Journal for Research in Mathematics Education
2024, Vol. 55, No. 4, 163–181

Looking Inside the Black Box: Measuring Implementation and Detecting Group-Level Impact of Cognitively Guided Instruction

Robert Schoen, Wendy Bray, Claire Riddell, Charity Buntin, and Naomi Iuhasz-Velez
Florida State University

Walter Secada
University of Miami

Eva Yujia Li
Yale University

Studies have found that some teacher professional development programs that are based on Cognitively Guided Instruction (CGI) can increase student mathematics achievement. The mechanism through which those effects are realized has been theorized, but more empirical study is needed. In service of this need, we designed a novel measure of instructional practice to assess the extent to which observable features of mathematics instruction are consistent with the principles of CGI. We describe the conceptual foundations and first use of the instrument, which we call M-CLIPS. We found that teachers involved in the first 2 years of a CGI program were using methods consistent with the principles. In contrast, instructional practice in the comparison condition was mostly inconsistent with those principles.

Keywords: Cognitively Guided Instruction; Instructional practice; Classroom observation; Problem solving; Classroom discourse; Assessment

Several experimental and quasi-experimental studies of mathematics professional development (PD) programs that are based on Cognitively Guided Instruction (CGI) have reported positive effects on students (Carpenter et al., 1989; Fennema et al., 1996; Jacobs et al., 2007; Schoen et al., 2020, in press; Schoen, Rhoads, et al., 2022; Villaseñor & Kepner, 1993). These studies postulated that the effects of CGI PD on students are mediated by the teacher, primarily through instructional practice (Carpenter et al., 1989; Schoen, Bray, et al., 2022; Schoen, Rhoads, et al., 2022). CGI books have offered some principles and suggestions for instructional practice (Carpenter et al., 1999, 2003, 2015, 2017; Empson & Levi, 2011), but they did not prescribe a specific model of classroom implementation. Likewise, the CGI PD programs in these studies did not provide a curriculum and were not prescriptive regarding how teachers should use what they learned in CGI PD.

The combination of the existence of studies demonstrating a positive impact on student learning and the lack of explicit guidance for the implementation of CGI creates a “black box” problem with respect to understanding the mechanisms through which CGI affects learning. Without empirical data collected through observation of instructional practice, we must rely on theory to explain how the CGI PD changes classroom instruction and student learning. Small studies can offer detailed explanations for how these factors appear to interact in individual schools or classrooms. However, measurement of instructional practice using relatively efficient and objective instruments is needed to support arguments about generalizability and an examination of how mechanisms interact similarly or differently in various settings.

In this article, we describe our recent efforts to identify and measure observable indicators of mathematics instructional practice that are consistent with CGI principles. These efforts resulted in the development of a novel observational measure of instructional practice that we call Mathematics-Cognition, Language, Interaction, and Problem Solving (M-CLIPS). The purpose of this article is to describe what M-CLIPS measures and whether it can provide insight into instructional practice in CGI and practice-as-usual classrooms. The following research questions guided the study:

1. What can the data from M-CLIPS tell us about how instructional practice differs within and between groups of teachers who have and have not participated in the first 2 years of a CGI-based program?
2. Can the M-CLIPS instrument detect group-level differences in mathematics instruction enacted by teachers participating in a CGI-based PD program compared with those who are not?

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant No. 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.

Brief Background on Cognitively Guided Instruction

The unifying premise of CGI is that children enter school with experiential knowledge and both formal and informal mathematical knowledge that can serve as a foundation for learning much of the early elementary mathematics curriculum (Carpenter et al., 1996, 1999). A CGI approach to teacher PD is based on the notion that children often view mathematics differently than adults do and that striving to understand the child's perspective is an important part of teaching. The developers of CGI advocate for children to be encouraged to solve problems in ways that make sense to them and for instruction to build on children's ways of thinking as their ideas become more sophisticated or abstract. The teacher's role in implementation of CGI is to provide instruction that leverages and elevates children's ways of knowing and understanding for use as a foundation to build new knowledge.

CGI PD develops teachers' capacities to attend to the details in their students' mathematical thinking by supporting them to construct and test models of student thinking in specific content domains. Teachers learn a great deal about mathematics language and content through CGI PD, and they do so by observing and discussing how children think about mathematics (Schoen, Bray, et al., 2022). Although CGI PD programs refrain from telling teachers exactly how to use what they learn in the PD program to teach, the resulting change in teachers' pedagogical content knowledge and beliefs "provides a context in which teachers can interpret and apply general pedagogical knowledge" (Carpenter et al., 1996, p. 4). The previously mentioned empirical studies demonstrating that CGI can increase student achievement are compelling and in need of explanation. Flexibility in the interpretation of CGI in practice may be one quality that makes CGI powerful and robust, but it also presents a challenge with respect to understanding the mechanisms through which CGI affects students. Researchers and practitioners alike want to know what changes in instructional practice are expected to be implemented as a result of participation in CGI PD programs and how those changes ultimately relate to student outcomes. The following sections describe our attempts to conceptualize core principles of CGI, relate those principles to extant research on instructional practice, and translate the principles into measurable features of mathematics teaching and learning environments.

Deciding What to Look For Inside the Black Box of Mathematics Instruction

Mathematics instruction can be conceptualized in myriad ways and involves a complex orchestration of behaviors, decisions, classroom norms, and more. Driven by the goal of measuring observable features of instructional practice that could represent the CGI principles in action, we knew that we would have to make decisions about how to narrow the focus to a few key features. We studied the definitive CGI books and dozens of CGI-related research articles to seek guidance on how to conceptualize implementation of CGI in ways that would be consistent across various interpretations and examples of CGI programs. Over the past decade, we also collectively observed hundreds of days of CGI PD and hundreds of classroom lessons facilitated by teachers who had participated in CGI PD as well as teachers who had not participated in CGI PD. This gave us valuable insight into the contrasting approaches to mathematics instruction in practice-as-usual classrooms and in classrooms of teachers who had been involved in CGI PD. We then drafted overarching principles and observable aspects of classroom instruction that seemed to be consistent with CGI principles and received critical feedback on these drafts from CGI experts and members of our project advisory board, including authors of the definitive CGI books.

Conceptual Framework for M-CLIPS

Our review of CGI-related literature, observation of CGI PD, and observation of classroom instruction led to our identification of the following four core principles regarding observable features of CGI implementation in mathematics lessons:

1. Problem solving plays an integral role in the teaching and learning of mathematics.
2. The teacher attends to students' mathematical thinking.
3. Students engage with their peers' ideas about mathematics.
4. The teacher supports students' understanding and meaningful use of the language of mathematics.

Our thesis is that features of these four principles are present and observable in mathematics lessons when CGI is implemented. These principles and associated instructional practices, described in subsequent sections, are informed by CGI-related research but not determined solely by it. They also reflect emergent principles and practices in the broader field of mathematics teaching and learning.

Principle 1: Problem Solving Plays an Integral Role in the Teaching and Learning of Mathematics

Problem solving involves grappling with tasks for which the path to a solution is not already known by the student or prescribed by the teacher (National Council of Teachers of Mathematics [NCTM], 2000). During problem solving, students engage in the work of making sense of problems, deciding how to approach solving them using what they already know, and reflecting on their progress. Discussing teaching through problem solving, Carpenter et al. (2015) emphasized, "the critical consideration is that students are engaged in deciding how to solve problems using what makes sense to them" (p. 78).

Research on CGI classrooms illustrates how the teaching and learning of fundamental aspects of the elementary mathematics curriculum (e.g., meaning and application of basic operations, base-ten number system, fact fluency, and word problems) can be taught through problem solving (Bray & Blais, 2017; Bray & Maldonado, 2018; Fennema et al., 1993; Hiebert et al., 1996). Students are presented with problems that they do not yet know how to solve and are encouraged to construct their own ideas and solutions through reasoning and sense making. The students' methods of solving problems can then, in turn, be used to make key mathematical ideas more salient, usually during a discussion or reflection phase of a lesson. Worked examples are often provided and examined by the class and represent the mathematical ideas of the class members, not some outsider's ideas. The resulting strategies for solving the problem typically do not conform to a single method used by all the students, although a high degree of commonality can exist among strategies and notation used by individuals in a classroom.

Mathematical tasks (or problems) implemented through problem solving typically allow multiple entry points and possible solution paths, allowing students at varying levels of mathematical understanding to engage in high levels of cognitive complexity (Stein et al., 2009). The students' role in working on high-level tasks involves some combination of reasoning and sense making, creativity, planning, mathematical modeling, providing explanations and justifications, making generalizations, or analysis of mathematical ideas. Teachers intentionally budget sufficient time and make available a variety of tools to support students' efforts to devise their own ways of approaching assigned problems. Teachers support student engagement in tasks in ways that maintain the cognitive complexity for students, even when students experience struggle (Henningsen & Stein, 1997).

Teachers who implement a problem-solving approach to mathematics instruction must honor the diversity in students' perspectives, thoughts, and ways of approaching problems. Teachers demonstrate genuine appreciation for student efforts that involve novel methods or interpretations, concrete models, or heuristical strategies, including those that end in partial or incorrect solutions. Teachers convey a sense that variation in perspective is to be celebrated, and the process of problem solving creates opportunities to learn from one another and celebrate each individual's contributions to their own learning and that of the class. In such environments, the teacher engages with students' mathematical work and ideas with the utmost respect (Reeve & Jang, 2006; Strati et al., 2017). The teacher consistently conveys curiosity and interest in students' ideas and confidence in their abilities as problem solvers (Kazemi & Hintz, 2014). The teacher is careful to consistently position all students' contributions as valuable for their own learning and that of the class.

Principle 2: The Teacher Attends to Students' Mathematical Thinking

In the complex environment of a mathematics classroom, many factors compete for the teacher's attention (Sherin et al., 2011). A hallmark feature of CGI is a deliberate choice by the teacher to attend to the details in students' thinking processes during and after the process of solving problems with the intention to use what is learned to inform teachers' subsequent instructional decisions (Carpenter et al., 1989, 2015; Fennema et al., 1993; Jacobs et al., 2022). Through CGI PD, teachers learn to observe and interact with students with a focus on gathering information about how individual students are solving and thinking about mathematics problems. Students' final answers to a given mathematics problem are important, but the final answer (and its correctness) is just one of many pieces of information gathered by the teacher about student understanding.

Teachers can gather information about individual students' mathematical thinking by being attentive to how students manipulate materials, record their ideas on paper, or use words and gestures to describe their thoughts (Sherin & Lynn, 2019). Although passive observation offers the opportunity to gather some evidence of children's thinking, high levels of attending often involve teacher questioning and in-the-moment teacher–student interactions that serve to further elicit, support, or extend the thinking of that specific student (Empson & Jacobs, 2008; Fraivillig et al., 1999). Jacobs and Empson (2016) stressed the importance of teachers making sense of a child's current thinking before attempting to build on it. This involves eliciting information about the details of a student's existing strategy by posing starter questions to initiate conversation (e.g., How are you thinking about this problem?) and pressing students for an explanation of specific parts of their problem-solving process (e.g., What did you do first? What do these marks on your paper represent? Can you tell me what you heard in your mind when you were doing that?). Teachers may also ask follow-up questions that probe the limits of a student's understanding (e.g., What would happen if. . . ?). Although these questions might be asked to extend students' thinking, their main purpose can be to support teacher assessment of student thinking, which can inform the teacher about student understanding or interests. A student's responses can help a teacher to pinpoint where that student's thinking on that problem may fit in the CGI frameworks for types of strategies.

Principle 3: Students Engage With Their Peers' Ideas About Mathematics

In CGI classrooms, peer interaction is a key mechanism to support learning. Peer interaction involves students in expressing their own mathematical ideas and actively engaging with the ideas generated by other students (Carpenter et al., 2015; Fennema et al., 1996; Franke et al., 2015). Students in CGI classrooms do learn from the teacher, but they also learn from their peers and through participation in the interactive, social aspects of doing mathematics. Such interactions provide motivation for students to make their own thinking explicit, learn from the ideas of others, and actively reflect on the underlying structure of mathematical concepts (Chapin & O'Connor, 2007; Mercer & Howe, 2012).

Collaborative groups and teacher-facilitated discussion are two interactive classroom structures that have been found through previous research to postively affect student learning (Mercer & Howe, 2012; Webb et al., 2014; Woods et al., 2006). This body of research emphasized, however, that the quality and content of interactions matter. To reap the benefits of collaborative groups, students must express and justify their own ideas and actively engage in listening to, making sense of, and responding to the ideas of group members (Ing et al., 2015). Similarly, research on social interaction patterns in teacher-facilitated discussion has suggested that greater student involvement is related to higher levels of expressed mathematical student thinking and achievement (Webb et al., 2014; Woods et al., 2006). Furthermore, the degree to which students give detailed explanations and the degree to which they engage with the mathematical ideas of others are predictive of student learning outcomes (Webb et al., 2019).

The teacher plays a critical role in cultivating and reinforcing social norms for classroom interactions focused on mathematics (Franke et al., 2015; Mercer & Howe, 2012; Walshaw & Anthony, 2008; Yackel & Cobb, 1996). Teachers must create respectful and equitable learning environments in which students' ideas are taken seriously, no one's ideas are derided or ignored, and all students are welcome in the conversation (Chapin & O'Connor, 2007). Teachers orient students' attention to one another's ideas by positioning students and their ideas as important resources for learning, including during moments of struggle (Franke et al., 2015; Gresalfi, 2009; Kazemi & Hintz, 2014; Webb et al., 2008). Researchers have identified specific conversational actions that teachers can use to support productive discussion, such as revoicing, inviting to add on, and providing sufficient wait time (Chapin & O'Connor, 2007). Ing et al. (2015) found that teachers' support for student participation, defined as their actions to elicit student thinking and support students with engaging in the mathematical ideas of their peers, positively predicted student participation, which in turn positively predicted student achievement.

Principle 4: The Teacher Supports Students' Understanding and Meaningful Use of the Language of Mathematics

Mathematical language may be interpreted broadly to include both expressive and receptive communication in spoken, written, and physical forms. Meaningful use of mathematical language is part and parcel of learning mathematics content. In the context of the CGI classroom, teachers and students communicate about and draw connections among mathematical ideas through spoken and written words, pictures, physical models, notations, and other forms of language. CGI teachers intentionally create abundant opportunities for students to express their mathematical thinking privately and publicly.

In CGI classrooms, teachers intentionally foster students' abilities to express and elaborate their mathematical ideas through various representations, including words, writing, and mathematical models. The act of describing, explaining, and elaborating mathematical ideas helps students clarify and internalize key ideas, forge associations with related ideas, and identify gaps or weaknesses in understanding (NCTM, 2000). Teachers ask probing questions that support students in providing additional mathematical detail (Franke et al., 2009). They also support students' development of increased clarity and precision through strategies such as revoicing (Forman et al., 1997; Michaels & O'Connor, 2015; Moschkovich, 2015; O'Connor & Michaels, 1993) and co-constructing additional written or symbolic detail to reflect students' thinking. When students share their problem-solving strategies, teachers provide ample time for students to give complete and detailed explanations (Webb et al., 2019).

External representations of mathematics concepts are always present in the mathematics classroom in some form. These external representations may be verbal, visual, contextual, physical, or symbolic manifestations of students' thinking or mathematical concepts. Interpretation and use of various representations of mathematics concepts—and developing an understanding of the connections among those various representations—is of central importance in mathematics learning. Some scholars define *understanding* as the ability to recognize a mathematical idea embedded in different representations and to translate the idea from one representation to another (Haciomeroglu & Schoen, 2008; Lesh et al., 1987). Pape and Tschoshanov (2001) argued that the development of mathematical understanding is an interactive process in which internal and external representations are created and connected through language within a social context. In CGI classrooms, teachers engage students in dialogue about the connections among different mathematical representations to deepen their mathematical understandings (Hiebert & Carpenter, 1992; NCTM, 2014).

Interaction Among the Four Principles

The four principles are conceptualized and discussed separately in the previous sections, but these principles interact in many ways when put into practice. Peer interaction involves expressive and receptive communication. Peer interaction creates the opportunity to practice and acquire language skills and other important interpersonal communication skills that transcend mathematics. When teachers attend to student thinking, they require students to use language to express their thoughts—verbally or in writing. Problem-solving situations—wherein the teacher does not provide guidance about how to solve a problem—can help students develop confidence, agency, and autonomy, but they can also create important and unique opportunities for teachers to attend to student thinking in ways that telling students how to solve problems cannot. Furthermore, interaction and communication among problem solvers can enhance their problem-solving experiences and abilities. In other words, these four principles may be conceptualized as distinct aspects of instruction, but the implementation of the four principles can occur in many simultaneous and harmonic ways.

Alignment of Preexisting Instruments and the Four Principles

We examined several classroom observation instruments with the goal of finding one to measure teacher practices in mathematics instruction that would meet the following two criteria: (a) a focus on the four identified aspects of classroom instruction that are aligned with central principles of CGI, and (b) suitability for relatively efficient use at a large scale. We first reviewed instruments explicitly designed to study aspects of CGI-based classroom instruction (for example, see Empson & Jacobs, 2021; Fennema et al., 1996; Franke et al., 2009; Jacobs & Empson, 2016). These CGI-focused instruments yield useful insight into instructional practice aligned with CGI principles, but they require time-intensive data collection techniques, often involving repeated classroom observations and teacher interviews. Those techniques make them difficult to use on a large scale. Moreover, the conceptualization of CGI-aligned instruction has evolved over the decades to emphasize interaction among classroom members (Principle 3) in ways that were not represented in some of the earlier approaches to measuring instructional practice in CGI-related studies (Carpenter et al., 2015).

We also reviewed mathematics classroom observation instruments that were designed to enable use at a large scale, including the Instructional Quality Assessment (IQA; Boston, 2012), the Mathematical Quality of Instruction instrument (MQI; Hill et al., 2008), the Teaching for Robust Understanding framework (TRU; Schoenfeld, 2018), the Mathematics Scan (M-Scan; Walkowiak et al., 2018), and others (cf. Nava et al., 2019; Reinholz & Shah, 2018). Each of these preexisting instruments has its merits, and most had some degree of alignment with the CGI principles we wanted to measure.[1] Many measure cognitive demand, for example, which can be an indicator of the presence or absence of problem solving (Principle 1). Attending to student thinking and adjusting instruction using what is learned about the student (Principle 2) is a fundamental principle of CGI—that is, after all, what "cognitively guided instruction" means—and can be considered as a method of engaging in formative assessment. Indeed, some of the extant instruments address formative assessment, although each instrument interprets formative assessment differently. For example, the M-Scan contains a rubric titled Responsiveness to Student Thinking within its Mathematical Accuracy dimension, but that rubric focuses on the extent to which teachers handle misconceptions and errors. CGI emphasizes the more asset-oriented framing of student conceptions as partial understandings (Franke et al., 2020). The IQA assumes a specific lesson structure involving student engagement in a task and then a whole-class discussion of the task. It has rubrics that measure teacher moves related to teachers attending to student thinking (e.g., linking), but those rubrics are situated in class discussion and would need to be reinterpreted to include examples of teachers attending to student thinking in ways that occur outside a whole-class discussion.

Ultimately, we concluded that the preexisting measures were partially aligned with these four CGI principles, but no individual measure was fully aligned with them. We considered selecting from several different instruments to create a composite instrument. After trying that approach, we concluded that the shift in thinking required of the observers to use rubrics with different design features was too challenging. Consequently, we undertook the project of developing a novel observation instrument.

Design of the M-CLIPS Instrument

In the service of our goal for M-CLIPS to efficiently measure identified aspects of mathematics instruction in the context of large-scale studies, we imposed a few design constraints. As mentioned earlier, M-CLIPS would involve only observation and, therefore, would not require or allow the observer to interview teachers or ask them to explain their actions or decisions. M-CLIPS was intended to be useable in real-time classroom observation. This latter quality did not preclude the use of video, but it did create limitations on what could be coded. Each individual rubric was designed to measure a singular, observable topic associated with the four principles, and we aimed for a middle ground between a low-inference and a high-inference approach.

The M-CLIPS instrument is designed to measure each of the four principles by rating nine associated aspects of instruction using an observation of a single mathematics lesson.[2] Each of the nine constructs is briefly described in Table 1 with the name of its corresponding rubric. Although we think these nine observable features are likely to be related to a single

[1] A full review of the alignment of these instruments to the four CGI principles is beyond the scope of this article. Several instruments were discussed in a special issue of *ZDM* (Charalambous & Praetorius, 2018).

[2] Although the word "lesson" can refer to events that can take place over multiple days or are integrated into multiple components of the school day, our operational definition of *mathematics lesson* for the sake of observation and measurement is the time spent on mathematics instruction during the official window allocated to mathematics on a single school day.

Table 1

M-CLIPS Items, Their Associated Constructs, and Four Principles of CGI-Aligned Instruction

Construct (item name)	Definition of associated construct
	Principle 1: Problem solving
	Problem solving plays an integral role in the teaching and learning of mathematics.
Autonomy (AUTO)	The teacher facilitates students' use of their emerging understandings of mathematical ideas to devise their own ways to solve mathematics problems.
Variation (VAR)	Students exhibit variation and individual differences in their thoughts and approaches to solving mathematics problems.
Respect (RSPT)	The teacher demonstrates respect and appreciation for each and every student's abilities, perspectives, and contributions when they are solving mathematics problems.
Cognitive Complexity (CC)	Enacted mathematical tasks engage students in high levels of cognitive complexity.
	Principle 2: Cognition
	The teacher attends to students' mathematical thinking.
Attend (ATND)	The teacher attends to the details in students' mathematical thinking processes.
	Principle 3: Interaction
	Students engage with their peers' ideas about mathematics.
Teacher Support for Peer Interaction (TSPI)	The teacher creates opportunities and provides support for students to interact with one another's mathematical perspectives and ideas in order to advance their individual and collective understanding of mathematics.
Peer Interaction (PRI)	Students interact with one another to support the advancement of their individual and collective understanding of mathematics.
	Principle 4: Language
	The teacher supports students' understanding and meaningful use of the language of mathematics.
Express-Elaborate (EE)	The teacher actively supports students with developing and refining their abilities to express and elaborate mathematical ideas.
Connecting Representations (CR)	The teacher facilitates opportunities for students to examine conceptual connections among different external representations of mathematical concepts to support the advancement of their understanding and use of mathematics content and language.

latent trait (i.e., implementation of CGI in mathematics lessons), we defined each as distinctly as possible so that they could be conceptualized as nine independent components of instructional practice.

Each construct is rated on a 0–5-point rubric. The rubrics describe three levels: low (0–1), medium (2–3), and high (4–5). Within each level, a main level descriptor and indicators of observable behavior were developed through an iterative process of (a) a preliminary characterization of the aspect of instruction as aligned with CGI, (b) the identification of substantive ways that instruction related to the particular aspect may differ in classrooms of teachers who have and have not participated in CGI PD, (c) the description of observable indicators for the high, medium, and low levels, and (d) the revision of those descriptions on the basis of use with video data as well as feedback from external experts.

We attempted to calibrate the rubrics to enable M-CLIPS to discriminate among real examples of instructional practice that are not consistent with CGI principles, examples of instructional practice that are consistent with CGI principles but not fully incorporating all aspects of the principles into their practice, and exemplary implementation of the principles. Ratings of 0 or 1 indicate little or no evidence of alignment of instructional practices with CGI principles. Ratings of 2 or 3 indicate the existence of some elements of practice that are consistent with the relevant CGI principle. Ratings of 4 or 5 signify high levels of consistency with the given CGI principle. The highest rating (5) indicates that the lesson presents an exemplar of that particular aspect with respect to implementation of CGI principles in practice.

M-CLIPS is a rater-mediated instrument and requires observers to have a high degree of expertise in mathematics and mathematics teaching. Observers are trained to apply a common methodology to determining a rating for each rubric. The middle-ground approach between low-inference and high-inference design occasionally requires the observer to use a holistic approach to scoring.

Figure 1 presents the Autonomy rubric to provide an example of one of the nine rubrics. The Autonomy rubric is designed to measure the extent to which the teacher invites students to devise their own ways to solve mathematics problems. This is a fundamentally important distinction of a pedagogical approach in which students learn through problem solving.

Figure 1

M-CLIPS Autonomy Rubric

Autonomy (AUTO) **The teacher facilitates students' use of their emerging understandings of mathematical ideas to devise *their own ways* to solve mathematics problems.**	
High (4, 5) *Strong to expert implementation of principle*	The teacher consistently creates and sustains opportunities for students to solve mathematics problems in any way they can and that makes sense to them. All of the following are present: a. Substantive part(s) of the lesson involve students in attempting to solve problems that the teacher has not shown them how to solve. b. The teacher refrains from telling or showing students how they *should* solve problems, including when students experience difficulty.* c. *All* students have opportunity to solve problems in their own ways (and decide how they will try to solve the problem). d. The pace of instruction affords sufficient time for students to solve problems using strategies with understanding. *The teacher responds to student difficulties in ways that stimulate rather than dominate student thinking—for example, by helping students to understand the problem, reflect on their approach, or by making modifications to the problem that bring the mathematics closer to the students' understanding (e.g., reframe in meaningful context, modify numbers). The students remain in control of devising *their own ways* to solve problems.
Medium (2, 3) *Weak to moderate implementation of principle*	The teacher encourages or creates some opportunities for students to solve problems in ways of their own choosing, but those opportunities are limited. The High rating is not warranted for one or more of the following reasons: a. A substantial portion of the lesson involves explicit instruction on and/or practice of specific strategies for solving certain types of problems. b. The teacher allows flexibility for some students to solve problems in their own ways, while others are expected to solve problems in prescribed ways. c. The teacher provides insufficient time or tools for students to make sense of the problem. d. The teacher invites students to solve problems by selecting a strategy from a menu of options that were previously instructed by the teacher or curriculum materials. e. The teacher tells the students that they can use any strategy they want to use but subsequently directs them to use specific strategies that are preferred by the teacher. f. Teacher response to student struggle tends to direct or heavily guide students toward the teachers' thinking (rather than stimulating the students' thinking).
Low (0, 1) *Contradicts principle*	The teacher does not encourage students to solve problems by using strategies of their own choosing or invention. The teacher consistently shows or tells students how to solve problems (i.e., explicit instruction) and expects students to practice and use the instructed strategy. The teacher directs students to use a particular solution method that is known, prescribed, or heavily prompted by the teacher or textbook. A small amount of divergent thinking may be tolerated, such as when it involves ways to complete intermediate steps in an overall strategy. (1) The teacher insists that students use a specific way of solving a given problem or actively discourages invented strategies or other forms of divergent thinking. Independent thinking or alternative methods of solving a problem are ignored, discouraged, or marginalized. (0)

Low levels of autonomy do not present examples of applying the problem-solving principle to mathematics instruction. A rating of 1 indicates instruction in which the teacher shows or tells students how to solve problems and expects students to practice and use the instructed strategy.

At the medium level, teachers offer some autonomy with problem solving. Medium ratings may indicate a teacher who has released some amount of control over how students approach problems. Maybe they do this for only one or two problems, but not for every problem, or maybe they create opportunities for students to solve problems in their own ways but then respond to student struggle with a more directive approach to teaching.

The high level is characterized by the teacher's consistent and sustained effort at facilitating students' autonomy in their problem solving and efforts to solve problems using strategies with understanding. When a lesson is rated as having a high level of autonomy, substantive parts of lessons involve all students in solving problems that the teacher has not shown them how to solve, and evidence suggests that the teacher's invitation to "solve it in any way that makes sense to you" is genuine. The teacher who supports autonomy budgets sufficient time for students to grapple with problems and, even when students experience difficulty, refrains from telling or showing students how they should solve problems.

The complete set of M-CLIPS rubrics is available for viewing and download online (https://osf.io/mk657). An interpretation and use statement (Carney et al., 2022) can be found in the Appendix.

Methods and Materials

Data for the present study consist of ratings of videos of mathematics lessons collected as part of a larger randomized controlled trial of the effect of a CGI PD program on teachers, teaching, and students.

Setting and Sample

Teachers from 22 elementary schools from two Florida school districts participated in the study. The first district served approximately 200,000 students in a mix of urban, suburban, and rural settings. The second district, adjacent to the first, was a suburban district that served approximately 60,000 students. Both districts served a diverse student population, with more than half of their students belonging to racially or ethnically minoritized groups.

Teacher participants could consent to participate in the larger study without consenting to participate in the video study. To be eligible for the video study, a teacher had to teach first- or second-grade mathematics in one of the 22 schools, the teacher had to consent to participate in the video study, and at least two thirds of the students in the teacher's class had to have positive, informed consent for the video component of the study. Only teachers who participated in both years of the study were eligible for random selection for the video component in the 2nd year of the study. Sampling procedures used random selection and were conducted separately in each year. We selected a random sample of up to two teachers per grade level in each school with eligible teachers/classrooms.

The resulting sample for the video study consisted of a total of 42 teachers/classrooms in the 1st year and 47 teachers/classrooms in the 2nd year. Of those, 37 teachers were video recorded on two occasions in the 1st year. By design, no teachers were video recorded twice during the 2nd year. In total, 22 teachers were video recorded in both years. Tables 2 and 3 give teacher demographics by condition and grade for teachers in the sample for the 1st and 2nd years of the video study, respectively.

At least three randomly assigned observers rated each of the 126 available videos. In cases in which more than three observers rated a video, three of the observers' ratings were randomly selected from the full set of complete ratings and used in the analyses. This ensured that each video had exactly three sets of ratings and that each set of ratings was weighted equally. We consider this set of exactly three observations for each of the 126 videos to be the full sample for the purpose of the current study.

Table 2

Teacher Demographics for the Video Sample in Study Year 1

	Sample size by condition		Sample size by grade		Total
Characteristic	CGI ($n = 26$)	Comparison ($n = 16$)	Grade 1 ($n = 22$)	Grade 2 ($n = 20$)	($N = 42$)
		Years of teacher experience			
3 or fewer	8	4	6	6	12
4 or more	18	12	16	14	30
		Gender			
Female	24	16	21	19	40
Male	2	0	1	1	2
		Race/ethnicity			
Asian/Pacific Islander	0	0	0	0	0
Black	2	0	1	1	2
White	23	15	21	17	38
Hispanic	2	1	0	7	3

Note. No demographic information was missing for teachers in the sample. Race/ethnicity totals may exceed sample size because teachers could select all that apply.

Table 3

Teacher Demographics for the Video Sample in Study Year 2

Characteristics	Sample size by condition: CGI ($n=24$)	Sample size by condition: Comparison ($n=23$)	Sample size by grade: Grade 1 ($n=27$)	Sample size by grade: Grade 2 ($n=20$)	Total ($N=47$)
Years of teacher experience					
3 or fewer	3	5	3	5	8
4 or more	21	18	24	15	39
Gender					
Female	24	23	27	20	47
Male	0	0	0	0	0
Race/ethnicity					
Asian/Pacific Islander	0	0	0	0	0
Black	1	0	0	1	1
White	22	21	27	16	43
Hispanic	2	3	2	3	5

Note. No demographic information was missing for teachers in the sample. Race/ethnicity totals may exceed sample size because teachers could select all that apply.

The CGI PD Program

The CGI PD program was designed and delivered by Teachers Development Group (http://www.teachersdg.org) under the direction of Linda Levi, who was the Director of CGI Initiatives at Teachers Development Group at the time. The entire program consisted of 8 days of teacher workshops per year (4 days in summer, 2 in fall, 2 in winter) for 3 years. Research-based frameworks for problem types and strategies are salient features of CGI and were referred to throughout the program. The program provided teachers with opportunities to learn about robust and predictable developmental progressions that described how children's knowledge and understanding of mathematics become more sophisticated over time. The content of the workshops focused on number (whole numbers and base ten) and operations as well as algebraic thinking. Workshop leaders guided teachers to construct their own understanding of the frameworks through in-depth analysis and discussion of student work and student thinking as they solved mathematics problems. Teachers participated in two classroom-embedded workshop days (Levi, 2017; Nielsen et al., 2016; Schoen & Champagne, 2017) per year and were introduced to the purposeful pedagogy model (Jaslow & Evans, 2012). 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. Schoen, Bray, et al. (2022) described the CGI program in more detail.

Data Collection

During the 1st year, two mathematics lessons on 2 separate days were video recorded in each teacher's classroom between mid-December and mid-February. Teachers in the CGI condition had all attended 6 of 8 days of the 1st year of the CGI PD workshops before the lessons were recorded.

During the 2nd year, one mathematics lesson was video recorded in each teacher's classroom between mid-December and mid-February. At the point of video recording, all teachers in the CGI condition whose classrooms were videotaped had attended 8 days of the 1st year of CGI PD and 6 of the 8 days of the 2nd year of PD. To avoid bias stemming from different times of the school year or days of the week, observation days were scheduled so that CGI and comparison classrooms were video recorded on the same days.

Participating teachers knew when lessons would be observed and video recorded. For all video recorded lessons, teachers were asked to teach in the same manner that they would have taught if an observer were not present. Approximately 90% of the lessons (113 out of 126) focused on number and operations. The other 13 lessons focused on measurement and were close to evenly divided between the CGI and comparison classrooms (six CGI, seven comparison).

The participating teachers indicated the beginning and end of the lesson to the camera operator. The duration of the lessons in both conditions was usually about 60 min. The minimum duration in CGI and comparison classrooms was 27 and 30 min, respectively. The median and mean for both conditions were between 59 and 61 min. The maximum duration in the CGI and comparison classrooms was 98 and 95 min, respectively.

Table 4

Summary of the Rubric Ratings for the Full Sample

Rubric	*N*	Min	Max	Mean	*SD*
Autonomy (AUTO)	378	0	5	2.59	1.52
Variation (VAR)	378	0	5	2.66	1.45
Respect (RSPT)	378	0	5	2.63	1.45
Cognitive Complexity (CC)	378	0	5	2.65	1.36
Attend (ATND)	378	0	5	2.71	1.36
Teacher Support for Peer Interaction (TSPI)	378	0	5	2.47	1.17
Peer Interaction (PRI)	378	0	5	2.29	1.07
Express-Elaborate (EE)	378	0	5	2.63	1.33
Connecting Representations (CR)	378	0	5	2.94	1.22

The first 20 videos in the 1st year were filmed by members of the research team. The remaining 106 videos were filmed by professional videographers using two vantage points (one stationary camera positioned for a broad view of the students and presentation space in the classroom, another camera following the teacher) and two microphones (one lapel microphone on the teacher, one shotgun microphone on the fixed camera). The two videos per class were combined into a single frame using picture-in-picture technology, and the combined frame was used for rating purposes.

Observer Training and Certification

Fourteen potential observers, including five coauthors of this article, participated in 12 hr of training on the M-CLIPS instrument. Training was distributed across four sessions in a synchronous online instruction format. Three of the M-CLIPS authors served as the trainers for these sessions. Between sessions, potential observers completed additional practice scoring exercises.

Following training, potential observers who were not members of the author team completed a certification process in which they were required to apply the M-CLIPS instrument to two video-recorded mathematics lessons. Specifically, observers rated the lessons using each of the nine M-CLIPS rubrics and gave detailed justifications for their ratings for each; these were blind-reviewed by the author team. Observers were certified if their ratings were within the acceptable ranges as determined by the maximum and minimum ratings of the author team, and their written justifications demonstrated high degrees of understanding of the principles and rubrics. Of the nine potential observers who were not members of the M-CLIPS author team, five were certified on their first certification attempt, three were certified through a second certification attempt after two additional training sessions, and one potential observer was never certified because of a failure to meet the criteria during either attempt.

Certified observers were then assigned videos to rate. Observers viewed an assigned lesson once without pausing while simultaneously writing field notes on paper. They then considered evidence from their field notes concerning each of the nine rubrics that compose the M-CLIPS instrument and the scoring guidance protocol to assign an ordinal rating of 0–5 for each rubric. These ratings were documented on a paper rating sheet and entered into a database using FileMaker (version 18) software. Observers were not made aware of the treatment conditions of the classrooms in the videos.

M-CLIPS Measurement Properties

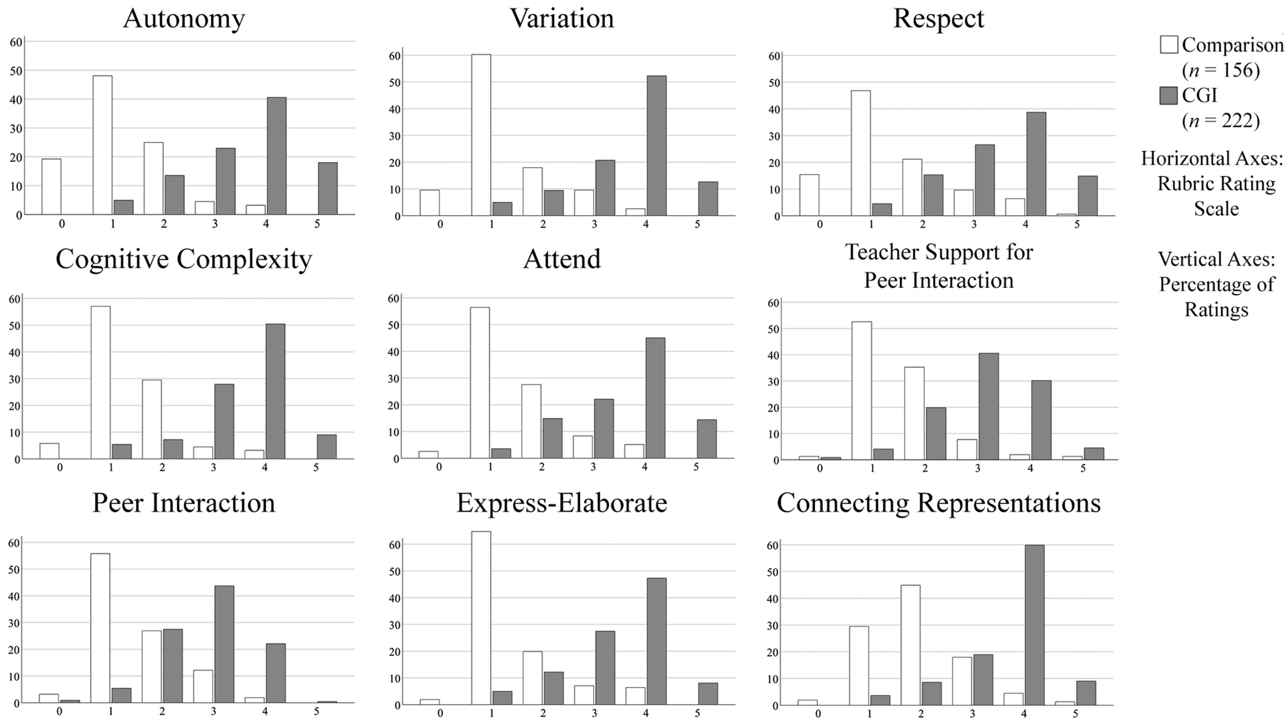
We assume unidimensionality of the underlying construct (i.e., implementation of CGI principles) as measured by the nine M-CLIPS rubrics, and analyses of empirical data presented elsewhere (Schoen et al., 2021) supported that assumption. Employing a many-facet Rasch model using the sample of 47 videos from the 2nd year and the pilot version of M-CLIPS, reliability for three facets (i.e., observers, lessons, and items) was estimated to be .78, .98, and .98, respectively.

Data Analysis and Results

We now turn our focus to the analysis of M-CLIPS data and interpretation of the results to answer our two research questions. Because of the study's exploratory nature, some of the interim results influenced subsequent data analysis decisions. Thus, we decided to describe our analysis methods and the associated results in the same section. All descriptive statistics were produced using IBM SPSS Statistics (Version 26) software.

Figure 2

Sample Distribution of the Ratings on Each M-CLIPS Rubric Disaggregated by Treatment Condition



Ratings on Individual Rubrics

We conducted initial exploratory analyses to examine the individual ratings on each of the nine M-CLIPS rubrics. Table 4 presents descriptive statistics. Observers used the full range of the rating scale on all nine rubrics with the videos in this sample, suggesting that the range of ratings for each rubric is reasonably well calibrated to the range of practices that may be observed in classrooms. The Autonomy ratings had the most variation. Peer Interaction had the lowest overall mean rating and the least variation. Teacher Support for Peer Interaction had the second lowest mean and second least amount of variation. Connecting Representations had the highest overall ratings.

We disaggregated the data by treatment condition (i.e., CGI, comparison) and examined the frequency distributions of the ratings for each of the nine rubrics (Figure 2). More than half (59%) of the videos were from CGI-condition classrooms, so we standardized the results by reporting the percentage within each condition to adjust for the imbalance between treatment conditions and enable informal comparison.

The center of the distribution within each rubric was substantially higher for the videos of classrooms in the CGI condition than for those in the comparison condition. The modal rating in the comparison condition was 1 for eight of the nine rubrics (the modal rating was 2 for Connecting Representations). In sharp contrast, the modal rating in the CGI condition was 4 in eight of the nine rubrics (the modal rating was 3 for Peer Interaction).

The frequency of the ratings for each of the M-CLIPS rubrics, disaggregated by condition, is given in Table 5. This visualization of the data again also shows a clear distinction between the observed instructional practice in the lessons implemented by the CGI and comparison groups. Almost all (95%) of the ratings in the CGI classrooms were in the medium to high levels (ratings 2–5), indicating at least a weak implementation of the CGI principles. Conversely, we found that 59% of the ratings of lessons delivered by teachers in the comparison condition were in the low levels (i.e., ratings of 0 or 1), indicating that instructional practices in classrooms in the control group did not present evidence of alignment with the CGI principles.

Overall, patterns across the rubrics varied between treatment groups. Whereas the two Interaction rubrics were the most difficult (i.e., lowest mean rating) of rubrics within the CGI condition, those two rubrics were mid-level difficulty within the comparison sample. Within the comparison condition, all four Problem Solving rubrics had a lower mean rating than the Interaction rubrics. Autonomy had the lowest mean rating and the highest incidence of ratings of 0 within the comparison group, implying that teachers in the comparison condition consistently told students how to solve problems. Connecting Representations had the highest mean rating in both conditions, implying that multiple representations of mathematics

Table 5

Frequency (and Proportion) of Ratings for Each M-CLIPS Rubric, Disaggregated by Treatment Condition

	Comparison condition ($n = 156$ ratings, three observer ratings per rubric for 52 videos)							CGI condition ($n = 222$ ratings, three observer ratings per rubric for 74 videos)						
		Low		Medium		High			Low		Medium		High	
Rubric	Mean	0	1	2	3	4	5	Mean	0	1	2	3	4	5
AUTO	1.24	30 (.19)	75 (.48)	39 (.25)	7 (.04)	5 (.03)	0 (.00)	3.53	0 (.00)	11 (.05)	30 (.14)	51 (.23)	90 (.41)	40 (.18)
VAR	1.35	15 (.10)	94 (.60)	28 (.18)	15 (.10)	4 (.03)	0 (.00)	3.58	0 (.00)	11 (.05)	21 (.09)	46 (.21)	116 (.52)	28 (.13)
RSPT	1.47	24 (.15)	73 (.47)	33 (.21)	15 (.10)	10 (.06)	1 (.01)	3.44	0 (.00)	10 (.05)	34 (.15)	59 (.27)	86 (.39)	33 (.15)
CC	1.42	9 (.06)	89 (.57)	46 (.29)	7 (.04)	5 (.03)	0 (.00)	3.50	0 (.00)	12 (.04)	16 (.07)	62 (.28)	112 (.50)	20 (.09)
ATND	1.57	4 (.03)	88 (.56)	43 (.28)	13 (.08)	8 (.05)	0 (.00)	3.52	0 (.00)	8 (.04)	33 (.15)	49 (.22)	100 (.45)	32 (.14)
TSPI	1.60	2 (.01)	82 (.53)	55 (.35)	12 (.08)	3 (.02)	2 (.01)	3.09	2 (.01)	9 (.04)	44 (.20)	90 (.41)	67 (.30)	10 (.05)
PRI	1.54	5 (.03)	87 (.56)	42 (.27)	19 (.12)	3 (.02)	0 (.00)	2.82	2 (.01)	12 (.05)	61 (.27)	97 (.44)	49 (.22)	1 (.00)
EE	1.52	3 (.02)	101 (.65)	31 (.20)	11 (.07)	10 (.06)	0 (.00)	3.40	0 (.00)	11 (.05)	27 (.12)	61 (.28)	105 (.48)	18 (.08)
CR	1.97	3 (.02)	46 (.29)	70 (.45)	28 (.18)	7 (.04)	2 (.04)	3.62	0 (.00)	8 (.04)	19 (.09)	42 (.19)	133 (.60)	20 (.09)
Total	1.52	95 (.07)	735 (.52)	387 (.28)	127 (.09)	55 (.04)	5 (.00)	3.39	4 (.00)	92 (.05)	285 (.14)	557 (.28)	858 (.43)	202 (.10)

ideas were used in both conditions, although the average rating for that rubric was considerably higher in the CGI group than in the control group.

The incidence of extreme ratings also differed by treatment condition. A rating of 5 was assigned 201 times in the CGI sample, which amounted to 10% of the ratings within that condition. The 5s were most common in the Autonomy, Variation, Respect, and Cognition rubrics. Conversely, a rating of 5 was assigned only five times (< 0.5% of the ratings) in the comparison condition. The five ratings of 5 in the comparison condition occurred in the Respect, Teacher Support for Peer Interaction, and Connecting Representations rubrics. Meanwhile, a rating of 0 occurred only four times for lessons in the CGI group, but 0s were assigned 95 times to lessons in the control group. Most of the 0s in the control group occurred in the four rubrics associated with the Problem Solving principle (Principle 1). No lessons in the CGI group received a rating of 0 on rubrics associated with Principles 1, 2, or 4. Two instances of a rating of 0 for each of the two rubrics associated with Principle 3 (i.e., Teacher Support for Peer Interaction and Peer Interaction) occurred in the CGI condition.

M-CLIPS Total Scores

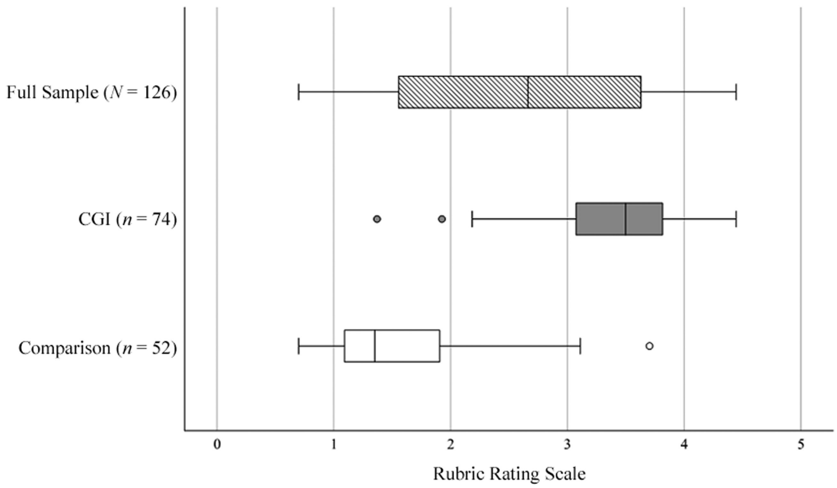
After looking at the data for individual rubrics, we focused on generating and analyzing overall M-CLIPS scores. We first computed the arithmetic mean of the nine rubric ratings for each video and observer, resulting in an M-CLIPS score. This scoring method produces scores on a scale with a theoretical range that matches the raw rating for each rubric and supports the use of the cut score of 2 for meaningful interpretation of the M-CLIPS score. This process resulted in three M-CLIPS scores per lesson. We then computed the arithmetic mean of those three scores so that each of the 126 videos had exactly one score—the mean value of the three M-CLIPS scores. For brevity, we simply refer to this value as the M-CLIPS score. Figure 3 presents a visualization of the data for the full sample and for the CGI and comparison subsamples.[3]

When disaggregated into the CGI and comparison conditions, the boxplots reveal clear differences in the M-CLIPS scores in the CGI and comparison condition. Less than one quarter of the data overlap. More than three quarters of the M-CLIPS scores for comparison lessons were lower than the nonoutlier M-CLIPS scores in the CGI group. Conversely, three quarters of the M-CLIPS scores for CGI-condition lessons were higher than the maximum score in the comparison group. Although the full range of observed ratings was used at the rubric level, we do not see any obvious ceiling or floor effects in the M-CLIPS scores.

Figure 3 uses the standard method of identifying outliers that are more than one-and-a-half times the interquartile range from the median. By this method, values lower than 2 in the CGI group were considered outliers, and only the outliers in

Figure 3

Boxplots of M-CLIPS Scores for Full Sample and Disaggregated by Treatment Condition



Note. M-CLIPS scores are the arithmetic mean of the three M-CLIPS scores calculated from the rubric ratings assigned by the three observers for each video.

[3] We note that the measures of variability (i.e., standard deviation, range, and interquartile range [IQR]) are affected by the decision to report the mean of the three scores rather than all three observations per video; the resulting variability statistics are therefore smaller in magnitude than they would have been had we used all three scores for each video to calculate the estimate. Outliers are identified using the standard rule of the median ±1.5 × IQR.

the CGI group had M-CLIPS scores lower than 2. More than three quarters of the M-CLIPS scores in the comparison condition were less than 2. These scores imply that instructional practices in the CGI group were generally consistent with CGI principles, but the instructional practices in the comparison group were not.

Testing for Group-Level Differences in M-CLIPS Scores

To examine the statistical evidence for the ability of M-CLIPS to detect differences in instructional practice, we compared the average M-CLIPS ratings between the CGI and comparison groups. The research design created several dependencies in the data. Some teachers were selected for the lesson observation only in Year 1, some in both years, and others only in Year 2. Lessons taught by the same teacher might not be independent. Each video was rated by three observers, creating another dependency. Teachers in the CGI condition in the Year 1 data had participated in one year of the CGI program, whereas teachers in the CGI condition in the Year 2 data had participated in 2 years of the CGI program, thereby creating a systematic difference between the 2 years of data. Last, teachers are nested within schools, and lessons taught by teachers in the same schools may correlate. To address the multiple levels of dependencies in the data, we specified multilevel models for our analyses. We also separated the 2 years of data to test for group differences in each year of the study to eliminate dependencies stemming from repeated teacher participation in data collection (within or across years) and acknowledge the differences in intervention dosage teachers received in Years 1 and 2.

We specified one model per year to produce estimates for the difference in average M-CLIPS ratings between CGI treatment and control groups. The mean of the nine M-CLIPS rubric ratings served as the dependent variable in both models. The independent variable was the treatment indicator (1 = CGI, 0 = comparison). We specified hierarchical linear models (HLMs) using the lmerTest package in R (v3.1–3; Kuznetsova et al., 2017) to account for dependencies stemming from nested observations. Raters were nested within lessons, lessons were nested within teachers, and teachers were nested within schools (Raudenbush & Bryk, 2002). For the Year 1 model, we specified a four-level HLM and included random intercepts for lessons, teachers, and schools. For Year 2, we specified a three-level HLM and included random intercepts for lessons and schools (we did not need to account for teacher random effects in Year 2 because each teacher in the Year 2 sample was observed exactly once in that year). Estimates for the treatment indicator were positive in each wave of data collection (Year 1: $b = 1.85$, $t = 8.08$, $p < .001$; Year 2: $b = 2.05$, $t = 12.76$, $p < .001$). This indicates that, on average, M-CLIPS ratings in the CGI group were higher than those of the control group in both years of data collection, and the differences were statistically significant.

For each year, we calculated an intraclass correlation (ICC) to evaluate the interrater reliability of M-CLIPS ratings on lessons using the ICC package in R (v2.4.0; Wolak et al., 2012). We computed the ICC(1,1) using single measurements, absolute agreement, and a one-way random effects model because lessons were randomly assigned to different sets of raters (Koo & Li, 2016). ICCs of raters on lessons were .79 in Year 1 (95% CI: [.72, .85]) and .85 in Year 2 (95% CI: [.77, .91]), indicating acceptable levels of interrater reliability in both years.

Discussion

The purpose of this study was to use M-CLIPS to examine the extent to which teachers participating in the CGI and comparison groups were implementing observable features of mathematics instruction in a manner consistent with CGI principles. Results reported in this article provide some evidence to support the argument that M-CLIPS can be used to detect the implementation of CGI in mathematics lessons. The arithmetic mean of the ratings is an acceptable method of scoring, yielding a continuous variable with a theoretical minimum of zero and a maximum of 5. As described in the interpretation and use statement in the Appendix, scores less than 2 indicate little or no evidence of implementation of CGI. Scores of 2 or greater imply that the teacher is implementing the practices described in the nine rubrics at a level that would be consistent with at least a novice level of implementation of CGI. Scores in the range of 4–5 indicate exemplary practice of CGI. Our results indicate that the instructional practices of teachers in the CGI group were consistent with CGI principles in the 1st and 2nd year of their participation in the CGI PD program, but the instructional practices of the majority of teachers in the comparison groups were not.

M-CLIPS scores detected statistically significant differences in CGI-related instructional practices between groups of teachers assigned to CGI or a comparison condition. These group differences were detected in the 1st year of the program and again during the 2nd year. These findings provide some initial evidence that the M-CLIPS instrument may be able to discriminate among instructional practices that are consistent with CGI and those that are not.

We observed the most drastic distinction between the CGI condition and the comparison condition with respect to the four rubrics associated with problem solving, indicating that classrooms in which teachers had participated in the CGI PD were more consistent with a problem-solving approach to teaching and learning. The four rubrics associated with problem solving focused on student autonomy, variation in approaches students used to solve problems, the teacher demonstrating respect for students, and cognitive complexity of the enacted mathematical tasks. With four of the nine rubrics associated with the problem-solving principle contributing to the overall M-CLIPS score, the high scores on the problem-solving-oriented rubrics in the CGI group probably greatly influenced the treatment–control contrast.

The CGI program encourages the use of problem solving as an integral component of mathematics instruction, and these data provide evidence that teachers responded positively to the CGI program by implementing more problem solving in their classrooms. At the same time, we found that instruction in comparison classrooms was generally inconsistent with a problem-solving approach. The data suggest that the teachers in the CGI condition were learning to pose problems with higher cognitive complexity and maintain higher cognitive complexity as the lesson unfolded, the latter of which is atypical of U.S. mathematics instruction (Stigler & Hiebert, 2009). Through the M-CLIPS ratings from the Autonomy, Variation, and Respect rubrics, we can see that the teachers in the CGI program were using instructional practices that provide students with more agency and autonomy in their mathematics learning and may be transitioning toward a more asset-oriented approach to teaching.

Data from the Variation rubric indicate that students in comparison classrooms used a narrower set of strategies to solve problems than those in CGI classrooms. We think this factor is probably not something teachers could change for the day they were observed, especially when they did not know that this aspect of their class would be monitored. That is, students would probably not increase the variation in the types of strategies they used to solve problems simply in response to being observed. Thus, this indicator suggests that teachers in CGI classrooms encouraged more agency and autonomy in their classrooms, even on the days preceding the observed lessons.

The data from the Attend rubric imply that teachers in the comparison group primarily engaged in classroom assessment that focused on whether students produced correct answers or used a prescribed procedure correctly. They rarely asked questions to probe student thinking or understanding. The data also imply that teachers in the CGI program, conversely, were attending to students' thinking processes. These results imply that teachers could make these shifts in their instructional practice toward a problem-solving approach and attend to student thinking within the first 2 years of their participation in the PD program.

The rubrics associated with the Interaction principle appear to be the most difficult for teachers in the CGI condition to implement at high levels, as evidenced by lower mean scores on these rubrics compared with those associated with other principles. Orchestrating classroom interaction among students to engage them in discussion about their own mathematical ideas may require more time or a different intervention from the one in this study.

The most common rating for eight of the nine rubrics in comparison lessons was 1, but the most common rating for Connecting Representations in comparison lessons was 2. This suggests that multiple forms of nonverbal representations of mathematics are usually present in the practice-as-usual condition, at least to some minimal degree.

Limitations and Next Steps

The four principles we describe are not isomorphic to the principles described in the definitive CGI books because we limited our scope to those principles of CGI that could be observed during mathematics instruction. For example, one fundamentally important component of CGI implementation—adjusting the instructional plan in response to student interests, experience, and understanding—is conspicuously absent from the set of principles. We made the difficult choice not to measure that component because we thought doing so would not be feasible without spending considerable time in a given classroom and engaging in conversation with teachers about their decisions.

We created four, one, two, and two rubrics corresponding to each of the four CGI principles, respectively. As a result, the method of scoring presented herein weights the problem-solving principle (Principle 1) more heavily than the other principles in the aggregate score. Conversely, the cognition principle (Principle 2) is weighted the least.

Generalizations of these results should be appropriately limited. This study was conducted only in first- and second-grade classrooms, and we cannot know from the current study whether these results would be replicated in other grade levels. We studied only one CGI PD model, and the results of this study cannot be generalized to the other CGI PD models that exist. Teachers did not know what we were looking for when we conducted the observations, but they did know that we were coming. We cannot know whether the instructional practice that we observed is representative of the other days of mathematics instruction in the sample classrooms. However, we can reasonably expect that all teachers, irrespective of condition, probably attempted to demonstrate their best effort at doing what they thought the observers wanted to see.

The CGI PD program may affect individual teachers differently. This study was not designed to enable fine-grained distinctions among individuals. In addition, the current study does not estimate the size of the impact on instructional practice; it tests only for group differences. Further research is needed to estimate the effect of CGI on instructional practice, the effects of these facets of instructional practice on student learning, and the heterogeneity in those effects.

Observations occurred through video of classroom instruction. We have not yet used M-CLIPS in real-time observations. Real-time observations may provide the observers with a different perspective than the video camera can, especially for some rubrics (e.g., Variation, Peer Interaction). The video camera possibly limited the ability of the raters to observe interaction, thereby introducing a downward bias to the ratings. Because we had, at most, only one or two lessons observed in each classroom, our ability to examine the test–retest reliability or stability/error in measurement within a single classroom is also limited at this time. Scoring procedures for the M-CLIPS assumed unidimensionality of the underlying construct (i.e., implementation of CGI principles). Due diligence is needed to evaluate the validity of this assumption and other aspects of structural validity (Flake et al., 2017). These are all important areas for future research.

Conclusions

This study offers a conceptual framework for observable features of mathematics instructional practices that are aligned with CGI principles. These instructional practices are also consistent with aspects of mathematics instruction that many in the mathematics education community consider to be desirable and productive. We consider the four principles to be substantive and durable, but they are not intended to be comprehensive or limiting, and they may ultimately be subject to reinterpretation.

On the basis of this study, M-CLIPS appears to be able to detect group-level differences between the instructional practice of teachers who have been involved in the CGI PD program and those who have not. Results indicate that teachers in the CGI program were implementing mathematics instruction in ways that are consistent with CGI principles, but teachers in the control group mostly implemented mathematics lessons in ways that were inconsistent with CGI principles. This study supports the assertion that M-CLIPS measures the kind of instructional shifts expected as a result of CGI PD.

This study represents a major step toward conceptualizing and measuring some key instructional practices that are consistent with CGI principles. We are hopeful that this work may ultimately provide both empirical data to help identify observable features of mathematics instructional practice that improves student learning and a method of measuring those practices that can be used to support large-scale research and further discovery regarding approaches to mathematics instruction that support student learning.

References

Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal*, *113*(1), 76–104. https://doi.org/10.1086/666387

Bray, W. S., & Blais, T. V. (2017). Stimulating base-ten reasoning with context. *Teaching Children Mathematics*, *24*(2), 120–127. https://doi.org/10.5951/teacchilmath.24.2.0120

Bray, W. S., & Maldonado, L. A. (2018). Foster fact fluency with number string discussions. *Teaching Children Mathematics*, *25*(2), 86–93. https://doi.org/10.5951/teacchilmath.25.2.0086

Carney, M. B., Bostic, J., Krupa, E., & Shih, J. (2022). Interpretation and use statements for instruments in mathematics education. *Journal for Research in Mathematics Education*, *53*(4), 334–340. https://doi.org/10.5951/jresematheduc-2020-0087

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. https://doi.org/10.1086/461846

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

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

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), 499–531. https://doi.org/10.3102/00028312026004499

Carpenter, T. P., Franke, M. L., Johnson, N. C., Turrou, A. C., & Wager, A. A. (2017). *Young children's mathematics: Cognitively guided instruction in early childhood education.* Heinemann.

Carpenter, T. P., Franke, M. L., Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Heinemann.

Chapin, S. H., & O'Connor, C. (2007). Academically productive talk: Supporting students' learning in mathematics. In W. G. Martin, M. E. Strutchens, & P. C. Elliott (Eds.), *The learning of mathematics* (pp. 113–128). National Council of Teachers of Mathematics.

Charalambous, C. Y., & Praetorius, A.-K. (Eds.). (2018). Studying instructional quality in mathematics through different lenses: In search of common ground [Special issue]. *ZDM*, *50*(3).

Empson, S. B., & Jacobs, V. R. (2008). Learning to listen to children's mathematics. In P. Sullivan & T. Wood (Eds.), *The international handbook of mathematics teacher education: Vol. 1. Knowledge and beliefs in mathematics teaching and teaching development* (pp. 257–281). Sense.

Empson, S. B., & Jacobs, V. R. (2021). Exploring teachers' responsiveness to children's fraction thinking and relationships to fraction achievement. In D. Olanoff, K. Johnson, & S. Spitzer (Eds.), *Proceedings of the 43rd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1449–1458). PME-NA.

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. https://doi.org/10.5951/jresematheduc.27.4.0403

Fennema, E., Franke, M. L., Carpenter, T. P., & Carey, D. A. (1993). Using children's mathematical knowledge in instruction. *American Educational Research Journal*, *30*(3), 555–583. https://doi.org/10.3102/00028312030003555

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Forman, E. A., McCormick, D. E., & Donato, R. (1997). Learning what counts as a mathematical explanation. *Linguistics and Education*, *9*(4), 313–339. https://doi.org/10.1016/S0898-5898(97)90004-8

Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in Everyday Mathematics classrooms. *Journal for Research in Mathematics Education*, *30*(2), 148–170. https://doi.org/10.2307/749608

Franke, M., McMillan, B., Johnson, N. C., & Turrou, A. C. (2020). Connecting research on children's mathematical thinking with assessment: Toward capturing more of what children know and can do. In S. Ryan, M. E. Graue, V. L. Gadsden, & F. J. Levine (Eds.), *Advancing knowledge and building capacity for early childhood research* (pp. 55–73). American Educational Research Association.

Franke M. L., Turrou, A. C., Webb, N. M., Ing, M., Wong, J., Shim, N., & Fernandez, C. (2015). Student engagement with others' mathematical ideas: The role of teacher invitation and support moves. *The Elementary School Journal*, *116*(1), 126–148. https://doi.org/10.1086/683174

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education*, *60*(4), 380–392. https://doi.org/10.1177/0022487109339906

Gresalfi, M. S. (2009). Taking up opportunities to learn: Constructing dispositions in mathematics classrooms. *Journal of the Learning Sciences*, *18*(3), 327–369. https://doi.org/10.1080/10508400903013470

Haciomeroglu, E. S., & Schoen, R. C. (2008). Mathematical continuity: Identifying, exposing, and closing the gaps of understanding. *Dimensions in Mathematics*, *28*(1), 9–12.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*(5), 524–549. https://doi.org/10.2307/749690

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). National Council of Teachers of Mathematics.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, *25*(4), 12–21.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*(4), 430–511. https://doi.org/10.1080/07370000802177235

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2015). Student participation in elementary mathematics classrooms: The missing link between teacher practices and student achievement? *Educational Studies in Mathematics*, *90*(3), 341–356. https://doi.org/10.1007/s10649-015-9625-z

Jacobs, V. R., & Empson, S. B. (2016). Responding to children's mathematical thinking in the moment: An emerging framework of teaching moves. *ZDM*, *48*(1), 185–197. https://doi.org/10.1007/s11858-015-0717-0

Jacobs, V. R., Empson, S. B., Jessup, N. A., Dunning, A., Pynes, D., Krause, G., & Franke, T. M. (2022). Profiles of teachers' expertise in professional noticing of children's mathematical thinking. *Journal of Mathematics Teacher Education*. Advance online publication. https://doi.org/10.1007/s10857-022-09558-z

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, A. L. (2012). *Purposeful pedagogy and discourse instructional model: Student thinking matters most*. Arkansas Department of Education.

Kazemi, E., & Hintz, A. (2014). *Intentional talk: How to structure and lead productive mathematical discussions*. Stenhouse.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lesh, R., Post, T. R., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 33–40). Lawrence Erlbaum.

Levi, L. (2017, October 19). Classroom embedded work: An alternative to observation lessons. *Teaching is problem solving*. https://teachingisproblemsolving.org/classroom-embedded-work-an-alternative-to-observations-lessons

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, *1*(1), 12–21. https://doi.org/10.1016/j.lcsi.2012.03.001

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. In L. B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 347–362). American Educational Research Association. https://doi.org/10.3102/978-0-935302-43-1_27

Moschkovich, J. N. (2015). Scaffolding student participation in mathematical practices. *ZDM*, *47*(7), 1067–1078. https://doi.org/10.1007/s11858-015-0730-3

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. https://www.nctm.org/PSSM

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*.

Nava, I., Park, J., Dockterman, D., Kawaski, J., Schweig, J., Quartz, K. H., & Martinez, J. F. (2019). Measuring teaching quality of secondary mathematics and science residents: A classroom observation framework. *Journal of Teacher Education*, *70*(2), 139–154. https://doi.org/10.1177/0022487118755699

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

O'Connor, M. C., & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly*, *24*(4), 318–335. https://doi.org/10.1525/aeq.1993.24.4.04x0063k

Pape, S. J., & Tchoshanov, M. A. (2001). The role of representation(s) in developing mathematical understanding. *Theory Into Practice*, *40*(2), 118–127. https://doi.org/10.1207/s15430421tip4002_6

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology*, *98*(1), 209–218. https://doi.org/10.1037/0022-0663.98.1.209

Reinholz, D. L., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education*, *49*(2), 140–177. https://doi.org/10.5951/jresematheduc.49.2.0140

Schoen, R. C., Bray, W. S., Tazaz, A. M., & Buntin, C. K. (2022). *Description of the cognitively guided instruction professional development program in Florida: 2013–2020*. Florida State University. https://doi.org/10.33009/fsu.1643828800

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*. https://www.teachingisproblemsolving.org/whats-next-stories

Schoen, R. C., LaVenia, M., Tazaz, A. M., Gersten, R., & Smolkowski, K. (in press). 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. *The Elementary School Journal*.

Schoen, R. C., LaVenia, M., Tazaz, A. M., 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*. Florida State University. https://doi.org/10.33009/fsu.1601237075

Schoen, R. C., Li, L., Yang, X., Guven, A., & Riddell, C. (2021, September 26–29). *Using a many-facet Rasch model to gain insight into measurement of instructional practice in mathematics* [Paper presentation]. Annual conference of the Society for Research in Educational Effectiveness.

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. https://doi.org/10.33009/fsu.1653430141

Schoenfeld, A. H. (2018). Video analyses for research and professional development: The teaching for robust understanding (TRU) framework. *ZDM*, *50*(3), 491–506. https://doi.org/10.1007/s11858-017-0908-y

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through the teachers' eyes*. Routledge. https://doi.org/10.4324/9780203832714

Sherin, M. G., & Lynn, J. (2019). Attending to evidence of students' thinking during instruction. *Mathematics Teaching in the Middle School*, *24*(7), 430–435. https://doi.org/10.5951/mathteacmiddscho.24.7.0430

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

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development* (2nd ed.). Teachers College Press.

Stigler, J. W., & Hiebert, J. (2009). *The teaching gap: Best ideas from the worlds' teachers for improving education in the classroom*. Free Press.

Strati, A. D., Schmidt, J. A., & Maier, K. S. (2017). Perceived challenge, teacher support, and teacher obstruction as predictors of student engagement. *Journal of Educational Psychology*, *109*(1), 131–147. https://doi.org/10.1037/edu0000108

Villaseñor, A., Jr., & Kepner, H. S., Jr. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education*, *24*(1), 62–69. https://doi.org/10.5951/jresematheduc.24.1.0062

Walkowiak, T. A., Berry, R. Q., Pinter, H. H., & Jacobson, E. D. (2018). Utilizing the M-Scan to measure standards-based mathematics teaching practices: Affordances and limitations. *ZDM*, *50*(3), 461–474. https://doi.org/10.1007/s11858-018-0931-7

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, *78*(3), 516–551. https://doi.org/10.3102/0034654308320292

Webb, N. M., Franke, M. L., Ing, M., Chan, A., De, T., Freund, D., & Battey, D. (2008). The role of teacher instructional practices in student collaboration. *Contemporary Educational Psychology*, *33*(3), 360–381. https://doi.org/10.1016/j.cedpsych.2008.05.003

Webb, N. M., Franke, M. L., Ing, M., Turrou, A. C., Johnson, N. C., & Zimmerman, J. (2019). Teacher practices that promote productive dialogue and learning in mathematics classrooms. *International Journal of Educational Research*, *97*, 176–186. https://doi.org/10.1016/j.ijer.2017.07.009

Webb, N. M., Franke, M. L., Ing, M., Wong, J., Fernandez, C. H., Shin, N., & Turrou, A. C. (2014). Engaging with others' mathematical ideas: Interrelationships among student participation, teachers' instructional practices, and learning. *International Journal of Educational Research*, *63*, 79–93. https://doi.org/10.1016/j.ijer.2013.02.001

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, *3*(1), 129–137. https://doi.org/10.1111/j.2041-210X.2011.00125.x

Woods, T., Williams, G., & McNeal, B. (2006). Children's mathematical thinking in different classroom cultures. *Journal for Research in Mathematics Education*, *37*(3), 222–255.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458–477. https://doi.org/10.5951/jresematheduc.27.4.0458

Authors

Robert Schoen, Learning Systems Institute and School of Teacher Education, Florida State University, Tallahassee, FL 32306; rschoen@fsu.edu

Wendy Bray, Learning Systems Institute, Florida State University, Tallahassee, FL 32306; wbray@fsu.edu

Claire Riddell, Learning Systems Institute and School of Teacher Education, Florida State University, Tallahassee, FL 32306; criddell@fsu.edu

Charity Buntin, Learning Systems Institute and School of Teacher Education, Florida State University, Tallahassee, FL 32306; cbauduin2@lsi.fsu.edu

Naomi Iuhasz-Velez, Learning Systems Institute, Florida State University, Tallahassee, FL 32306; niuhaszvelez@fsu.edu

Walter Secada, School of Education and Human Development, University of Miami, Coral Gables, FL 33146; wsecada@miami.edu

Eva Yujia Li, Office of Institutional Research, Yale University, New Haven, CT 06520; evayujia.li@yale.edu

Submitted May 24, 2023
Accepted September 15, 2023
doi:10.5951/jresematheduc-2022-0130

APPENDIX

Interpretation and Use Statement for M-CLIPS Observation Instrument (v. 2023-11-03)

Construct and articulation
What is the construct being measured?
Implementation of Cognitively Guided Instruction (CGI) in mathematics lessons.

Why is the construct important to measure?
We identified a need for greater insight into classroom implementation of CGI and data that will support empirical study of implementation, outcomes, and the theory of change for CGI programs.

Operationalization and Administration
How is the construct measured?
Certified raters observe live or video recorded mathematics lessons. Duration of lessons may vary but are usually 30–75 min.

Who is the target population?
Teachers and students engaged in the teaching and learning of mathematics.

What is the context for administration?
At this time, M-CLIPS has been administered only by researchers for research purposes using video recorded lessons.

What costs are associated with the instrument?
The M-CLIPS instrument is available to the public, free of charge, through a Creative Commons–by Attribution license. Raters must successfully complete M-CLIPS observer training to demonstrate qualifications.

Scores and Usage
How are scores calculated?
Scoring may use the sum or arithmetic mean of the raw ratings for the nine rubrics or a many-facet Rasch model. The former method is simpler for scoring and interpreting scores. The latter has many affordances, such as statistical adjustment for rater severity and separate calculation of reliability for raters and the measure, but requires various design elements, such as linking across raters.

How should scores be interpreted?
Raw rubric ratings of less than 2 indicate that the implementation of the mathematics lesson is inconsistent with CGI principles. Ratings of 2 or 3 indicate weak or moderate implementation of CGI principles, respectively. Ratings of 4 or 5 indicate strong implementation of CGI principles; 5 is reserved for exemplary approaches.

How should scores be used?
M-CLIPS scores can be used to quantify the implementation of core principles of CGI in mathematics lessons.

What cautions or warnings should be considered?
M-CLIPS should not be used to evaluate individuals or make high-stakes decisions. Scores may be interpreted differently if the observed teachers are aware of the nine rubrics in M-CLIPS and whether the observers are using M-CLIPS in their observations. Real-time observation and rating of lessons with M-CLIPS will need to be done before we know whether the instrument can be used in that way and whether the results are comparable between video and real-time, in-classroom observations.