

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

Robert C. Schoen
Mark LaVenia
Amanda M. Tazaz

The research and development reported here was supported by the Florida Department of Education, through Award Numbers 371-2355B-5C001, 371-2356B-6C001, and 371-2357B-7C004 to Florida State University. The opinions expressed are those of the authors and do not represent views of the Florida Department of Education.

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

Copyright 2018, Florida State University. All rights reserved. Requests for permission to use this test should be directed to Robert Schoen, rschoen@lsi.fsu.edu, FSU Learning Systems Institute, 4600 University Center C, Tallahassee, FL, 32306

**Effects of the First Year of a Three-Year CGI Teacher Professional
Development Program on Grades 3–5 Student Achievement
A Multisite Cluster-Randomized Trial**

Research Report No. 2018-25

Robert C. Schoen

Mark LaVenia

Amanda M. Tazaz

Florida Center for Research in Science, Technology, Engineering, and Mathematics (FCR-STEM)
Learning Systems Institute
Florida State University
Tallahassee, FL 32306
(850) 644-2570

Introduction

Relative to the rest of the world, U.S. students' performance in mathematics decreases between grade 4 and grade 8 (Mullis, Martin, Foy, & Hooper, 2016). The introduction of abstract concepts such as rational number has been theorized by members of the National Mathematics Advisory Panel (2008) as a likely cause of the dip in student performance in these grade levels, and proficiency with fractions has been identified as an important gatekeeper for success in Algebra (Booth & Newton, 2012; Empson & Levi, 2011; Siegler et al., 2012; Siegler, Thompson, & Schneider, 2011).

Professional development for teachers is one possible strategy for improving student achievement. Regrettably, a recent trend toward rigorous evaluation of the impact of highly regarded teacher PD programs on student learning has yielded mostly null findings (Garet et al., 2011; Garet et al., 2016; Jacob, Hill, & Corey, 2017; Jayanthi et al., 2017; Santagata et al., 2010). In a review of the emerging evidence of the effects of teacher PD programs on student learning, Wilson (2013) concluded, "it is nearly impossible to isolate the effects of PD on student learning" (p. 312).

One exception to the trend of null effects have been programs based on Cognitively Guided Instruction (CGI), which have had positive effects on student's achievement (Gersten et al., 2014). Until now, three extant studies employing experimental designs have reported positive effects on student achievement in mathematics. Carpenter et al. (1989) reported positive effects on first-grade students' achievement on problem-solving tasks within the domain of whole-number addition and subtraction. Jacobs et al. (2007) reported positive effects on first-through sixth-grade students' understanding of the equals sign and use of strategies that rely on algebraic relationships to solve equations. The third study—and the first one to be conducted by a third-party evaluator—reported results suggesting potentially positive effects on first-grade students' problem-solving abilities ($p < .10$) and potentially negative effects on second-grade students computational abilities ($p < .10$) in the first year of the program (Schoen, et. al 2017)

Although results concerning the impact of CGI PD on teachers and students are promising, much work is still needed in the area of research and evaluation in that CGI innovations are outpacing even the formidable body of CGI-related research and evaluation. Over the past thirty years, CGI has continually evolved and expanded. The developers of the CGI program have maintained a focus on increasing teachers' understanding of children's thinking as a means to improve teaching and learning, where problem-solving serves as the organizing focus for instruction (Carpenter et al., 1999; Carpenter & Franke, 2004). That focus has not changed. Whereas the original CGI program focused on addition and subtraction on whole numbers at the first-grade level (Carpenter et al., 1989), the content of CGI expanded through the 1990s to include multiplication and division on whole numbers and base-ten number concepts (Carpenter et al., 1999; Fennema et al., 1996). In the 2000s, the scope of the CGI program expanded further to focus on algebraic thinking (Carpenter, Franke, & Levi, 2003; Jacobs et al., 2007). In the 2010s, Empson and Levi (2011) published frameworks to provide support for the

CGI approach in the domain of fractions and decimals, and Carpenter et al. (2016) published a similar book to support the CGI approach in the domain of counting and early number concepts.

Purpose Statement

The purpose of the present study was to estimate the impact of the first year of a three-year teacher professional development program designed for grades 3–5 mathematics teachers and their students. The study is guided by the following research question.

What is the effect of the CGI 3–5 program on grade 3, 4, and 5 student achievement as measured by the Spring 2016 Grades 3–5 Elementary Mathematics Student Assessment (EMSA)?

Other mechanisms in the theory of change for the CGI program, such as the impact of the program on teacher knowledge and beliefs and the role of those factors as mediators of the effect of the program on student achievement, are addressed in the broader evaluation and research study and are beyond the scope of the present manuscript.

Description of the CGI 3–5 Program

The teachers in this study participated in the first year of a three-year CGI 3–5 professional development program. This particular model consisted of five consecutive seven-hour days during summer 2015; two consecutive six-hour days in fall 2015; and two consecutive days six-hour days in winter 2016. The CGI 3–5 program was designed and taught by certified CGI instructors at the Teachers Development Group under the direction of Linda Levi. Dr. Levi was the Director of CGI Initiatives at TDG and one of the co-authors of the three definitive CGI books as well as the CGI Guide for Workshop Leaders (Carpenter et al., 1999; Carpenter, Franke, & Levi, 2003; Empson & Levi, 2011; Fennema et al., 1999).

The CGI 3–5 program was designed to focus teachers' attention on their students' mathematical thinking and to provide teachers with principled frameworks for understanding their students' thinking. Teachers are introduced to two, complementary, researched-based frameworks during the PD workshops:

- Problem Types Frameworks, which describe how the structure of a problem influences how children think about the mathematical concepts embedded in the problem; and
- Solution Strategy Frameworks, which describe the developmental progressions of children's mathematical thinking as determined by their strategies for solving problems within the problem-type framework.

The frameworks addressed in the CGI 3–5 program describe children's thinking about: fraction quantities; operations with fractions; whole number multiplication and division; and base-ten concepts for whole numbers and decimals. In addition to developing an understanding of these

frameworks, teachers are supported to use what they learn about students to drive instructional decisions.

CGI 3–5 Program Theory of Change

Figure 1 presents the CGI 3–5 program theory of change. The program aims to focus teachers' attention on the details in students' cognitive processes as they solve problems and encourages them to use what they learn about students to drive instructional decisions. It is thought to have a direct effect on teachers' mathematical knowledge for teaching and beliefs about mathematics teaching and learning and an indirect effect on classroom instruction and student learning in mathematics. These effects build incrementally over the course of the school year (and across multiple school years for the multiyear program).

Teachers learn about learning progressions related to student thinking and practice using these progressions to guide their instructional practice. During the classroom-embedded components of the professional-development workshops, teachers learn how to use problem solving as the organizing principle for instruction and orchestrate classroom discussions involving student sharing their various solution strategies and learning from one another. Teachers' understanding of fractions and other mathematics concepts along with associated conventions of mathematical notation grows through in-depth study of children's thinking processes.

Classroom instruction following the CGI approach (Carpenter et al., 2015) uses a formative-assessment process wherein teachers observe students solving problems and explaining their thinking processes. Teachers use these observations to draw inference about students' mathematical understanding. Students in CGI classrooms learn mathematics by engaging in problem solving, explaining their problem-solving strategies to the teacher and to their peers, and listening to various ways of solving problems.

Methods

Setting

One hundred forty-nine grades 3–5 teachers—representing 32 schools, nine public school districts, and the geographic, socioeconomic, and cultural diversity of the central and northern regions of the state of Florida—participated in the study. Students in the analytic sample represented grades 3, 4, and 5.

Enrollment

Teachers were enrolled through an informed-consent process to voluntarily participate in the professional-development program and associated research study through a Web-based survey deployed in spring 2015. To qualify for enrollment in the study, the applicants needed to be expecting to teach in a participating school district, and their expected teaching assignment for

SY 2015–16 needed to involve teaching mathematics to intermediate-grades students. At the start of SY 2015–16, teachers distributed an approved letter from the principal investigator to parents and guardians in accordance with the consent procedure approved by the participating school districts and the Institutional Review Board at Florida State University.

Randomization

Blocking on schools, individual teachers were assigned at random to the intervention group or the wait-list comparison group. Teachers assigned to the comparison condition participated in business-as-usual professional development and mathematics instruction during the study year and were invited to participate in the intervention program starting the summer after the student and teacher posttests were completed.

A total of 34 randomization blocks were used. The additional 2 blocks (greater than the number of schools) are attributable to two teachers being part of a rolling assignment process (to meet the target sample size); both teachers were late applicants in schools where there were teachers who were already randomized into the study. The number of participating teachers per school ranged from three to eight. The decision rule for the blocking procedure was, if the number of participating teachers at a school was three to five, two teachers were randomly assigned to treatment and the others to control; if the number of participating teachers at a school was six or more, three were randomly assigned to treatment and the others to control. For the two rolling-assignment blocks, teachers had a 50 percent chance of assignment to treatment. This approach resulted in 69 of the 149 teachers being assigned to treatment. As described in the data analysis section below, the covariate set for the impact analyses included dummy variables for randomization blocks, which is one of three methods considered acceptable by the What Works Clearinghouse (WWC, 2017) for accounting for different assignment probabilities.

Data Sources

Baseline student achievement was measured using the 2015 Fall Grades 3–5 Elementary Mathematics Student Assessment (EMSA) test during the first two weeks of the school year (Schoen, Anderson, Champagne, & Bauduin, 2018). Student posttest mathematics achievement was measured in May 2016 using the Spring 2016 3–5 EMSA test (Schoen, Anderson, & Bauduin, 2018). The tests were designed to assess student achievement in understanding fraction quantities and operations involving whole numbers and fractions.

Teachers provided consent to participate and completed pretest assessments during April–June, 2015, prior to participation in the summer training. Teachers completed posttest assessments in April–May, 2016. Students completed pretest assessments during the first two weeks of the 2015–16 school year. Students completed posttest assessments during an assessment window spanning late-April to late-May in 2016.

Sample Attrition and Cluster Baseline Equivalence

All impact estimates were conducted as an intent-to-treat analysis of randomized teachers and their respective students. We used a joiners model for the student sample, where all students with consent to participate who were enrolled in a randomized teacher’s classroom at the time of post-testing were eligible for inclusion in the analytic sample. The analytic sample included those students eligible for inclusion in the analytic sample who completed the outcome measure administered spring 2016. Teacher-level attrition was evaluated based on randomized teachers with students who met the inclusion criteria for the analytic sample relative to all randomized teachers. Student-level attrition was evaluated based on students in the analytic sample relative to all students on randomized teachers’ spring 2016 class rosters (excluding students within attriting teacher clusters), which were provided by the school districts’ and based on their administrative data.

Table 1 presents attrition rates for teacher clusters and students. The overall and differential rates for teacher clusters were 32.9% and 11.6%, respectively. According to WWC (2017) guidelines, this exceeds the boundary for acceptable threat of bias due to attrition. However, the overall and differential rates of 9.2% and 4.4%, respectively, for students are below the cautious WWC boundary for acceptable threat of bias due to attrition.

Equivalence of clusters at baseline was established using fall 2015 student achievement data for all students who were enrolled in a non-attriting teacher’s classroom spring 2016. The pretest and the posttest measures correlated at $r = .76, p < .001$. As presented in Table 2, using classroom means and standard deviations for the calculation, a group difference of $g = .21$ was calculated; using individual student scores and standard deviations for the calculation, a group difference of $g = .17$ was calculated. Parallel information is provided in Table A1 in Appendix A. Regardless of calculation method, results indicate the group difference at baseline was within acceptable range, providing statistical adjustment for pretest in the analysis, according to WWC guidelines (2017).

Data Analysis

Data were fit to a two-level model involving students nested in teachers with fixed effects for $n-1$ randomization blocks. Table 3 presents a description of the analytic models as fit using Mplus 8 (Muthén & Muthén, 1998–2017). Model 1 includes binary indicator covariates for student grade level, treatment condition, and randomization block. Model 2 adds a continuous covariate for pretest. All models used the Mplus MLR maximum likelihood with robust standard errors estimator. All impact analyses constituted a second stage of analysis, where the initial stage involved calculation of IRT-based scores for the pre- and posttest student achievement measures. The saved student-ability estimates were treated as observed variables in the impact analyses.

For all impact analyses, we used a full information maximum likelihood (FIML) approach, as described by Muthén, Muthén, and Asparouhov (2016), to avoid dropping cases missing data

on the pretest measure. FIML is one of several WWC-accepted approaches for handling missing data (WWC, 2017).

Sensitivity and Verification

Additional analyses were conducted using a complete-case sample with respect to student pre- and posttest data. Models 1 and 2 were fit to the complete-case sample, as were exploratory models to test whether effects of treatment were moderated by student grade level or baseline performance on the student pretest. Using the same data, these models were fit independently by the projects' external evaluator using Stata v. 15.1 (StataCorp, 2017) and R v. 3.4.4 (R Development Core Team, 2014) lme4 package (Bates, Maechler, Bolker, & Walker, 2015).

Results

Main Effects

The results of the statistical models designed to estimate the main effect of the program on student achievement are provided in Table 4. Using FIML with the full analytic sample, we estimated an effect for treatment in Model 2 of $g = 0.18$ ($p = .007$). Results of the sensitivity analysis using the complete-case sample are provided in Table 5. For the complete-case analysis, we estimated an effect for treatment in Model 2 of $g = 0.13$ ($p = .030$). The unadjusted standard deviations and analytic sample sizes are provided in the Appendix in Table A2.

Moderation Analyses

We did not find any statistically significant interactions between treatment and grade level or treatment and baseline achievement scores.

Sensitivity Analysis

Results from our analyses and those conducted by the external evaluators were practically identical and had no meaningful discrepancies apart from what would be expected by different software routines. Further, the inference drawn from the complete-case analysis did not differ from the inference drawn from the results of the model that used FIML, suggesting the results are robust to the influence of missing data on sample composition.

Discussion

The CGI 3–5 PD program was found to significantly impact student mathematics achievement ($p = .007$) in the first year of the three-year program. The point estimate of the main effect of the CGI program on student achievement ($g = .18$) was greater than many of the most effective

teacher professional-development programs that have been subjected to rigorous evaluation (Kennedy, 2016).

In a recent evaluation brief, Garet et al. (2016) discussed three teacher PD programs that did not translate to increases in student learning. They concluded that teacher PD can improve teachers' content knowledge and practice, to some degree, but "that the field does not yet fully understand how to ensure that teacher PD leads to measurable improvements in student learning" (p. 11). The findings reported in the present study provide a counterexample to this narrative and suggest that the field can learn from the design and implementation of the CGI 3–5 program.

Some of the leading ideas for improving student learning in fractions currently focus on the use of linear representations and relative magnitude of fractions, including identifying fractions as points on the number line (Booth, Newton, & Twiss-Garrity, 2014; Lewis & Perry, 2017, Torbeyns, Schneider, Xin, & Siegler, 2015). CGI 3-5 PD takes a different approach from these other initiatives. Rather than introducing fractions through part-whole models or as points on a number line, fractions are introduced as an extension of whole number division through Equal Sharing Problems, such as:

If there are 3 brownies, and 4 people want to share them equally, how much brownie should each person get?

In the CGI approach to instruction, students are supported in developing their intuitive strategies to solve mathematics problems involving fractions and, in the process, develop a strong understanding of fractional quantities and the properties of operations as they related to operations with fractions. Students are not shown how to solve these problems. Instead, they produce their own representations of fractional quantities in the process of solving each problem. Word problems are also used as vehicles for introducing fractions, operations on fractions and whole numbers, and equivalence relations. Word problems, and equations associated with these problems, continue to be the primary tools for teaching fractions concepts and operations after such concepts are introduced. For more information on this approach, see Empson and Levi (2011).

CGI programs have garnered considerable praise and attention from mathematics educators and mathematics teacher educators since the program's inception. These new results provide further evidence that the CGI approach to professional development may extend to higher grade levels. A more complete analysis of the theory of change and implementation of the CGI 3–5 program will be necessary to yield additional insight into the generalizable and scalable components of the design and implementation that resulted in the positive impact on students. Future work should involve (a) replication of this study, (b) a longer-term study to evaluate the impact of the full three-year program, and (c) a more complete analysis of the impact of the program on teachers' knowledge, beliefs, and behaviors, including classroom-instruction data, and the role of those factors in mediating the effect of the program on students.

References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology*, 37, 247–253.
- Booth, J. L., Newton, K. J., Twiss-Garrity, L. K. (2014). The impact of fraction magnitude knowledge on algebra performance and learning. *Journal of Experimental Child Psychology*, 118, 110–118.
- Carpenter, T.P.; Fennema, E., Franke, M.L., Levi, L., & Empson, S.B. (1999). *Children's Mathematics: Cognitively Guided Instruction*. Portsmouth, NH: 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), 385-531.
- Carpenter, T. P., & Franke, M. L. (2004). Cognitively Guided Instruction: Challenging the core of educational practice. In T. K. Glennan, S. J. Bodilly, J. R. Galegher & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 41-80). Santa Monica, CA: RAND Corporation.
- Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic & algebra in elementary school*. Portsmouth, NH: Heinemann.
- Carpenter, T. P., Franke, M. L., Johnson, N. C., Turrou, A. C., & Wager, A. A. (2016). *Young Children's Mathematics: Cognitively Guided Instruction in Early Childhood Education*. Heinemann Educational Books.
- Empson, S. B., and Levi, L. (2011). *Extending children's mathematics: Fractions and decimals*. Portsmouth, NH: 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, 458-477.
- Fennema, E., Carpenter, T. P., Levi, L., Franke, M. L., & Empson, S. B. (1999). *Children's mathematics: Cognitively Guided Instruction - A guide for workshop leaders*. Portsmouth, NH: Heinemann.
- Fuson, K. (1992). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*. Reston, VA: National Council of Teachers of Mathematics.
- Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., & Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development* (NCEE 2016-4010). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., and Doolittle, F. (2011). *Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation* (NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

- Gersten, R., Taylor, M. J., Keys, T. D., Rolhus, E., & Newman-Gonchar, R. (2014). Summary of research on the effectiveness of math professional development approaches. Tallahassee: Regional Educational Laboratory Southeast, Florida State University. Retrieved June 30, 2014 from http://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2014010.pdf.
- Jacob, R., Hill, H., & Corey, D. (2017). The impact of a professional development program on teachers' mathematical knowledge for teaching, instruction, and student achievement. *Journal for Research on Educational Effectiveness*, *10*(2), 379–407.
- 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.
- Jayanthi, M., Gersten, R., Taylor, M. J., Smolkowski, K., & Dimino, J. (2017). Impact of the Developing Mathematical Ideas professional development program on grade 4 students' and teachers' understanding of fractions (REL 2017–256). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 1–36.
- Lewis, C. & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261–299.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 international results in mathematics. Chestnut Hill, MA: Boston College.
- Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthen & Muthen.
- Muthén, L. K. and Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthen & Muthen.
- National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. US Department of Education: Washington, DC.
- R Development Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Santagata, R., Kersting, N., Givvin, K. B., & Stigler, J. W. (2011). Problem implementation as a lever for change: An experimental study for the effects of a professional development program on students' mathematical learning. *Journal of Research on Educational Effectiveness*, *4*, 1–24.
- Schoen, R. C., Anderson, D., & Bauduin, C. (2018). Elementary mathematics student assessment: Measuring the performance of grade 3, 4, and 5 students in number (whole number and fractions), operations, and algebraic thinking in fall 2015. (Research Report No. 2018-21). Tallahassee, FL: Learning Systems Institute, Florida State University.
- Schoen, R. C., Anderson, D., & Bauduin, C. (2018). Elementary mathematics student assessment: Measuring grade 3, 4, and 5 students' performance in number (whole numbers and fractions), operations, and algebraic thinking in spring 2016. (Research Report No. 2018-23). Tallahassee, FL: Learning Systems Institute, Florida State University.

- Schoen, R. C., Kisa, Z., & Tazaz, A. *Effects of the first two years of a professional development program on grade 1 and 2 teachers' mathematical knowledge for teaching and beliefs about mathematics teaching and learning: A randomized controlled trial*. Manuscript under review.
- Schoen, R.C., Lavenia, M., & Tazaz, A. (2017, March). Effects of a Two-year Cognitively Guided Instruction Professional Development Program on First- and Second-Grade Student Achievement in Mathematics. Paper presented at the spring conference of the Society for Research in Educational Effectiveness, Washington, DC.
- Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., et al. (2012). Early predictors of high school mathematics achievement. *Psychological Science*, 23, 691–697.
- Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, 62, 273–296.
- StataCorp. (2017). Stata statistical software: Release 15. College Station, TX: StataCorp LLC.
- Torbeyns, J. Schneider, M., Xin, Z., & Siegler, R. S. (2015). Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. *Learning and Instruction*, 37, 5–13.
- What Works Clearinghouse. (2017). Standards handbook, Version 4.0. Washington, DC: U. S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf
- Wilson, S. M. (2013). Science teacher professional development. *Science*, 340, 310-313.

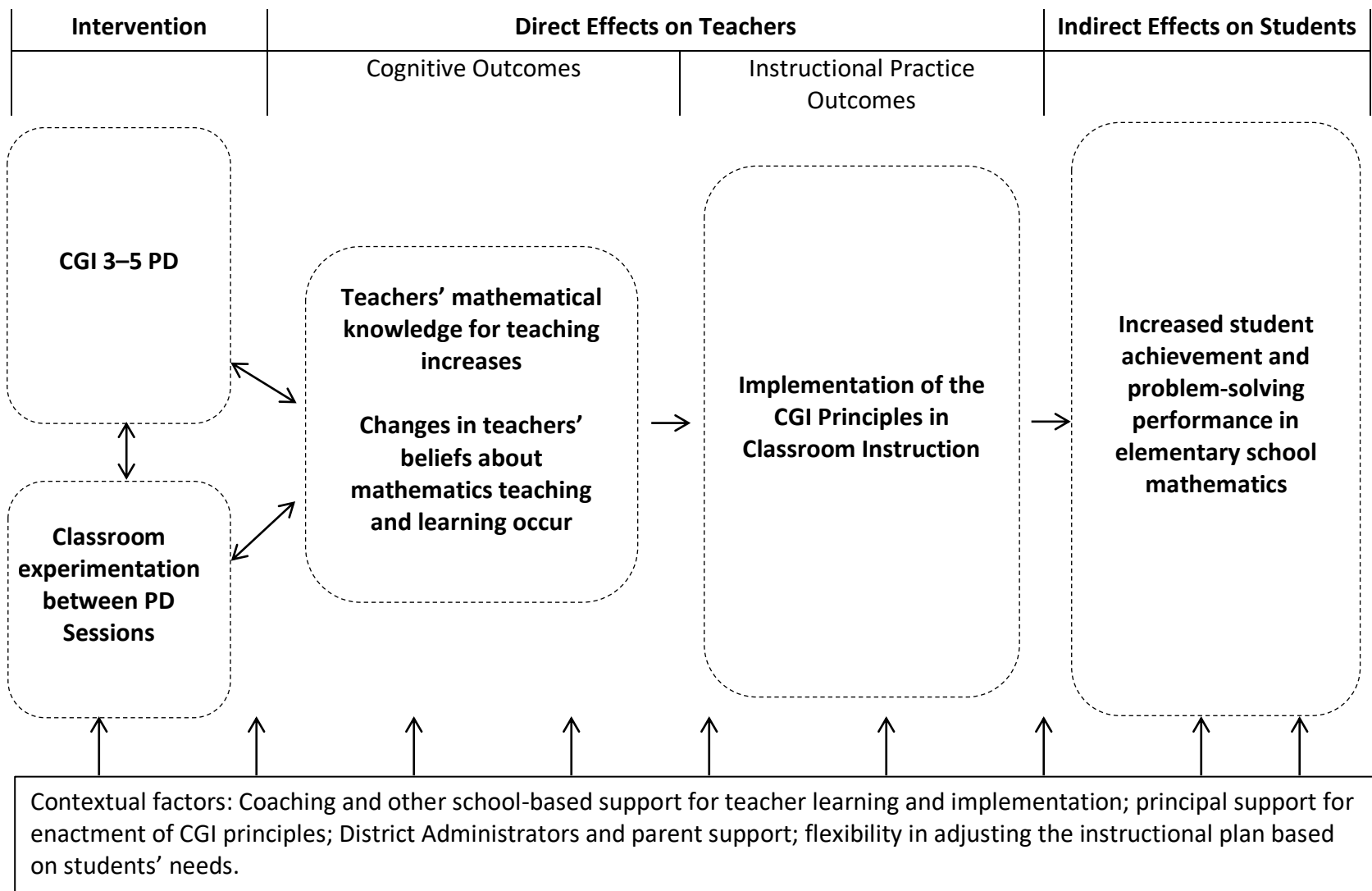


Figure 1. Theory of change for the CGI 3-5 program.

Table 1. Teacher Cluster Attrition and Students Representativeness

| | Analytic sample <i>N</i> | | | Reference sample <i>N</i> | | | Attrition/ Representativeness | |
|--|--------------------------|---------|-------|---------------------------|---------|-------|----------------------------------|--------------|
| | Treatment | Control | Total | Treatment | Control | Total | Overall | Differential |
| Randomly assigned teacher clusters | | | | | | | | |
| Teachers | 42 | 58 | 100 | 69 | 80 | 149 | 32.89% | 11.63% |
| Students (excluding those within attriting teacher clusters) | | | | | | | | |
| Students | 957 | 1,249 | 2,206 | 1,083 | 1,347 | 2,430 | 9.22% | 4.36% |

Note. Analytic sample is defined as all students with data on the outcome measure.

Table 2. Baseline Equivalence of Clusters

| Baseline <i>M</i> | | Hedges' <i>g</i> standard deviation calculation | | | | | Small-sample bias correction calculation | | |
|---|----------------------|---|-----------------------------------|----------------------|-----------------------------------|----------------------------|--|----------|----------|
| <i>m_i</i> | <i>m_c</i> | <i>n_i</i> | <i>SD_i²</i> | <i>n_c</i> | <i>SD_c²</i> | <i>SD_{pooled}</i> | <i>N</i> | ω | <i>g</i> |
| Using classroom means and standard deviations | | | | | | | | | |
| 0.193 | 0.002 | 43 | 0.979 | 52 | 0.743 | 0.922 | 95 | 0.992 | 0.205 |
| Using individual student scores and standard deviations | | | | | | | | | |
| 0.321 | 0.114 | 844 | 1.599 | 963 | 1.304 | 1.201 | 1,807 | 1.000 | 0.172 |

Note. Subscript *i* and *c* refer to the intervention and control group, respectively.

Table 3. *Description of Analytic Models as Fit using Mplus*

| Model | Construct/Variable | Variable description | Mplus modeling particulars |
|-------|--|---|--|
| M1 | Student characteristics | | |
| | Grade 4 | Binary variable indicating student was in Grade 4 | Modeled at Level-1, as fixed effect. Variance is estimated at Level-1. |
| | Grade 5 | Binary variable indicating student was in Grade 5 | Modeled at Level-1, as fixed effect. Variance is estimated at Level-1. |
| | Group characteristics | | |
| | Block | Vector of binary indicators for randomized block (school) ID, using Block 35 as the reference category. | Modeled at Level-2, as fixed effect. |
| | Workshop | Binary variable indicating group was assigned to workshop condition. | Modeled at Level-2, as fixed effect. |
| M2 | Student characteristics <i>All variables in M1 plus:</i> | | |
| | Pretest | Continuous variable for student mathematics achievement fall 2015 pretest, vertically scaled for grades 3-5 | Modeled at Level-1, as fixed effect. Variance is estimated at Level-1. |

Note. The Mplus MLR maximum likelihood with robust standard errors estimator is used for both models. The specification that variance for independent variables is estimation at Level-1 applies to only analyses using FIML to handle missing data; estimation of variance at Level-1 for independent variables is not necessary for complete case analysis.

Table 4. *Impact Analyses with the Full Sample using FIML to Handle Missing Data*

| | Model 1 | | | Model 2 | | |
|---------------------|-------------------|-------------|----------|-------------------|-------------|----------|
| | Coeff (SE) | Effect size | <i>p</i> | Coeff (SE) | Effect size | <i>p</i> |
| Fixed effects | | | | | | |
| Within classroom | | | | | | |
| Grade 4 | 0.423 (0.098) | — | < .001 | -0.022 (0.086) | — | .798 |
| Grade 5 | 1.069 (0.119) | — | < .001 | 0.022 (0.106) | — | .838 |
| Pretest | | | | 0.653 (0.025) | — | < .001 |
| Between classroom | | | | | | |
| Block ^a | — | — | — | — | — | — |
| Workshop | 0.189 (0.092) | 0.17 | .041 | 0.202 (0.075) | 0.18 | .007 |
| Intercept | -0.294 (0.137) | — | .032 | 0.005 (0.129) | — | .968 |
| Variance components | | | | | | |
| Within classroom | 0.800 (0.035) | | < .001 | 0.450 (0.016) | | < .001 |
| Between classrooms | 0.107 (0.026) | | < .001 | 0.058 (0.015) | | < .001 |

Note. Student $N = 2,206$; Teacher $N = 100$. Reported estimates are unstandardized. Only the effect size for the Workshop treatment variable is presented.

^aBlock indicates the vector of $n-1$ randomization blocks. Parameters for Block are not presented for visual simplicity.

^bCoefficients reported at the within-level for random slopes are the random intercepts estimated at the between-level.

Table 5. *Impact Analyses with the Complete-Case Sample*

| | Model 1 | | | Model 2 | | |
|---------------------|-------------------|-------------|----------|-------------------|-------------|----------|
| | Coeff (SE) | Effect size | <i>p</i> | Coeff (SE) | Effect size | <i>p</i> |
| Fixed effects | | | | | | |
| Within classroom | | | | | | |
| Grade 4 | 0.430 (0.106) | — | < .001 | -0.162 (0.079) | — | .040 |
| Grade 5 | 1.060 (0.129) | — | < .001 | -0.237 (0.094) | — | .011 |
| Pretest | | | | 0.722 (0.024) | — | < .001 |
| Between classroom | | | | | | |
| Block ^a | — | — | — | — | — | — |
| Workshop | 0.161 (0.099) | 0.14 | .106 | 0.147 (0.068) | 0.13 | .030 |
| Intercept | -0.225 (0.156) | — | .150 | 0.111 (0.128) | — | .386 |
| Variance components | | | | | | |
| Within classroom | 0.781 (0.037) | | < .001 | 0.452 (0.016) | | < .001 |
| Between classrooms | 0.106 (0.027) | | < .001 | 0.038 (0.009) | | < .001 |

Note. Student $N = 1,683$; Teacher $N = 92$. Reported estimates are unstandardized. Only the effect size for the Workshop treatment variable is presented.

^aBlock indicates the vector of $n-1$ randomization blocks. Parameters for Block are not presented for visual simplicity.

Appendix A
Table A1. Baseline Equivalence Information for Complete-Case Analysis ($n = 1,683$) reported in Table 5

| | Intervention group | Comparison group | Total |
|------------------------------------|--------------------|------------------|-------|
| Student sample size | 770 | 913 | 1,683 |
| Student-level EMSA Pretest Mean | 0.386304 | 0.092678 | |
| Student-level EMSA Pretest SD | 1.269690 | 1.139906 | |

Table A2. Unadjusted SDs for Analytic Sample for Complete-Case Analysis in Table 5

| | HLM coefficient | Analytic-sample unadjusted standard deviation | Analytic-sample size (n) |
|--------------------|-----------------|---|---------------------------------|
| Intervention group | | 1.127458 | 770 |
| Comparison group | 0.147 | 1.101253 | 913 |
| <i>Total</i> | | | 1,683 |