apply)
   - ☐ 7
   - ☐ 8
   - ☐ Other: ___________________
2. What courses are you teaching this semester? (Select all that apply)
   - ☐ 7^{th} Grade Math
   - ☐ 8^{th} Grade Math
   - ☐ Pre-Algebra
   - ☐ Algebra
   - ☐ Computer Science
   - ☐ Other: _____________________

Open-Ended Questions

3. Did you teach mathematical generalization or pattern finding to your students this semester?

☐ Yes

☐ No

[if Yes] Please briefly describe the lesson(s) you taught that addressed generalization or pattern finding.

4. Did you teach proof-writing or communicating with mathematical language to your students this semester?

☐ Yes

☐ No

[if Yes] Please briefly describe the lesson(s) you taught that addressed proof-writing or communicating with mathematical language.

5. Did you teach Python programming to your students this semester?

- ☐ Yes
- ☐ No

[if Yes] Please briefly describe the lesson(s) you taught that included Python.

6. Did you teach programming as a means for students to learn about mathematical concepts this semester?

- ☐ Yes
- ☐ No

[if Yes] Please briefly describe the lesson(s) you taught that addressed programming as a means for students to learn about mathematical concepts.

Appendix D: Student Assessment Items

Item #1 (multiple choice): "If *n* is any integer, which of the following expressions must be an odd integer?"

Response Options
n+1
2n
2n+1
3n
3n+1

Item #2 (multiple choice): "According to the pattern suggested by the four examples above, how many consecutive odd integers are required to give a sum of 144?"

1 + 3 = 4
1 + 3 + 5 = 9
1 + 3 + 5 + 7 = 16

Response Options
9
12
15
36
72

Item #3 (multiple choice): "If *n* represents an even number greater than 2, what is the next larger even number?"

Response Options
$n + 1$
$2n + 1$
$2n$
$n + 2$
n^2

Item #4 (multiple choice): "Which of the following is always an odd integer?"

Response Options
The product of two odd integers
The product of two consecutive integers
The sum of three even integers
The sum of two odd integers
The sum of three consecutive integers

Item #5 (open response, bounded): "If the product of 6 integers is negative, at most how many of the integers can be negative?"

Response Options
0
1
2
3
4
5
6

Responses for Item #6 (multiple choice): "Which expression is the greatest when *n* is a negative number?"

Response Options
$n - 2$
$2n$
n^2
$n/2$
$2/n$

Item #7 (multiple choice): "A car can seat *c* adults. A van can seat 4 more than twice as many adults as the car can. In terms of *c*, how many adults can the van seat?"

Response Options
$c + 8$
$c + 12$
$2c - 4$
$2c + 4$
$4c + 2$

Item #8 (multiple choice): "Each of the 18 students in Mr. Hall's class has *p* pencils. Which expression represents the total number of pencil's that Mr. Hall's class has?"

Response Options
$18 + p$
$18 - p$
$18 * p$
$18 / p$

Note: Overall n = 78.

Item #9 (multiple choice):

□	△
4	9
5	11
6	13
7	15

Which rule describes the pattern shown in the table?

A. □ + 5 = △
B. □ + □ = △
C. □ + □ + 1 = △
D. □ + □ + 2 = △

Correct answer: C

Appendix E: Student Survey Reliability testing

We ran reliability testing on the current pre-test and the fall 2020 pilot sample to compare to the research done on the Student Computer Science Attitude Survey.[3] This survey was tested and validated across 2010–16 for grade 8+ students for measuring five attitudinal constructs related to computer science. Since the survey was originally validated against mostly high school students, we wanted to check reliability against the study's population of middle schoolers during piloting year before using this survey for the implementation study. We sought to answer three main questions in the pilot:

1. Are the survey factors still reliable with our intended student population?
2. Do students appear to meaningfully respond to the survey?
3. Do we have significant risk of response ceilings or floors such that intervention impacts would be difficult to determine?

The factor structure of the original survey was broadly maintained in the pilot administration. Three of the five factors had very good reliability with our student population, one had good reliability, and one had minimally acceptable reliability.[4] Our reliability metrics for most factors were slightly below those of the original research (see Table below), which may be partially due to the differences in sample sizes. We consider only one factor, Belongingness, to be potentially problematic for our study. Our reliability calculation was barely acceptable (0.650) and significantly below that of the original research (0.850). We addressed our piloting results with the original survey author, and we determined that Belongingness may not be a meaningful construct for our study given the age group of our students and the design of CPR2. We decided not to edit the survey (i.e., in the same form used in prior validation research by its creator) or remove the Belongingness survey items. While removing these items would be unlikely to affect students' responses, we decided to err on the side of caution.

During the fall 2021 administration, we again ran reliability testing with this larger middle school sample to confirm the four factors we found reliable in the pilot study performed well with the larger sample. We again found the Confidence, Interest, Encouragement, and Usefulness factors had either very good reliability or good reliability. This further confirmed this is suitable instrument for this group of students. While Belongingness did barely achieve good reliability with the larger sample (0.703), we decided we would need to cautiously analyze these data when impact analyses are conducted given the lack of reliability in our pilot and the survey author's sense that this may not be meaningful for our age group or for the CPR2 intervention.

---

[3] Haynie, K.C. and Packman, S. (2017). *AP CS Principles Phase II: Broadening Participation in Computer Science Final Evaluation Report*. Prepared for The College Board and the National Science Foundation, February 12, 2017. Skillman, NJ.

[4] We consider Cronbach's alphas of 0.80+ to indicate very good reliability, 0.70+ good reliability, and 0.60+ minimally acceptable reliability per Nunally, J.C. (1967). *Psychometric Theory*. New York: McGraw-Hill.

Reliability Results

Psychometrics		Confidence	Interest	Belongingness	Usefulness	Encouragement
Fall 2021 Pretest	Cronbach's alpha	0.766	0.843	0.703	0.815	0.813
	n	843	840	843	839	841
Pilot Sample	Cronbach's alpha	0.850	0.892	0.650	0.853	0.778
	n	149	149	149	149	149
Research Reference	Cronbach's alpha	0.890	0.932	0.850	0.892	0.858
	n	802	802	803	802	805

Note: Factors were only assessed for a student if they replied to all five items for a factor, which is why the *n*'s may vary across factors.

Appendix F: Student Survey Items by Construct

Construct	Items (1-4 scale; 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree)
Factor 1: Confidence	I am sure I could do advanced work in computer programming.
	I have self-confidence when it comes to computer programming.
	I am confident that I can solve problems by using computing.
	I can learn computer programming without a teacher to explain it.
	I think I will do well in computer programming.
Factor 2: Interest	I like writing computer programs.
	I like to use computer programming to solve problems.
	The challenge of solving problems using computer programming appeals to me.
	I would take additional computer programming courses if I were given the opportunity.
	I hope that my future career will require the use of computer programming.
Factor 3: Belongingness	I feel I belong in computer programming.
	I feel comfortable in computer programming.
	I feel accepted by my peers in computer programming.
	I know a lot of students like me who are interested in computer programming.
	I know someone like me who uses computer programming in their work.
Factor 4: Usefulness	Skills used to understand computer science material can be helpful to me in understanding things in everyday life.
	Computer programming is a worthwhile and necessary subject.
	Knowledge of computer programming will help me earn a living.
	Learning to use computing skills will help me achieve my career goals
	I'll need a mastery of computer programming for my future work.
Factor 5: Encouragement	A friend or peer has encouraged me to study computer programming.
	Someone in my family has encouraged me to study computer programming.
	Someone I know has discussed with me the computer programming field.
	Someone I know has given me the desire to study computer programming.
	Someone I know has given me the desire to study computer programming.
	Someone I know has praised my work in computer programming.