2022

Impact of Cognitively Guided Instruction on Elementary School Mathematics Achievement: Five Years After the Initial Opportunity

Robert C Schoen, Christopher Rhoads, Alexandra Lane Perez, Amanda M Tazaz and Walter G Secada

Impact of Cognitively Guided Instruction on Elementary School Mathematics Achievement: Five Years After the Initial Opportunity

Robert C. Schoen[1,2]
Christopher Rhoads[3]
Alexandra Lane Perez[3]
Amanda M. Tazaz[1]
Walter G. Secada[4]

May 2022

[1]Learning Systems Institute; Florida State University
[2]School of Teacher Education, Florida State University
[3]Neag School of Education, Dept. of Educational Psychology
[4]School of Education and Human Development; University of Miami

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant number R305A180429 to Florida State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Suggested Citation

Schoen, R. C., Rhoads, C., Perez, A. L., Tazaz, A. M., & Secada, W. G. (2022). *Impact of Cognitively Guided Instruction on Elementary School Mathematics Achievement: Five Years After the Initial Opportunity*. Florida State University. Working Paper. https://doi.org/10.33009/fsu.1653430141

Correspondence regarding this paper can be sent to Robert Schoen at rschoen@lsi.fsu.edu.

Abstract

We studied the impact of a long-term teacher professional development program on elementary school mathematics achievement five years after the initial randomization using an intent-to-treat approach and all available achievement data for kindergarten through fifth-grade students. The intervention consisted of a randomized offer for teachers in 22 schools to participate in a professional-development program based on Cognitively Guided Instruction. The intervention had a small positive effect ($g = 0.03$) on mathematics achievement in the primary grades and a larger effect ($g = 0.16$) in the intermediate grades. Grade level was the only statistically significant moderator, with larger effects in higher grade levels. These results provide new evidence of a long-term effect of Cognitively Guided Instruction on student learning in mathematics.

Impact of Cognitively Guided Instruction on Elementary School Mathematics Achievement: Five Years After the Initial Opportunity

Billions of dollars are spent in the U.S. every year on teacher professional development (PD) programs (Fermanich, 2002; Odden et al., 2002; TNTP, 2015; U.S. Department of Education, 2014). Despite that massive investment, rigorous evaluation of the impact of those dollars on the various outcomes of interest has historically been uncommon, but the number of studies of teacher PD that employ experimental or strong quasi-experimental designs has increased over the past decade.

Referencing studies that do employ rigorous evaluation designs, researchers have drawn mixed conclusions about whether teacher PD programs increase student learning (Garet et al., 2016a; Gersten et al., 2014; Lynch et al., 2019; Pellegrini et al., 2021). Studying three different teacher PD programs in reading and mathematics, all three of which focused on increasing teacher content knowledge, Garet et al. (2016a) concluded that developers of PD programs seem to know how to design and implement PD programs that impact teacher knowledge, but "the field does not yet fully understand how to ensure that teacher PD leads to measurable improvements in student learning" (p. 11).

Having read this powerful statement by Garet et al., a reasonable person might ask why researchers should bother to continue to research whether teacher PD programs can improve student outcomes—especially teacher PD programs that focus on increasing teacher content knowledge. One reason to question that conclusion is the tendency of research studies to emphasize detecting effects on student achievement in the short-term (i.e., within weeks or months). We observe that the overwhelming majority of RCTs intended to evaluate the effect of teacher PD programs are designed to detect effects on students within the first year of the intervention. Fewer studies are designed to follow up on student learning in a second year, and fewer yet are designed to follow up after multiple years. It is possible, therefore, that effects on students could materialize after the researchers have stopped looking for them. This may be particularly true of teacher PD programs that are non-prescriptive in nature, address a broad range of curricular outcomes that are also of central importance to those in the counterfactual condition, or attempt to tackle persistent or particularly difficult problems of practice.

Evidence of Effectiveness of Teacher Professional Development Programs Based on Cognitively Guided Instruction

Mathematics teacher professional development (PD) programs based on Cognitively Guided Instruction (CGI) have been the subject of dozens of studies using qualitative, quantitative, and mixed-methods research designs. CGI ranks among the few mathematics PD programs that have been found to have a potentially positive impact on student learning. Several experimental studies of CGI PD programs have reported positive effects on students in the first year of implementation (Carpenter et al., 1989; Jacobs et al., 2007; Schoen et al., 2018a; under review), and one study reported mixed effects on students (Schoen et al., 2020).[1]

All five of the extant RCTs of CGI examined the effect of the program on student mathematics achievement in the first year of implementation. Consequently, none of those five studies can determine whether longer-term effects on students occurred. While some of those studies used standardized tests, such as the Iowa Test of Basic Skills (ITBS) and the Elementary Mathematics Student Assessment (EMSA; Schoen et al., 2017; 2018b), none of them used the

---

[1] Looking at effects of the program in the first year of implementation, Schoen et al. (2020) reported potentially positive effects on first-grade students' problem-solving abilities and potentially negative effects on second-grade students' computational abilities, although none of those effects were statistically significant when using the customary threshold of p=.05.

standardized or state-standardized tests in use by the participating schools to measure student achievement in mathematics at the time.

Schoen et al. (2020) randomly assigned 22 elementary schools in spring 2013 to a CGI condition or a comparison condition. For the subsequent five years, the teachers in the 11 schools assigned to the CGI condition had the opportunity to participate in a CGI PD program. This situation created a unique opportunity to study the longer-term impact of the CGI PD program on student achievement—a situation without precedent in the research literature.

Purpose of the Study and Associated Research Questions

The current study leverages the school-level randomization that occurred in spring 2013 to estimate the impact of the CGI intervention on school mathematics achievement after five years. The effect of CGI on school mathematics achievement in the fifth school year after randomization is the focus of the present study for several reasons. First, the 2017-2018 school year represented the fifth and final year that teachers in the CGI-condition schools had the opportunity to participate in the CGI program (Schoen et al., 2022). It also happened to be the first year that both school districts used the same progress-monitoring test for student mathematics achievement (i.e., i-Ready Diagnostic), which provided a common metric in all six grade levels across the full sample for the first time in 2017-2018. We note that the first-grade students in 2013-2014 were fifth graders (i.e., the highest grade level in the schools) in 2017-2018. Thus, another phenomenon occurred for the first time in that year, because that was the first school year when every student in those 11 schools could have had at least some opportunity to have been taught by a mathematics teacher who had participated in the CGI program. This context motivated an intent-to-treat study of the effects of school-level randomization to the opportunity to participate in the CGI PD program on the mathematics achievement of *all* students enrolled in participating schools during the 2017-2018 school year.

There are at least three distinct explanations for why so many previous studies of teacher PD have not found the interventions focused on teacher knowledge to have positive effects on student achievement. The first—and simplest—is that such effects do not exist. The second is that teachers require some time to adjust their teaching to be consistent with the PD they have just received, so the improvement in teaching practice may not appear until a few years after PD receipt. The third recognizes that, if exposure to a CGI-trained teacher does have a beneficial effect on the mathematics achievement of student~~s~~, it is unclear when we should expect that effect to be observable on standardized mathematics tests. We conjecture that it is possible that the beneficial effects of CGI could remain latent for some time and will only be revealed as students are exposed to more difficult mathematical content in later grades. Collectively, these factors provide a rationale for the current study, which explores the impact on student achievement across all the elementary grades five years after the initial randomization.

The present study investigates the effect of the opportunity for teachers to participate in the multiyear CGI PD program on student learning. Specifically, we estimate the impact of school-level randomization on school mathematics achievement after five continuous years of the availability of the opportunity to participate (in the schools randomized to the CGI condition). We define the intervention as the opportunity to participate, because all teachers were invited to participate, not all teachers did participate, and we are interested in the effect on all students in the school—whether or not their teacher(s) participated in the CGI program. The study was guided by the following research questions (RQs). We consider RQ1 and RQ2 to be confirmatory and RQ3 to be exploratory.

RQ1: What is effect of the opportunity for elementary mathematics teachers to participate in CGI professional development on early elementary (grades K–2) student achievement during the 2017-2018 school year as measured by the i-Ready Diagnostic?

RQ2: What is effect of the opportunity for elementary mathematics teachers to participate in CGI professional development on upper elementary (grades 3–5) student achievement during the 2017-2018 school year as measured by the i-Ready Diagnostic and the Florida Standards Assessment (FSA)?

RQ3: To what extent do the following variables moderate the impact of the opportunity for elementary mathematics teachers to participate in CGI professional development on student achievement as measured by the i-Ready and FSA exams: grade level, race/ethnicity, disability status, English-learner status, gifted status, and free and/or reduced price lunch status?

We hypothesized that the program would have a positive effect on student achievement on both the i-Ready and the FSA at all grade levels. We explored moderation with respect to the student demographic variables available in the dataset. We did not have specific, *a priori* hypotheses about how the CGI treatment would interact with these student level characteristics, although it seemed plausible that treatment effects might vary with respect to any of the variables listed, and we note that a previous study (Schoen et al., 2020) reported differences in effects by grade level.

Professional Development Based on Cognitively Guided Instruction

The first CGI PD program was created in the mid-1980s. That CGI PD program was designed to inform teachers about a robust set of frameworks for problem types and strategies that had recently crystallized following several decades of research on addition and subtraction problem types and associated student strategies (Carpenter & Moser, 1983; Fuson, 1992; Schoen et al., in-press). By that time, research had demonstrated that many children can solve word problems with understanding without first experiencing explicit instruction on how to do so (Baroody, 1987; Baroody & Ginsburg 1986). Researchers observerd that students can solve these problems by modeling the action and relations in stories told through the word problems, and that they use a predictable and sensible progression of strategies as their understanding and abilities in mathematics became more sophisticated (Carpenter et al., 1993; 1998). By the 2010s, almost every state in the U.S. had adopted mathematics curriculum standards that reflected the problem-types frameworks for word problems involving all four of the basic operations that had crystallized in the research literature during the 1980s and 1990s (NGA & CCSSO, 2010; Schoen et al., 2021) and that CGI programs and publications had been introducing to teachers since the mid-1980s.

Research predating the development of CGI PD programs indicates that experienced teachers have a considerable amount of knowledge about children's mathematical thinking, but that knowledge is often fragmented and disorganized, thereby limiting its use in teachers' decision making (Carpenter et al., 1988). That original CGI program set out to support teachers with building and organizing their understanding of children's mathematical thinking by introducing teachers to research-based taxonomies for types of word problems and progressions of associated strategies students use for solving those problems (Carpenter et al., 1989; Fennema et al., 1996). But the acquisition of knowledge of the formalized, research-based frameworks in CGI PD is not viewed as an end in and of itself. Rather, the frameworks serve as a lens to support teachers' interpretation and cultivation of instructional practices that build on their own students' thinking. Carpenter and Franke (2004) asserted that fundamental changes in teacher practice can result from understanding and building upon students' mathematical thinking.

Consequently, CGI PD aims to support teachers with organizing and building on their informal knowledge of children's thinking in specific content domains to construct and test models of student thinking such that they can use these models and information about their students' thinking processes to make instructional decisions. The goal of CGI was (and is) to stimulate teachers' engagement in practical inquiry that leads to better outcomes for students and generative learning for teachers (Franke et al., 1998).

CGI PD does not provide teachers with a curriculum to follow or prescribe how teachers should implement what they learn through CGI PD in the classroom. Fennema et al. (1996) described their approach as follows:

> Teachers had to decide how to consider students as they selected problems, how to question children, and how to organize their classrooms. In order to do this, they had to reflect on what the research-based model of children's thinking meant for their classroom with their own students. This was not easy. The teacher had to deal with the complexity of children's problem solving as well as the myriad of other factors that are always present. But by doing so, the teachers transformed the model and it became part of their knowledge. (p. 432)

Through their study of the initial models of CGI PD, Carpenter, Fennema, Peterson, and colleagues reported positive results and far-reaching potential of the CGI program for improving the teaching and learning of mathematics (Carpenter et al., 1989; 1996; Fennema et al., 1996). Subsequently, CGI has been embraced by many in the mathematics education community. More than 200,000 copies of the definitive CGI book (Carpenter et al., 1999; 2015) have been sold in the past two decades, and tens of thousands of prospective and practicing teachers have participated in CGI-based PD programs (Philipp et al., 2009; Schoen et al., 2022; Secada & Brendefur, 2000).

Many different models of CGI PD have subsequently appeared in the past four decades. The original CGI PD program only involved first-grade teachers and focused on addition and subtraction on whole numbers for a period of one school year (Carpenter et al., 1989). Subsequent CGI PD efforts have supported teachers at all elementary grade levels, and the focus of the subject matter has expanded to include a wider range of topics, including multi-digit addition and subtraction, single digit multiplication and division, base-ten concepts (Carpenter et al., 1999; 2015), algebraic reasoning (Carpenter et al., 2003), multi-digit multiplication and division, and fractions and decimals concepts and operations (Empson & Levi, 2011). While some core features of CGI and CGI PD tend to be consistent across different CGI-based programs, the particulars of the design and implementation of CGI PD programs—such as the structure and duration, mathematics content, role and qualifications of the learning leader(s), target participants, and focal activities—vary substantially. Fennema et al. (1999) published a Guide for CGI Workshop Leaders but acknowledged CGI teacher PD programs employ a variety of approaches, such as workshops (e.g., Carpenter et al., 1989; Fennema et al., 1996; Moscardini, 2014; Schoen et al., 2018a; 2020), teacher work groups (e.g., Franke & Kazemi, 2001), and blends of these approaches (Jacobs et al., 2007).

Description of the CGI PD Program in the Current Study

The CGI PD program that was the subject of the present study was designed and delivered by Teachers Development Group under the direction of Linda Levi. Dr. Levi was one of the coauthors of the three definitive CGI books (Carpenter et al., 1999; Carpenter et al., 2003; Empson & Levi, 2011) and served in the role of the director of CGI PD initiatives. Schoen et al. (2022) provided a thorough description of the intervention. A few select highlights are provided in the following paragraphs.

The design of the CGI PD program provided three years of support for each teacher, with 8 days of teacher workshops offered each year. The same number of days repeated for each of the three years of the program, resulting in 24 planned days spaced across three years. Each workshop day involves approximately seven hours of direct contact time. Four (and sometimes five) of the days per year of workshops were offered in the summer, and a series of two two-day follow-up sessions were offered in the fall and winter. The multi-year program extended the opportunities for teachers to continue to develop their understanding, beliefs, and instructional practice.

There were two distinct tracks in the program for teachers: K–2, or 3–5. Each track was designed to provide three years of support. As teachers revisited a given topic in subsequent sessions and years, they focused on the topic from a more advanced perspective that was meant to facilitate a deeper understanding. Because the substantive content each year built on what teachers learned in the preceding years, a teacher must have completed the first year of the program to become eligible for the second year and completed the second year of the program to be eligible for the third. Completion of the program in three consecutive years was recommended, but it was not required. Under normal conditions, completion of the program in three consecutive years is rarely achieved in practice due to discontinuity in funding, scheduling conflicts, or changes in teaching assignments. Teachers may participate in both of the (K–2 and 3–5) programs simultaneously, but it is not recommended for classroom teachers, because that would require them to be out of their classroom for as many as 8 days during the school year.

Using the CCSSM as a common touchstone for describing the focal topics in mathematics, the K–2 program focused on content in the domains of: operations and algebraic thinking, counting and cardinality, and number and operations in base ten. The content of the

CGI K–2 program aligned with the state curriculum standards for grades K–2 for the most part, but it also extended beyond the grade-level expectations in the state curriculum standards to address frameworks for story problems involving multiplication and division as well as the corresponding frameworks for associated student strategies. The grades 3–5 program focuses on the domains of: operations and algebraic thinking, number and base ten, and fractions. As such, the content coverage in each track focused on the mainstay of the elementary math curriculum (i.e., number and operations). The CGI program did not directly address domains such as geometry, measurement, data analysis, or probability.

The CGI PD program did not provide curriculum materials and was not prescriptive regarding how teachers should teach. Research-based frameworks for problem types and strategies are well-known salient features of CGI in general, and they were used explicitly throughout this CGI program. The CGI program provided teachers with opportunities to learn about robust and predictable developmental progressions that described how children's knowledge and understanding of mathematics becomes more sophisticated over time. The workshops also provided support for teachers to learn a well-developed professional vernacular for describing these progressions and the underlying mathematics.

Interactions with children around mathematics were an integral part of the CGI PD. In-person interviews with children were a key component in the summer sessions. Each of the two-day follow-up sessions included a classroom-embedded workshop day (Levi, 2017; Nielsen et al., 2016; Schoen & Champagne, 2017; Schoen et al., 2022) as well as introduction to new content. The lessons implemented during classroom-embedded days were structured around the instructional design model described by Smith and Stein (2011), although that connection was not made explicit to the participating educators.

Whereas the CGI PD program consisted mostly of workshops that focused on content (e.g., frameworks for problem types and strategies, mathematics language and content, curriculum standards), it did provide some process-oriented support for participating teachers through the classroom-embedded workshop days, introduction to the purposeful pedagogy model (Jaslow & Evans, 2012), and the provision of and practice using protocols for interviewing students to gather information about their mathematical understanding—including fact fluency. A small subset of grade-level teams also received in-school support in the form of Formative Assessment Collaborative Team meetings during the 2015-2016 school year (Bauduin et al., 2016; Bray et al., 2019).

Theory of Change for the CGI PD Program

The thesis of CGI is that children have experiential knowledge and informal knowledge of mathematics that can serve as the basis for developing a more formal understanding of the elementary mathematics curriculum (Carpenter et al., 1996; 1999; 2015). Children can build on their intuitive knowledge to develop progressively abstract understanding of and formalized strategies for addition, subtraction, multiplication, and division with single-digit whole numbers, multi-digit whole numbers and fractions, as well as the base-ten number system and fractions concepts (Carpenter et al., 1996). A CGI approach to teaching is mindful that children often view mathematics differently than adults do and that striving to understand the child’s perspective is an important part of teaching. Within this conceptual framework, a teacher’s role is to design and implement instruction in a way that leverages and elevates children’s ways of knowing and understanding such that they are used as a foundation for building new knowledge. There are myriad possible ways these broad principles can manifest in practice. In this section, we describe some of the key features of the CGI program and describe how those may ultimately—and indirectly—improve student outcomes.

Parallel to CGI's foundational assumptions (as described above) about children's abilities to construct knowledge of mathematics, CGI PD for practicing teachers aims to support teacher learning by activating and building on teachers' existing knowledge of children's thinking (Carpenter et al., 1996). CGI PD conceptualizes teaching as a problem-solving endeavor (Carpenter, 1989), in which teachers can use information gained by attending to the mathematical thinking of their students to further refine their knowledge of children's thinking and how instruction can be designed to support its further development (Franke et al., 1998, 2001). A teacher in the CGI program must do the work of determining how to organize classroom instruction in a way that makes sense in their own context. CGI teachers learn to select or create mathematical tasks that expose student thinking in relation to various learning goals, and they become increasingly adept at being responsive to student understanding and using students' ideas to support student learning.

Figure 1 depicts the program theory of change for the model of CGI PD that was designed and implemented in the current study. This theory of change implies that the only direct effects of the PD program are on the teacher, and the teacher mediates the subsequent effects of the program on instruction and student learning. The figure also implies that teacher experimentation in the classroom and teacher observation of changes in student learning can also mediate the impact of the program on teacher knowledge and beliefs and can even feed back into the substantive content of the PD workshops. This theory of change also acknowledges some of the contextual factors that may enhance or impede teachers' learning, changes to instructional practice, and student learning.

According to the theory of change, effects on teachers, teaching, and students occur through an iterative process over an extended period, both within and across school years. The teacher-change process occurs in an iterative manner over an extended duration of formal and informal experiences. The iterative back-and-forth between workshop and school-based experiences provides a supportive structure for implementation of and experimentation with new ideas. It also creates opportunities for teachers to situate their learning into their own practice. As a result, participating teachers play an active role in creating coherence between their daily work and the ideas they encounter in CGI PD. This dynamic is thought to allow for the changes in knowledge and beliefs that may occur through participation in the workshops to transfer into long-term, significant changes in instructional practice.

The CGI program increases teachers' mathematical knowledge for teaching and changes their beliefs about mathematics teaching and learning (Carpenter et al., 1996; Schoen & LaVenia, 2019; Schoen et al., 2017b; 2019; 2022). Teacher knowledge of mathematics increases in this CGI program through in-depth study, analysis, and discussion of observable features of student thinking—not necessarily by solving mathematics problems that are difficult for the teachers to solve. That mechanism differs from other programs that increase teacher knowledge through teachers solving mathematics problems at the boundaries of their own mathematics ability. Because the workshops involve extensive opportunity for teacher discussion about student thinking and mathematics concepts, the CGI PD program also provides teachers with opportunity to develop their professional vernacular and improve their ability to communicate effectively and efficiently with other educators about mathematics teaching and learning.

Interaction with children in a mathematics teaching and learning setting occurs both within and between the workshops. During the summer workshops in the first year of the program, participating teachers interview children soon after they have been introduced to a framework of solution strategies children typically use to solve various types of problems. Interviews with children during workshops serve multiple purposes. The interviews in that first set of workshops offer the participating teachers an opportunity to consider how their experiences with the children they interview compare to what they observed in the videos of children solving problems. Participating teachers are often surprised to find that the children they interview are solving problems in similar ways to those of the children in the videos (which might have been staged) and that the young children they interview can solve various types of word problems that they haven't been previously told how to solve. In this way, these initial experiences interviewing children reinforce the validity of the research-based frameworks, and in doing so, also often spark curiosity and interest in learning more about their own students and about the implications of this newfound perspective on children's thinking and abilities. That surprise and curiosity helps to maintain motivation to continue to learn about the thinking processes of individual children. The interviews also create opportunity for teachers practice just observing and learning about children without feeling compelled to instruct. Classroom interactions with students between workshops is also an integral part of the learning experience as teachers experiment with their new perspective on students and teaching and their curriculum. These experiences, which confirm the validity and potential usefulness of the research-based frameworks for many teachers, also catalyze changes in teacher beliefs about mathematics teaching and learning (Peterson et al., 1989; Schoen et al., 2015; manuscript under review).

The classroom-embedded workshop days provide an opportunity for teachers to see how the ideas they encounter in the workshops may be implemented in practice. With guidance from the workshop leaders, classroom-embedded workshops involve teachers in the following processes: formative assessment (with an emphasis on strategies students use to solve mathematics problems), setting near-term and long-term goals for students, planning for instruction in ways that build on student understanding and leverage variation in ways students solving problems to support differentiation of instruction and advance mathematical understanding during a mathematics lesson, teaching a lesson or observing the lesson being taught by another teacher, and reflecting on the lesson (Schoen & Champagne, 2017). The classroom-embedded workshop days deliberately slow down the practice of teaching. This allows teachers time to share ideas about how to put the ideas they have encountered in the workshops into action. This includes the opportunity to implement instruction that is responsive to student thinking and reflect on those experiences in a group setting. Classroom-embedded days provide the opportunity for teachers to engage in the process of implementation of mathematics instruction that is cognitively guided and is designed and implemented in a manner that is consistent with the thesis that teaching is a problem-solving activity.

CGI-aligned mathematics instructional practice centers on children's ways of thinking and reasoning about mathematics. CGI emphasizes mathematics as a sensemaking activity and advocates for a corresponding bottom-up (Hiebert & Carpenter, 1992) approach to teaching and learning. Rather than focusing attention on those things students do not know or things they cannot do—an orientation that privileges a deficit perspective on teaching and learning—CGI focuses on knowledge and skills that students do have and builds toward more sophisticated knowledge or abilities. In contrast with the deficit perspective, this latter perspective views all understandings—from the least sophisticated to the most sophisticated—as partial understandings and can be described as an asset-oriented approach to teaching (Hunt et al., 2020). In this learning environment, students engage in self- and peer-assessment as they express their ideas verbally and in writing, engage in mathematical modeling, study their peers' solutions to problems, and struggle to understand each other's ideas about mathematics.

In a CGI classroom, variation in student abilities is viewed as an asset, not a limitation. Variation in student thinking creates opportunity for meaningful discussion among members of the class, and classroom-embedded days provide teachers with opportunities to learn how to use this variation to improve teaching and learning. Students using more abstract strategies can learn from those who are using more concrete or literal strategies, and those who are using less sophisticated strategies can learn from those who are using more sophisticated strategies. Students often present their own worked examples for class to review and discuss during the resulting lessons.

In a CGI classroom, mathematics is viewed as a human endeavor. Students in CGI classrooms learn that meaningful use of symbols, words, and procedures is expected, and their voice and perspective is valuable and important. Each child is a contributor to the mathematics community in the classroom, and their ideas constitute useful contributions to the learning in that community. Students actively construct knowledge by making sense of problems, engaging in meaningful use of mathematical language, and analyzing solutions to problems that are offered by their peers.

The culture of the learning community in a CGI classroom emphasizes learning mathematics with understanding (Hiebert & Carpenter, 1992). Understanding of addition and subtraction, for example, is a natural extension of the more basic understanding of counting and whole number. Student understanding of mathematics concepts, such as the laws of operations, relations between numbers and operations, place-value concepts, etc., grows through the processes of problem solving, communicating, sensemaking, and reflection. In this constructivist-oriented, problem-solving-as-modeling (Carpenter et al., 1993) approach to teaching, ability to recall number facts and use them meaningfully occurs through students solving many problems with understanding—not by first expecting students to memorize a prescribed list of facts involving number and operations. Teachers look for opportunities to help students be aware of the algebraic foundations of elementary arithmetic, such as the commutative property of addition or the distributive property of multiplication over addition. These laws of operations are sometimes understood intuitively by students, such as when students notice that sum is the same regardless of the order in which the addition operation is applied on two numbers. The CGI teacher looks for ways to help students to formalize their intuitive understandings of these laws, which may serve to create a firm foundation for their future learning of algebra and other higher-level mathematics.

In summary, CGI PD is firmly rooted in constructivism, an asset-oriented approach to teaching, and a specific approach to engaging in ongoing, classroom-based, formative assessment in that teachers focus on the observable strategies that students use to solve problems and use this information to draw inference about student understanding and make decisions about next steps in instruction that will serve to advance their students' understanding of mathematics. This approach expects that deep and meaningful learning requires considerable effort over an extended period, and the CGI program developers maintain that this is true for students and teachers alike.

Methods

The current RCT was conducted in two Florida school districts with schools, teachers, and students participating voluntarily. The study used a matched-pair design where, within district, 22 schools were matched on the percentage of their students eligible for free or reduced lunch. School-level randomization occurred within these matched pairs in spring 2013 as described in Schoen et al. (2020). Schools assigned to the CGI PD condition, and teachers and students in those schools, are hereafter referred to as CGI schools, teachers, and students. Schools *not* assigned to the CGI PD condition in spring 2013, and teachers and students in those schools, are hereafter referred to as comparison schools, teachers, and students. The current study addresses the impact of school-level randomization on *all* K–5 students in the participating schools five years after the initial opportunity to participate in the CGI PD.

During the five-year period following the initial randomization, many educators were reassigned to teach at a different grade level, and some are no longer teaching in a participating school. However, many of those who changed grade levels did remain in the same school. Additionally, the participating school districts were able to provide the research team with demographic and mathematics achievement data for students enrolled in participating schools during the 2017-2018 school year. They were not, however, able to provide information linking students to their respective teachers. Because randomization occurred at the school level, all students in these schools (grades K–5) may have had the opportunity to have been taught by a teacher in the CGI program, mathematics assessment data were available for virtually all students in the schools, and teacher-student links are not available for the fifth year of the study, we take an intent-to-treat approach to this study.

Setting and Sample

The study occurred in 22 public schools in two adjacent school districts in Florida. Participation was neither required nor funded by their schools or school districts. During the first two years after randomization, eligibility was limited to grades 1 and 2 teachers and instructional support personnel, including math coaches, interventionists, and principals. Only the CGI PD program targeted towards K–2 teachers was offered during these two years. During the subsequent three years, all personnel in the 22 schools became eligible to participate (including personnel in schools randomized to the comparison condition). As shown in Table 1, many more educators in the schools originally randomized to the CGI condition took advantage of this opportunity across the five-year period. As of fall 2017, a total of 182 non-overlapping educators in the CGI-condition schools had directly participated in the CGI program (i.e., attended workshops). Only 34 teachers in the comparison-condition schools had similarly participated.

Participants

Schools were the unit of randomization of the offer to participate in the CGI PD intervention. The current study focuses on the impact of the intervention on student achievement in mathematics for all students in those schools in grades K–5 during the 2017-2018 school year, which was five years after the initial randomization of schools.

Research Design

This study used a blocked, school-randomized design with 22 schools assigned to the CGI or comparison condition during spring 2013. As noted previously, the study used a matched-pair design where, within district, schools were matched on the basis of the percentage of their students eligible for free or reduced lunch. Data analysis used all available data for students in randomized schools during the 2017-2018 school year.

Data Collection and Measures

During the 2017-2018 school year, both districts administered the i-Ready diagnostic assessment three times per year (roughly in the beginning, middle and end of the school year) to all students in grades K–5. Additionally, students in grades 3–5 took the Florida Standards Assessment (FSA) in Spring 2018. These two assessments are the outcome measures for our study. Across all grade levels, there was at least one valid mathematics achievement test score for 11,226 students within 14 schools in District 1 and 5,693 students within 7 schools in District 2. One school in the comparison condition was closed before the final year of the study, so that school could not be included in the analytic sample for the 2017-2018 school year. Table 2 provides a description of the students in the sample by condition. Table 3 breaks down the data provided by the district by grade level and treatment condition for the full sample. Table 4 provides a further breakdown by wave of data collection for each mathematics test.

In addition to the design variables (i.e., treatment condition, blocks, school of enrollment), we use student demographic and assessment data in our statistical models. The two partner school districts provided de-identified student-level demographic (e.g., grade level, exceptionality, language proficiency, race/ethnicity, free- and/or reduced-price lunch) and mathematics achievement data for all students in participating schools for the years between 2012 and 2018.

The Florida Standards Assessment (FSA) was the high-stakes, state-standardized test administered to all Florida students in grades 3–8 near the end of the year (FDOE, 2018). Starting in spring 2018, the FSA used a computer-based, online, fixed-form testing platform for grades 3–5. FSA is not administered to students in grades K–2. All schools in the study were required to administer the FSA every year as the state-standardized mathematics assessment in the state accountability system. The school accountability system was designed to align with the curriculum standards that were adopted by the state Board of Education. The state adopted the Common Core State Standards for Mathematics (CCSSM; NGA & CCSSO, 2010) in 2010 and replaced those standards with an almost identical set of elementary mathematics curriculum standards in 2014 called the Mathematics Florida Standards (MAFS). The FSA was designed to align with the MAFS. Using 2-parameter and 3-parameter logistic models based on item-response theory, the FSA yielded a within-grade score and a vertically scaled score using a second-order factor structure with an overall math score and subscale scores for the lower-order factors. The current study focuses on the vertically scaled score for the higher-order factor. Based on the full sample of over 200,000 students per grade level, the Florida Department of Education reported marginal reliability (Samejima, 1977; Sireci et al., 1991) of .93 for grades the grades 3 and 5 tests and .92 for the grade 4 tests (Florida Department of Education, 2018).

Both districts also administered the i-Ready Diagnostic (Curriculum Associates, 2018) three times per year (roughly in the beginning, middle and end of the school year) to virtually all students in grades K–5 during 2017–2018. The i-Ready Diagnostic also provided an overall mathematics achievement score as well as four subscale scores. The current study focused on the overall score, which was also vertically scaled across grade levels. Test scores were generated using models based on item-response theory, and the publisher of the i-Ready Diagnostic test reported marginal reliability estimates of .92 in grade K, .94 in grades 1 and 2, .95 in grade 3, and .96 in grades 4 and 5 based on a fall 2016 administration of the test with their own sample. The i-Ready Diagnostic was not designed to be a high-stakes measure of school accountability, but part of its appeal to school administrators is its strong predictive power for the spring administration of the FSA. The third-grade i-Ready Diagnostic mathematics test was an approved alternative to the FSA for promotion from third grade to fourth grade at the time of the study. During the 2018-2019 school year, 57 of the 67 public school districts in Florida used the i-Ready Diagnostic test (D. Chinn, personal communication, September 20, 2018), but the use of i-Ready is not limited to Florida schools. According to a press release in February 2017, Curriculum Associates—the publisher of the i-Ready Diagnostic—reported more than 4.2

million grades K–8 student users during the 2017-2018 school year, representing all 50 states and 10% of the students in the U.S.

Description of the Comparison Condition

In first year of the study, we offered an alternative intervention to the teachers in schools that were assigned to the comparison condition. Leaders in the curriculum department in District 1 requested a program that they wanted to promote in their schools that they called Bridge to STEM. Bridge to STEM was a local adaptation of a science program that was developed through an NSF-funded project called Ramps and Pathways (Zan & Geiken, 2010). The program was designed to support learning about force and motion in early childhood. The curriculum leaders in District 2 asked for the teachers in the comparison schools to have the option to participate in a program that the district called Data Talks in Science. Many of the teachers in the comparison group for District 1 did participate in summer training for the Bridge to STEM program. They received a set of lesson plans that were designed by teachers in their district and materials necessary to implement in their classrooms (e.g., blocks, ramps, marbles). Teachers were remunerated for their participation in two days of summer workshops to learn how to implement the lessons in Bridge to STEM. None of the teachers in the sample in District 2 elected to participate in the Data Talks in Science program. In subsequent three years, we did not offer an alternative intervention to CGI for the teachers or schools in the comparison condition.

Word problems serve a central role in CGI and were featured prominently in the state-adopted curriculum standards for mathematics during the five-year study period. In fact, the state curriculum standards referenced types of addition, subtraction, multiplication, and division word problems that were nearly identical to the problem-type frameworks that had been an integral part of CGI PD programs since the 1980s and 1990s. In an attempt to better understand how the counterfactual condition might be similar or different in the Carpenter et al. (1989) RCT of CGI and the present study, Schoen et al. (2021) found that first-grade textbooks in the Common Core Era contain many more word problems and substantially more variation in types of word problems than U.S. textbooks in the 1980s did, which corresponds to the period when the initial RCT of CGI was conducted. Schoen et al. (in-press) also found that the overall pattern of word problem difficulty remained similar in the Common Core Era as compared to previous decades and that, on average, the contrast in difficulty between the traditional types of problems and nontraditional types was less in the Common Core Era than it was in the 1980s.

We remind the reader that the teachers in the comparison schools did have the opportunity to participate in the CGI program (both K–2 and 3–5) during the third, fourth, and fifth years of the study. The extent to which they did participate in the program was reported in Table 1. Consistent with our “intent-to-treat” approach to data analysis, teacher participation in PD workshops plays no formal role in our statistical models. Instead, students residing in schools randomized to the CGI condition in Spring 2013 are assumed to have had, on average, greater exposure to CGI-based instruction than students residing in schools randomized to the comparison condition in Spring 2013. The teacher participation numbers reported in Table 1 supports this assertion, and the original randomization ensures that the greater participation did not occur for endogenous reasons. Due to the intent-to-treat approach to data analysis and the known treatment diffusion during the five-year period, it seems reasonable to assume that the estimated average treatment effects are conservative estimates of the true average impacts of school-level CGI exposure.

Data Analysis

Our main data analytic approach involves fitting hierarchical linear models (HLMs) with i-Ready or FSA scores as the dependent variable, a school-level indicator for randomization to the CGI condition as an independent variable, and additional covariates (described below) added as independent variables. When there is only one achievement measurement per student (i.e., when FSA scores are the outcome measure), then the models have two levels with students nested within schools. When there are multiple achievement measurements per student (i.e., when i-Ready scores are the outcome), then the models have three levels with measurement occasions (BOY, MOY, and EOY timepoints) nested within students nested within schools.

To account for the blocked randomization process, all statistical models included dummy variables (fixed effects) for block membership. We also included indicator variables for student grade. Given past research on the importance of covariates, especially pretests, for achieving reasonable levels of statistical power in school randomized studies (Bloom, Richburg-Hayes, & Rebeck Black, 2007), we wished to include additional covariates as independent variables. However, it is well known that conditioning on variables that have themselves been impacted by the intervention can cause bias in analyses of randomized trials (e.g., Montgomery et al., 2018). We decided that the values of demographic variables were unlikely to have been impacted by school-level randomization. Therefore, we also included the following variables in our models: indicators for free or reduced lunch eligibility, English-language learner status, gifted status, disability status, and student race/ethnicity (where White–non-Hispanic was the reference category, and the model included dummy variables for Black-non-Hispanic, Asian, Hispanic and Other). Finally, as an independent variable, we included the school-average for the fall administration of the i-Ready test for Kindergartners. Since Kindergartners had just entered the schools in fall 2017, their scores are unlikely to have been impacted by school's treatment condition. On the other hand, it would not be appropriate to condition on test scores collected from students in other grade levels, because those scores may have been impacted by exposure to CGI trained teachers. Correspondingly, we did not include fall Kindergarten i-Ready scores as an outcome measurement (but we did include fall i-Ready scores in the models for students in other grades).

We provide the model specification for the grade 3–5 model with i-Ready outcomes. The grade K–2 model with i-Ready outcomes is specified in an analogous manner. The grade 3–5 model with FSA outcomes is the same as the model with i-Ready outcomes, only without the residual term for measures nested within students. The grade 3–5 i-Ready outcome model is specified as:

$$Y_{ijk} = \beta_0 + \beta_1(FK_{AVG})_k + \beta_2(CGI)_k + \sum_{l=3}^{13} \beta_l \ (Block_l)_k + \sum_{l=14}^{21} \beta_l \ (X_l)_{jk} + \beta_{22}(Grade4)_{jk} + \beta_{23}(Grade5)_{jk} + v_k + u_{jk} + r_{ijk},$$

where Y_{ijk} represents the i-Ready score for the *i*th measurement occasion for the *j*th student in the *k*th school. $(FK_{AVG})_k$ represents the Fall average i-Ready math score for kindergartners for the *k*th school. $(CGI)_k$ is the indicator variable for randomization status of the kth school (with $(CGI)_k$=1 if the school was randomized to the CGI condition). The $(Block_l)$ variables are the indicator variables for the randomization blocks in the *k*th school. $(X_l)_{jk}$ represents the value of the lth demographic indicator variable for the *j*th student in the kth school. $(Grade4)_{jk}$ and $(Grade5)_{jk}$ are indicator variables equal to 1 is student j in school k was in the 4th grade (or 5th grade, respectively) in 2017-18. Finally, v_k, u_{jk} and r_{ijk} are, respectively, the school, student and measurement level residual error terms.

The HLM models described above were used to estimate overall average treatment effects on i-Ready scores in grades K–2 (research question 1) and on i-Ready scores and FSA scores in grades 3–5 (research question 2). In addition to these mean regressions, we fit linear quantile mixed models using the *lqmm* package in R (Geraci, 2014). These quantile regression models allowed us to explore heterogeneity in treatment effects across different percentiles of the achievement distribution while accounting for the nesting of students in schools (and measurement occasions in students, where applicable). Model specifications were the same ones used for the mean regressions, with the exception that block dummy variables were omitted. The *lqmm* package computes standard errors using a block-bootstrap approach. With only two schools per block, including the block dummies resulted in non-identified models across certain bootstrap replications.

To answer research question 3, we added terms to the models representing an interaction between the variable of interest and the randomization indicator variable. We only explored these interactions using the standard mean regression HLM models. If the moderating variable of interest was binary (gifted, disabled, FRL, or ELL), only a single additional interaction term was needed. Because there were five race/ethnicity categories, four additional interaction terms were needed in this case. Because there are three grade levels within each of the upper and lower grade bands, two additional terms were needed to explore moderation by grade. Finally, we fit an additional model where we pooled together all data from grades K–5, treated grade as a continuous variable ranging from 0–5, and looked at the interaction between this continuous grade variable and the randomization indicator.

We report all treatment effects both in the original units and standardized (as effect sizes). To compute the standard deviation for the denominator of effect size calculations, we fit the same HLM models used to estimate the unstandardized treatment effects, however, these models contained only dummy variables for randomization status, grade, and the interaction between grade and randomization status. The variance components from these models were added up and the square root was used as the effect size denominator. This is equivalent to standardizing on the pooled standard deviation of the outcome variable within groups defined by grade and treatment condition. Quantile regression effect size estimates were standardized using the same denominator as mean regression estimates. That is, the variance components were extracted from the appropriate two or three level mean regression model.

Results

We only report estimated coefficients for independent variables involving randomization status (either individually or as an interaction). The results for mean regression HLM models estimating average treatment effects in the K–2 and 3–5 grade bands are presented in Table 6. The estimated average treatment effect in the 2017-2018 school year of attending a school randomized to the CGI condition in Spring 2013 was positive but very small in the K–2 grade band. However, for both i-Ready and FSA outcomes, the estimated average treatment effect was much larger in the 3–5 grade band. Effect sizes of 0.16 were estimated for both outcome measures in the 3–5 grade band, and the estimate for the i-Ready outcome was statistically significant at the conventional threshold of $p = .05$.

Table 7 presents the results of the quantile mixed effects regression analyses. In the K–2 grade band, the estimates are negative and small at most quantiles. However, positive estimates are obtained at the eighth and ninth deciles, with the estimate at the eighth decile much larger in absolute value relative to the other deciles. In the 3–5 grade band, estimates are remarkably consistent across all quantiles of the outcome distribution. Except for one of the nine deciles with FSA as the outcome, all estimates are positive. There appear to be somewhat larger estimates at low deciles of the i-Ready test, whereas the opposite pattern is observed for the FSA test. We note that the decile regression estimates are noisier than the mean regression estimates, and these seeming patterns could simply be the result of noise in the data. None of the quantile regression estimates were statistically significant at the conventional 0.05 level.

Our next set of analyses examined differences in treatment effects as a function of student grade within each of the K–2 and 3–5 grade bands. The reader may be interested in both differences in the average impact of CGI across grades and in grade-specific effect sizes. Table 8 presents both pieces of information. The left-hand column presents treatment effect estimates in the original metric for the reference category, the estimated deviation from the average effect in the reference category for students in other grades, and the statistical significance of the corresponding coefficients. The right-hand column presents grade specific average treatment effect estimates in the effect-size metric. In the K–2 band, treatment effect estimates were small and positive at grade K and were virtually zero at grades 1 and 2. For the i-Ready test at the 3–5 grade band, effect-size estimates were much larger in grade 4 and 5 than they were in grade 3. Both the grade 4 and 5 effect sizes were statistically significantly larger than the effect-size estimate at grade 3. For the FSA test, the effect size at grade 4 was significantly larger than the effect size at grade 3, but the effect size at grade 5 was virtually identical to that at grade 3.

We also pooled all data from grades K–5 together using the i-Ready test as an outcome. We treated grade as a continuous variable with levels from 0–5 and explored the interaction with the randomization indicator. Results are in Table 9. There is a strong, statistically significant positive trend in the effect sizes as grade level increases.

A final set of analyses looked at variation in treatment effects as a function of the following student-level variables: race/ethnicity, EL status, gifted student status, disability status, and eligibility for free or reduced-price lunch (FRL). Table 10 provides the results. Results are again presented in an analogous fashion to those in Table 7. That is, the columns with an "Estimate" heading report an average CGI effect for a reference category, deviations from the reference effect in other categories, and statistical significance of the relevant coefficients. The columns with a "Group ES" heading report the group-specific CGI effect in standardized units. There is a fair amount of consistency among the treatment effect estimates. None of the estimated moderating effects for these variables were statistically significant, but there was a marginally significant result associated with the Black coefficient in the K–2 grade band. These results suggest a possibility that Black children may benefit more from CGI relative to other racial or ethnic groups across both grade bands and both test types. There isn't much evidence for a moderating effect of the ELL, gifted or disability variables. We should also note that it is plausible that the gifted and disability variables are not entirely exogenous to the intervention. Because teachers play a role in recommending who is tested for exceptionality, it is plausible that randomization to CGI may have impacted the types of children identified as gifted and/or disabled during the five-year period.

Discussion

The present study was designed to address the two major shortcomings of the previous five RCTs of CGI PD programs: the short-term duration of the studies, and use of assessment instruments for student outcomes that were not used in practice by the schools in the study. We studied the effect of a CGI PD program after five years and conceptualized the intervention as the opportunity for teachers in elementary schools to participate in one or more years of a three-year-long CGI PD intervention. We selected the mathematics assessment used by the state of Florida for their school accountability system—the ultimate in a high-stakes test—to measure student outcomes. Because those data are only available for grades 3–5, we also estimated the effect of the CGI PD intervention using i-Ready Diagnostic, a progress monitoring assessment that is vertically scaled across all six grade levels, has high reliability estimates, and is used by many school districts within and outside the state of Florida. Both mathematics assessments were administered to almost all eligible students in the schools, resulting in low rates of measurement attrition. The results of this study are therefore maximally relevant to the manner in which schools assess student learning in mathematics.

The estimated average effect size ($g = 0.03$) on i-Ready Diagnostic scores for grades K–2 was very small, positive, and not statistically significant for the available sample. According to Kraft (2020), an effect size of 0.03 is equivalent to the average effect-size estimate of all mathematics interventions studied in RCTs funded by the U.S. Department of Education. Given the small effect-size estimate and lack of statistical significance, if the study had only looked for treatment effects at grades 1 and 2 (the grade levels of participating teachers in the original study), we may have concluded that CGI has little or no discernible impact on school mathematics achievement.

The estimated average effect sizes for grades 3–5 were positive and much larger than the effect-size estimates for grades K–2. At $g = 0.16$, the estimated effect sizes using the three available i-Ready Diagnostic scores and the one FSA score were identical in magnitude, but only the i-Ready result was statistically significant. The i-Ready Diagnostic was used to measure student achievement at the three time points over the course of the year, while the FSA was administered only one time. The three data points for i-Ready Diagnostic outcome may have served to reduce measurement error, which could explain the smaller p-value. The magnitude of the effect-size estimates obtained for the two grades 3–5 measures are above average for experimental studies in mathematics (Kraft, 2020) and may be substantively important, especially when considering that we used the standardized and state-standardized tests that were used by school districts as the outcome of interest.

We don’t have pre-intervention mathematics achievement scores for individual students in grades 1–5 in this study, because only the fifth-grade students were present in the schools before randomization occurred and the commencement of the intervention. This precludes our ability to examine whether the effects of the CGI program vary by pre-intervention levels of mathematics achievement. Quantile regression, however, enabled us to examine the effect of the program across students with different achievement levels, and results suggested a positive effect on FSA scores at eight of nine deciles. We conclude that the positive effects appear to be relatively homogenous across achievement levels.

The only clear moderator of the effect of the CGI condition on student mathematics achievement appears to be grade level, but we note that the point estimate of the impact was larger for Black students than for White students for both grade bands and for both outcome measures. (See Table 10 for details). Otherwise, the effects appeared to be homogeneous across other student demographic variables.

These new findings parallel the results of three previous one-year RCTs conducted in Florida in the sense that the CGI program had small average impacts on student achievement in the primary grade levels (i.e., K–2) and larger effects in the intermediate (i.e., 3–5) grade levels in those studies (Schoen et al., 2018a; 2020; manuscript in progress). Even after five years, we continue to see small, positive, not statistically significant (at the customary $p<.05$ threshold) effects in the early grades. We again see larger effects in the higher grade levels. This overall pattern of larger effect-size estimates in the intermediate grades is further confirmed by the statistically significant treatment-by-grade-level interaction that we found in this study.

We don't know why the CGI program has a larger effect on student achievement in the intermediate grades than in the primary grades. It could be that the effect of CGI on individual students is cumulative. Many studies report that early math achievement predicts later math achievement. Perhaps there is something about the experience of being in a CGI-trained teachers' classroom in early elementary that presents in the form of higher achievement scores (as compared to their peers in the comparison-condition schools) in later grades. We also note that the students in higher grade levels would have had more opportunity to have had at least one CGI-trained teacher, which could also contribute to the cumulative effect.

The assessments in the intermediate grade levels assess higher-level mathematics abilities than those used in the primary grades, including persistently challenging topics such as fractions. We also presume more variation in mathematics performance in the student population at the higher grade levels than at lower grade levels. It could be that the assessment tools are able to better discriminate among ability levels in the intermediate grades than they are in the primary grades.

Early exposure to multiplicative concepts could be another plausible explanation for the larger effects of the CGI program on intermediate grades mathematics achievement. Driven by arguments promoting a narrower focus at each grade level and less overlap in content across years (Schmidt et al., 2002), the policy context since mid-2000s pushed primary curriculum to be very different from intermediate. As a result, the mathematics curriculum in the early elementary grades (i.e., K–2) during the Common Core era focuses largely on addition and subtraction, and the intermediate grades focus on multiplicative ideas, including multiplication and division on whole numbers as well as fractions (NGA & CCSSO, 2010; Schoen et al., 2021). Teachers in the CGI program learn that many kindergarten students can successfully solve story problems involving multiplicative situations (Carpenter et al., 1993). They also learn how story problems involving multiplication and division with groups of ten can be used to support learning of place-value concepts—a topic that is emphasized heavily in the grades K–2 curriculum standards and is also based on an inherently multiplicative concept. In results not presented here, analysis of content coverage in videos of mathematics instruction collected during the first and second year of the study period suggested that the teachers in CGI schools were covering more above-grade content, and the teachers in comparison schools were covering more below-grade content (Schoen et al., 2015). It is possible that teachers are introducing multiplicative concepts earlier than the standards call for students to begin learning them, and this earlier exposure in CGI schools manifested in increased performance when the curriculum later addressed those multiplicative topics.

Another plausible explanation of the larger effect sizes in higher grade levels is that the CGI program influences student attitudes about mathematics and their identity as a person who is capable of doing mathematics. It is plausible that positive impacts on attitude and sense of identity could have a long-term effect on student achievement. Rather than focusing attention on misconceptions or errors, the CGI program focuses teacher attention on what students do know, what they can do, and the funds of knowledge that they bring to the classroom from their experiences outside of school. Perhaps this asset orientation and general respect for students and their problem-solving abilities results in a more positive, supportive, and productive experiences in mathematics, which could plausibly translate to higher achievement over a long period of time. The present study did not measure students' mathematics attitudes, identities, confidence, anxiety, or enjoyment, so we cannot investigate the plausibility of this potential explanation with empirical data at this time, but this is an important area for future research.

According to the theory of change for the CGI program, teacher attributes (e.g., knowledge, beliefs) and behaviors (e.g., instructional practice) mediate the effect of the program on student achievement. Because we lacked teacher-student links, we did not have data available to perform formal mediation analyses in the current study. We know from previously published studies that the CGI program did have a positive effect on teacher's mathematical knowledge for teaching (Schoen et al., 2015; 2017; 2019) and beliefs about mathematics teaching and learning (Schoen et al., manuscript under review) in the first two years of the study. We have also found that the CGI program did change instructional practice as measured by the M-CLIPS (Li & Schoen, 2020; Schoen et al., 2021) during the first two years of the study. Some of the individual links in that causal chain of events appear to have materialized, but a comprehensive study of the role of those mediating factors is still needed.

The CGI program that served as the intervention in this study is designed to be a three-year teacher PD program. Although previous RCTs of the CGI program have been done, no study has been done to evaluate the individual contribution of each of the three years of the program. That study needs to be done to provide important guidance to schools who have limited budgets for teacher PD. We cannot yet say, for example, whether it would be a better use of resources to provide twelve teachers with one year of CGI PD, four teachers with three years of CGI PD, or some other combination. More research is needed to better understand the effect of each year of the CGI PD intervention on teachers and the impact of those opportunities.

Conclusions

The results of this study provide new evidence in support of a causal argument that the CGI program can yield a positive, statistically significant, and substantively important impact on student achievement, especially in the intermediate grades. The impact of the CGI program on early elementary student achievement is less clear, because the estimated effects are not statistically significant, but the point estimates were all positive, suggesting that the program may also have a small, positive impact on mathematics learning in the early elementary grades.

We consider these results to be both compelling and in need of explanation. The theory of change for CGI hypothesizes that meaningful learning and changes in instructional practice require an extended time to occur and suggests that any effects on students are mediated by the teacher. The current study does not have the benefit of empirical data to perform mediation analyses. We therefore must continue to rely on theory to explain the mechanisms that resulted in higher achievement in the CGI schools for the time being. Further conceptualization and empirical study of those mediators hypothesized in the program theory of change is needed.

The design of almost all extant RCTs designed to estimate the effect of teacher PD programs on student achievement assume that a measurable effect of intervention on student achievement will materialize during the year in which the PD intervention is provided. But teaching is a complex endeavor. Experienced teachers have developed their habits and routines over multiple years of practice, and inexperienced teachers are often overwhelmed with just trying to survive the day and year. The CGI program is designed to span multiple years, and the program is not prescriptive with respect to instructional practice. Teachers may require a year or more of practice before they learn how to use what they have learned through the CGI program to improve teaching and learning. The results of this study imply that more studies of promising PD programs should investigate the long-term effects of those programs on student learning. More research may be needed to identify the conditions under which long-term evaluation of impact of such programs is warranted.

References

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Baroody, A. J. (1987). *Children's mathematical thinking: A developmental framework for preschool, primary, and special education teachers.* Teachers College Press.

Baroody, A. J., & Ginsburg, H. P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (p. 75–112). Lawrence Erlbaum Associates, Inc.

Bauduin, C., Schoen, R. C., Bray, W., Champagne, Z., Iuhasz-Velez, N., & Tazaz, A. (2016). Formative Assessment Collaborative Team (FACT) meetings facilitator's guide (Research Report No. 2016-05). Florida State University. https://doi.org/10.17125/fsu.1493410046

Banilower, E. R., & Smith, P. S. (2006, April 3–6). *Measuring teachers' knowledge for teaching force and motion concepts.* Paper presented at the National Association for Research in Science Teaching (NARST) 2006 Annual Meeting. San Francisco, CA.

Bloom, H., Hill, C., Rebeck Black, A.., & Lipsey, M. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289–328.

Bloom, H.S., Richburg-Hayes, L. and Rebeck Black, A. (2007). Using Covariates to Improve Precision for Studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29(*1), 30–59.

Bray, W., Johnson, J., Rivera, N., Fink, L-A., Bauduin, C., & Schoen, R. C. (2019). Unlocking mathematical understanding together through FACT meetings. *Teaching Children Mathematics, 25*(6), 370–377.

Carpenter, T. P. & Moser, J. M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), *The acquisition of mathematical concepts and processes* (pp. 7–44). Academic Press.

Carpenter, T. P. (1989). Teaching as problem solving. In R. I. Charles & E. Silver (Eds.), Research agenda in mathematics education: The teaching and assessing of mathematical problem solving (pp. 187–202). National Council of Teachers of Mathematics and Erlbaum.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4), 385–531.

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Levi, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes. *Journal for Research in Mathematics Education, 24*(5), 428–441.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal, 97*(1), 3–20.

Carpenter, T.P. Fennema, E., Franke, M.L., Levi, L., & Empson, S.B. (1999). *Children's Mathematics: Cognitively Guided Instruction.* Heinemann.

Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education, 29*(1), 3–20.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking Mathematically: Integrating Arithmetic and Algebra in Elementary School.* Heinemann.

Carpenter, T. P., & Franke, M. L. (2004). Cognitively Guided Instruction: Challenging the core of educational practice. In T. K. Glennan, S. J. Bodilly, J. R. Galegher & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 41–80). RAND Corporation.

Carpenter, T. P., Fennema, E. E., Franke, M. L., Empson, S. B., & Levi, L. W. (2015). *Children's mathematics: Cognitively guided instruction* (2nd ed). Heinemann.

Clarke, D. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947–967.

Curriculum Associates. (2018). *i-Ready assessments technical manual.* Author.

Empson, S. B. & Levi, L. (2011). *Extending children's mathematics: Fractions and decimals.* Heinemann.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403–434.

Fermanich, M. L. (2002). School spending for professional development: A cross-case analysis of seven schools in one urban district. *The Elementary School Journal*, *103*(1), 27–50.

Florida Department of Education. (2018). *Florida Standards Assessments 2017-2018: Annual Technical Report, Volumes 1–7.* Author.

Franke, M. L., Carpenter, T., Fennema, E., Ansell, E., & Behrend, J. (1998). Understanding teachers' self-sustaining, generative change in the context of professional development. *Teaching and Teacher Education, 14*(1), 67–80.

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal, 38*(3), 653–689.

Fuson, K. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning. National Council of Teachers of Mathematics.

Franke, M. L. & Kazemi, E. (2001). Learning to teach mathematics: Focus on student thinking. *Theory Into Practice, 4*0(2), 102–109.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/pdf/20084030.pdf

Garet, M. S., Heppen, J., Walters, K., Smith, T., & Yang, R. (2016a). Does content-focused teacher professional development work? Findings from three Institute of Education Sciences studies. (NCEE Evaluation Brief No. 2017-4010). Institute of Education Sciences.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., et al. (2016b). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development* (NCEE 2016-4010). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/pubs/20164010/pdf/20164010.pdf

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches.* (REL 2014-010). US Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs

Geraci, M. (2014). Linear quantile mixed models: The lqmm package for Laplace quantile regression. *Journal of Statistical Software, 57*(13), 1–29. doi:10.18637/jss.v057.i13

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*(3), 381–391.

Hiebert, J. & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). Macmillan.

Hunt, J. H., Martin, K., Khounmeuang, A., Silva, J., Patterson, B., & Welch-Ptak, J. (2020). Design, development, and initial testing of asset-based intervention grounded in trajectories of student fraction learning. *Learning Disability Quarterly.* https://doi.org/10.1177/0731948720963589

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education, 38*(3), 258–288.

Jaslow, L. & Evans, E. L. (2012). Purposeful pedagogy and discourse instructional model: Student thinking matters most. Arkansas Department of Education. Retrieved May 4, 2015 from https://sites.google.com/a/archford.org/arch-ford-mathematics/purposeful-pedagogy-and-discourse-model

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253.

Levi, L. (2017, October 19). Classroom embedded work: An alternative to observation lessons. *Teaching Is Problem Solving.* https://www.teachingisproblemsolving.org/blog/classroom-embedded-work/

Levi, L., & Empson, S. (2011). *Extending Children's Mathematics: Fractions and Decimals*. Heinemann.

Li, L., & Schoen, R. C. (2020, October 23). *Using a Many-Facet Rasch Model to Gain Insight into Measurement of Mathematics Instructional Practice.* Presented at the first annual symposium on observing and measuring how teachers implement CGI in their classrooms. [Presentation delivered virtually.]

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis, 41*(3), 260–293.

Montgomery, J., Nyhan, B. and Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, *62*(3), 760-775.

Moscardini, L. (2014). Developing equitable elementary mathematics classrooms through teachers learning about children's mathematical thinking: Cognitively Guided Instruction as an inclusive pedagogy. *Teaching and Teacher Education, 43*, 69–79.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* Authors.

Nielsen, L, Steinthorsdottir, O. B., & Kent, L. B. (2016). Responding to student thinking: Enhancing mathematics instruction through classroom based professional development. *Middle School Journal, 47*(3), 17–24, https://doi.org/10.1080/00940771.2016.1135096

Odden, A., Archibald, S., Fermanich, M., & Gallagher, H. A. (2002). A cost framework for professional development. *Journal of Education Finance*, *28*(1), 51–74.

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open, 7*(1), 1–29.

Philipp, R. A., Ambrose R., Lamb, L. L. C., Sowder, J. T., Schappelle, B. P., Sowder, L, Thanheiser, E., & Chauvot, J. (2007). Effects of Early Field Experiences on the Mathematical Content Knowledge and Beliefs of Prospective Elementary School Teachers: An Experimental Study. *Journal for Research in Mathematics Education, 38*(5), 438–476.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement, 1*(3), pp. 233–247.

Schoen, R. C., Anderson, D., & Bauduin, C. (2017). *Elementary mathematics student assessment: Measuring the performance of grade K, 1, and 2 students in number, operations, and equality in spring 2016* (Research Report No. 2017-22). Learning Systems Institute, Florida State University. http://doi.org/10.17125/fsu.1534964774

Schoen, R. C., & Champagne, Z. (Eds.) (2017). *What's Next? Stories of teachers engaging in collaborative inquiry focused on using student thinking to inform instructional decisions.* Retrieved from https://www.teachingisproblemsolving.org/whats-next-stories/

Schoen, R. C., Bray, W., Wolfe, C., Nielsen, L., & Tazaz, A. M. (2017). Developing an assessment instrument to measure early elementary teachers' mathematical knowledge for teaching. *The Elementary School Journal, 118*(1), 55–81. https://doi.org/10.1086/692912

Schoen, R. C., LaVenia, M., & Tazaz, A. M. (2018a). *Effects of the first year of a three-year CGI teacher professional development program on grades 3–5 student achievement: A multisite cluster-randomized trial.* (Research Report No. 2018-25). Florida State University. http://doi.org/10.33009/fsu.1562595733

Schoen, R. C., Anderson, D., Riddell, C. M., & Bauduin, C. (2018b). *Elementary Mathematics Student Assessment: Measuring the performance of grade 3, 4, and 5 students in number (whole numbers and fractions), operations, and algebraic thinking in fall 2015* (Research Report No. 2018-24). Florida State University. https://doi.org/10.33009/fsu.1581609234

Schoen, R. C., Kisa, Z., & Tazaz, A. M. (2019, March). *Beyond the horizon: Examining the associations among professional development, teachers' subject-matter knowledge, and student achievement.* Paper presented at the spring conference of the Society for Research in Educational Effectiveness, Washington, DC.

Schoen, R. C., & LaVenia, M. (2019). Teacher beliefs about mathematics teaching and learning: Identifying and clarifying three constructs. *Cogent Education OA, 6*(1). https://doi.org/10.1080/2331186X.2019.1599488

Schoen, R. C., LaVenia, M., Tazaz, A., Farina, K., Dixon, J. K., & Secada, W. G. (2020). *Replicating the CGI Experiment in Diverse Environments: Effects on Grade 1 and 2 Student Mathematics Achievement in the First Program Year* (Research Report No. 2020–02). Florida State University. https://doi.org/10.33009/fsu.1601237075

Schoen, R. C., Champagne, Z., Whitacre, I., & McCrackin, S. (2021). Comparing the number and distribution of additive word problems in first-grade U.S. textbooks in the 1980s and the Common Core era. *School Science and Mathematics, 121*(2), 110–121. https://doi.org/10.1111/ssm.12447

Schoen, R. C., Bray, W.S., Tazaz, A. M., & Buntin, C. K. (2022). *A description of the Cognitively Guided Instruction professional development program in Florida: 2013–2020.* Learning Systems Institute, Florida State University. https://doi.org/10.33009/fsu.1643828800

Schoen, R. C., Whitacre, I., & Champagne, Z. M. (In-press). Relative difficulty of additive word problems for U.S. first graders: Synthesizing and updating the literature. *Journal for Research in Mathematics Education.*

Schmidt, W.H., Houang, R., & Cogan, L. (2002). A coherent curriculum: The case of mathematics. American Educator, 1-18.

Secada, W. G. & Brendefur, J. L. (2000, Fall). CGI student achievement in Region VI: Evaluation findings. *The Newsletter of the Comprehensive Center–Region VI, 5*(2). Insert pp. 1–4.

Sireci, S.G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*, 237–247.

Smith, M. S. & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussions.* National Council of Teachers of Mathematics.

TNTP. (2015). *The mirage: Confronting the hard truth about our quest for teacher development*. https://tntp.org/assets/documents/TNTP-Mirage2015.pdf

U.S. Department of Education. (2014). *Fiscal year 2014 budget summary and background information.* Author. http://www2.ed.gov/about/overview/budget/budget14/summary/14summar

Vacc, N. N. & Bright, G. W. (1999). Elementary preservice teachers' changing beliefs and instructional use of children's mathematical thinking. *Journal for Research in Mathematics Education, 30*(1), 89–110.

Villaseñor, A., & Kepner, H. S. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education, 24*(1), 62–69.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education, 24*, 173–209.

Whitehurst, G. J. (2004) Making education evidence based: Premises, principles, pragmatics, and politics. Retrieved from: https://ies.ed.gov/director/pdf/20040426.pdf

Whitehurst, G. (2018) The Institute of Education Science: A model for federal research offices. *Annals of the American Academy of Political and Social Science, 678,* 124–133.

Zan, B., & Geiken, R. (2010). Ramps and pathways: Developmentally appropriate, intellectually rigorous, and fun physical science. *YC Young Children, 65*(1), 12–17.

Table 1

Number of Teachers in the 22 Schools Who Participated in the CGI Program During Each Year of the Study, Split by Grade Level

	CGI condition			Comparison condition			
School year	K–2	3–5	Total	K–2	3–5	Total	*Total*
2013-2014	107	0	*107*	0	0	*0*	*107*
2014-2015	103	0	*103*	0	0	*0*	*103*
2015-2016	52	8	*60*	21	8	*29*	*89*
2016-2017	37	14	*51*	10	11	*21*	*72*
2017-2018	10	2	*12*	0	0	*0*	*12*
Total	*309*	*24*	*333*	*31*	*19*	*50*	*383*

Note. Individuals may appear in more than one row of the table if they participated in the CGI program in multiple years. A total of 182 unique individuals in the CGI-condition schools participated in the CGI program (i.e., attended workshops) during the five-year period. Only 34 unique individuals in the comparison-condition schools had similarly participated.

Table 2

School Demographics Based on Available Student Data in the Full Sample

	CGI schools (11 schools)		Comparison schools (10 schools)	
	Mean	s.d.	Mean	s.d.
Race/Ethnicity				
% American Indian	0.2	0.2	0.1	0.1
% Asian	5.7	7.0	5.0	4.1
% Black or African American	23.9	23.0	12.2	4.2
% Hispanic	35.6	19.8	50.0	14.7
% Multi-racial	2.5	1.5	2.8	1.4
% White	32.1	19.6	29.9	15.7
% English Learners	13.2	11.3	23.3	13.5
% FRL	64.4	20.7	71.3	17.8
% Gifted	6.6	4.0	5.3	2.0
% Disabled	12.6	4.4	12.4	3.5
% Female	49.5	1.5	48.2	1.0
Number of students in school	689.4	281.2	933.6	272.8

Note. %FRL=Percentage of students in the school who were eligible for free or reduced-price lunch. For race/ethnicity, if a student was Hispanic, then they were categorized as Hispanic, regardless of race.

Additionally, any (non-Hispanic) child with indicators for two or more race categories were categorized as multi-racial. Therefore, race/ethnicity categories were defined to be mutually exclusive, which is consistent with how the Florida Department of Education collected and reported race/ethnicity information at the time of the study.

Table 3

Number of Students Represented in the Full Available Sample for the 2017-2018 School Year, Disaggregated by Treatment Condition

	CGI schools		Comparison schools	
	n	Within-condition percentage	*n*	Within-condition percentage
Grade level				
Kindergarten	1,103	15.0	1,389	15.3
First-Grade	1,213	16.5	1,535	16.9
Second-Grade	1,198	16.3	1,418	15.6
Third-Grade	1,250	17.0	1,609	17.7
Fourth-Grade	1,332	18.1	1,576	17.4
Fifth-Grade	1,269	17.2	1,550	17.1
Mathematics assessment				
i-Ready BOY (K through Fifth Grade)	6,969	91.2	8,292	88.8
i-Ready MOY (K through Fifth Grade)	6,992	92.2	8,355	89.5
i-Ready EOY (K through Fifth Grade)	6,563	86.5	7,607	81.5
FSA (Third through Fifth Grade)	3,664	95.1	4,420	93.3

Table 4

Number of Students with Available Data by Test, Grade Level, and Treatment Condition

	i-Ready Timepoint 1		i-Ready Timepoint 2		i-Ready Timepoint 3		FSA	
Grade level	CGI	Comp.	CGI	Comp.	CGI	Comp.	CGI	Comp.
K	1,034	1,215	1,042	1,241	1,027	1,241	–	–
1	1,139	1,402	1,143	1,405	1,118	1,376	–	–
2	1,137	1,303	1,128	1,314	1,106	1,317	–	–
3	1,182	1,482	1,183	1,497	1,051	1,423	1,184	1,507
4	1,267	1,459	1,270	1,461	1,151	1,348	1,267	1,469
5	1,210	1,431	1,226	1,437	1,110	902	1,213	1,444

Note. Comp.= comparison condition

Table 5

Average Scaled Score (and Standard Deviation) by Test, Grade Level, and Treatment Condition

	i-Ready Timepoint 1		i-Ready Timepoint 2		i-Ready Timepoint 3		FSA	
Grade level	CGI	Comp.	CGI	Comp.	CGI	Comp.	CGI	Comp.
K	348.2	346.7	366.6	364.0	384.5	380.7	–	–
	(22.3)	(22.0)	(23.9)	(22.9)	(25.2)	(25.1)		
1	378.4	377.3	397.7	395.0	412.3	410.7	–	–
	(25.0)	(25.6)	(26.2)	(26.6)	(27.4)	(27.4)		
2	408.9	405.2	425.8	423.6	439.0	436.4	–	–
	(26.2)	(27.1)	(26.8)	(26.6)	(26.6)	(26.8)		
3	434.1	428.0	451.1	444.2	467.7	459.4	307.8	302.5
	(24.2)	(26.8)	(25.4)	(27.4)	(27.9)	(30.0)	(20.7)	(22.7)
4	460.6	454.0	476.8	466.5	491.4	478.7	323.1	316.1
	(25.7)	(26.7)	(26.2)	(27.9)	(29.4)	(32.6)	(23.1)	(23.4)
5	476.9	470.8	488.0	481.9	497.6	490.3	329.1	325.4
	(27.9)	(29.4)	(28.6)	(29.7)	(30.2)	(33.6)	(23.8)	(24.3)

Note. Comp.= comparison condition

Table 6

Average Effects of School Randomization to CGI Five Years after Randomization

	Estimate (SE)	Effect size
i-Ready: Kindergarten through Second Grade	0.89 (2.11)	0.03
i-Ready: Third through Fifth Grade	4.96 (1.73)	0.16*
FSA: Third through Fifth Grade	3.63 (2.03)	0.16

Note. ***$p < 0.001$, **$p<0.01$, *$p<0.05$, ~$p<0.10$

Table 7

Estimated Average Effects at Quantiles of the Outcome Distribution

Test and grade level	Decile	Estimate (SE)	Effect size
i-Ready: Kindergarten through Second Grade			
	1	-0.94 (3.14)	-0.03
	2	-0.58 (2.44)	-0.02
	3	-0.15 (2.25)	-0.01
	4	-1.32 (1.95)	-0.04
	5	-1.00 (2.33)	-0.03
	6	-0.63 (2.36)	-0.02
	7	-0.72 (2.67)	-0.02
	8	4.59 (2.75)	0.15
	9	0.84 (3.42)	0.03
i-Ready: Third through Fifth Grade			
	1	6.95(3.42)	0.22
	2	6.00 (3.24)	0.19
	3	5.1 (2.54)	0.16
	4	5.22 (3.05)	0.17
	5	2.3 (1.86)	0.07
	6	3.31 (2.71)	0.11
	7	4.05 (3.13)	0.13
	8	3.13 (1.99)	0.10
	9	1.91 (3.86)	0.06
FSA: Third through Fifth Grade			
	1	3.31 (3.35)	0.15
	2	-1.06 (2.44)	-0.05
	3	1.81 (2.76)	0.08
	4	2.27 (2.36)	0.10
	5	2.16 (2.09)	0.09
	6	3.16 (2.56)	0.14
	7	3.57 (2.68)	0.16
	8	4.21 (2.76)	0.19
	9	2.64 (2.96)	0.12

Note. *** $p < 0.001$, **$p<0.01$, *$p<0.05$, ~$p<0.10$

Table 8

Variation in Treatment Effects as a Function of Student Grade (within Each Grade Band)

Outcome measure and grade level	Estimate (SE)	Grade-specific CGI effect size
i-Ready: Kindergarten through Second Grade		
Kindergarten (reference category)	1.62 (2.25)	0.06
Grade 1 deviation	-1.58 (1.21)	0.00
Grade 2 deviation	-0.41 (1.22)	0.05
i-Ready: Third through Fifth Grade		
Grade 3 (reference category)	2.72 (1.87)	0.09
Grade 4 deviation	4.05 (1.15)	0.21^{***}
Grade 5 deviation	2.64 (1.16)	0.18^{*}
FSA: Third through Fifth Grade		
Grade 3 (reference category)	2.80 (2.11)	0.12
Grade 4 deviation	2.47 (0.99)	0.23^{*}
Grade 5 deviation	-0.03 (1.01)	0.12

*Note: *** p <0.001, **p<0.01, *p<0.05, ~p<0.10*. While the coefficient estimates are expressed as deviations from the reference category, the effect size estimates are the estimated effect size within a particular grade.

Table 9

Results for Research Question 4: i-Ready Kindergarten through Fifth Grade Analysis

Outcome measure and grade level	Estimate (SE)
i-Ready: Kindergarten through Fifth Grade	
Intercept (Kindergarten)	-1.42 (1.75)
Grade Slope	1.70 (0.21) ***

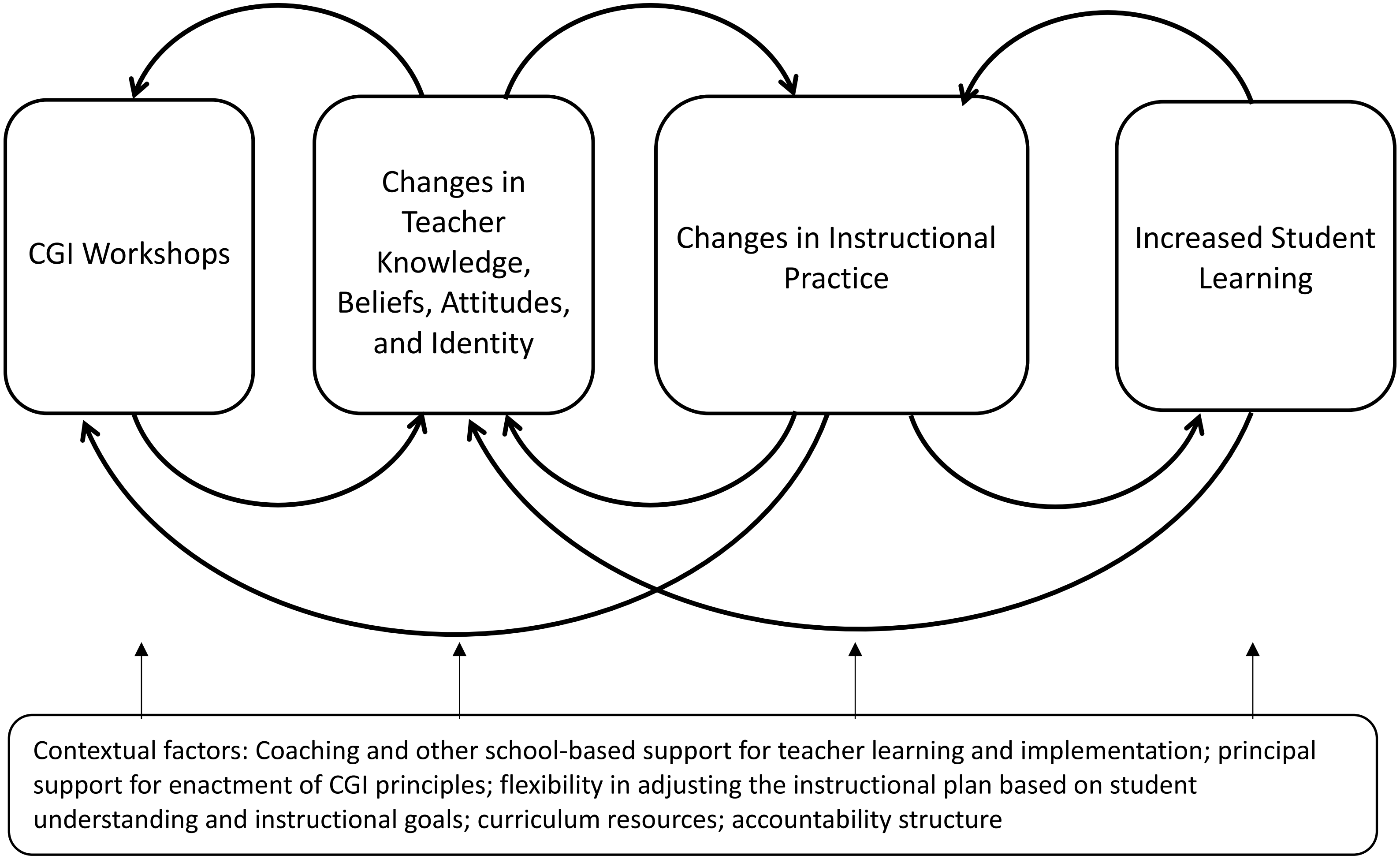
*Note: *** p <0.001, **p<0.01, *p<0.05, ~p<0.10*

Table 10

Results for Research Question 4: Race, ELL, Gifted, SWD, and FRL Moderators

	i-Ready Grades K–2		i-Ready Grades 3–5		FSA Grades 3–5	
	Estimate (SE)	Group ES	Estimate (SE)	Group ES	Estimate (SE)	Group ES
Race						
White (ref.)	0.37 (2.37)	0.01	5.01 (1.87)	0.16*	4.27 (2.12)	0.18
Asian (dev.)	1.96 (2.16)	0.08	-1.57 (2.12)	0.11	-0.95 (1.83)	0.14
Black (dev.)	2.69 (1.58)	0.10~	2.16 (1.55)	0.23	0.66 (1.36)	0.21
Hispanic (dev.)	-0.15 (1.22)	0.01	-0.72 (1.20)	0.14	-1.49 (1.05)	0.12
Other (dev.)	1.14 (2.81)	0.05	0.50 (3.04)	0.18	-1.69 (2.66)	0.11
ELL						
No (ref.)	1.01 (2.14)	0.04	4.95 (1.75)	0.16*	3.60 (2.04)	0.16
Yes (dev.)	-0.63 (1.37)	0.02	0.07 (1.48)	0.16	0.23 (1.28)	0.17
Gifted						
No (ref.)	0.82 (2.12)	0.03	5.14 (1.73)	0.17*	3.83 (2.02)	0.17
Yes (dev.)	2.47 (2.87)	0.12	-1.83 (1.58)	0.11	-1.96 (1.37)	0.09
Disability						
No (ref.)	0.91 (2.12)	0.03	4.99 (1.75)	0.16*	3.72 (2.04)	0.16
Yes (dev.)	-0.19 (1.66)	0.02	-0.23 (1.44)	0.15	-0.63 (1.28)	0.13
Free/Reduced Price Lunch						
No (ref.)	0.67 (2.22)	0.02	4.71 (1.87)	0.15*	4.52 (2.16)	0.20~
Yes (dev.)	0.34 (1.12)	0.04	0.38 (1.10)	0.16	-1.35 (0.96)	0.14

*Note: *** p <0.001, **p<0.01, *p<0.05, ~p<0.10*. “Ref” refers to the reference category for a particular comparison. “Dev” refers to the deviation of the estimated treatment effect for that category from the reference category. While the coefficient estimates are expressed as deviations from the reference category, the “Group ES” estimates are the estimated effect size within the category in question.



Contextual factors: Coaching and other school-based support for teacher learning and implementation; principal support for enactment of CGI principles; flexibility in adjusting the instructional plan based on student understanding and instructional goals; curriculum resources; accountability structure

Figure 1. *CGI program theory of change.*