12-17-2021

Impact Evaluation of the Lucy Calkins Units of Study Program

Béla Figge
Georgia State University

Tim Sass
Georgia State University

Recommended Citation

Figge, B., & Sass, T. (2021). Impact Evaluation of the Lucy Calkins Units of Study Program. Georgia Policy Labs. https://doi.org/10.57709/30728968

GEORGIA
POLICY LABS

Georgia State University | Andrew Young School of Policy Studies

Impact Evaluation of the Lucy Calkins Units of Study Program

Béla Figge and Tim R. Sass

Metro Atlanta Policy Lab for Education

December 2021

Background and Motivation

The Lucy Calkins Units of Study Program (hereafter "Lucy Calkins Program") is a widely used program that consists of materials and methods for teaching reading and writing in Grades K–8. It was developed by Professor Lucy Calkins and her colleagues in the Reading and Writing Project—a center at Teachers College, Columbia University. The Project developed the curriculum and teaching methods and provides professional development for teachers implementing the program.

A metro-Atlanta school district ("the district") has used the Lucy Calkins Program for many years, with adoption at the discretion of individual school leaders. Table 1 summarizes the implementation of the Lucy Calkins Program in traditional elementary schools throughout the district. Four elementary schools first implemented the Reading Program in school year (SY) 2014–15, and 20 schools implemented the Writing Program the same year. By SY 2018–19, all but six traditional elementary schools in the district implemented a Lucy Calkins Reading and/or Writing Program. Some schools fully implemented the program; some implemented only parts of the program; some took a "hybrid" approach, combining the Lucy Calkins Program with another reading/writing instructional approach; and others did not utilize the Lucy Calkins Program at all.

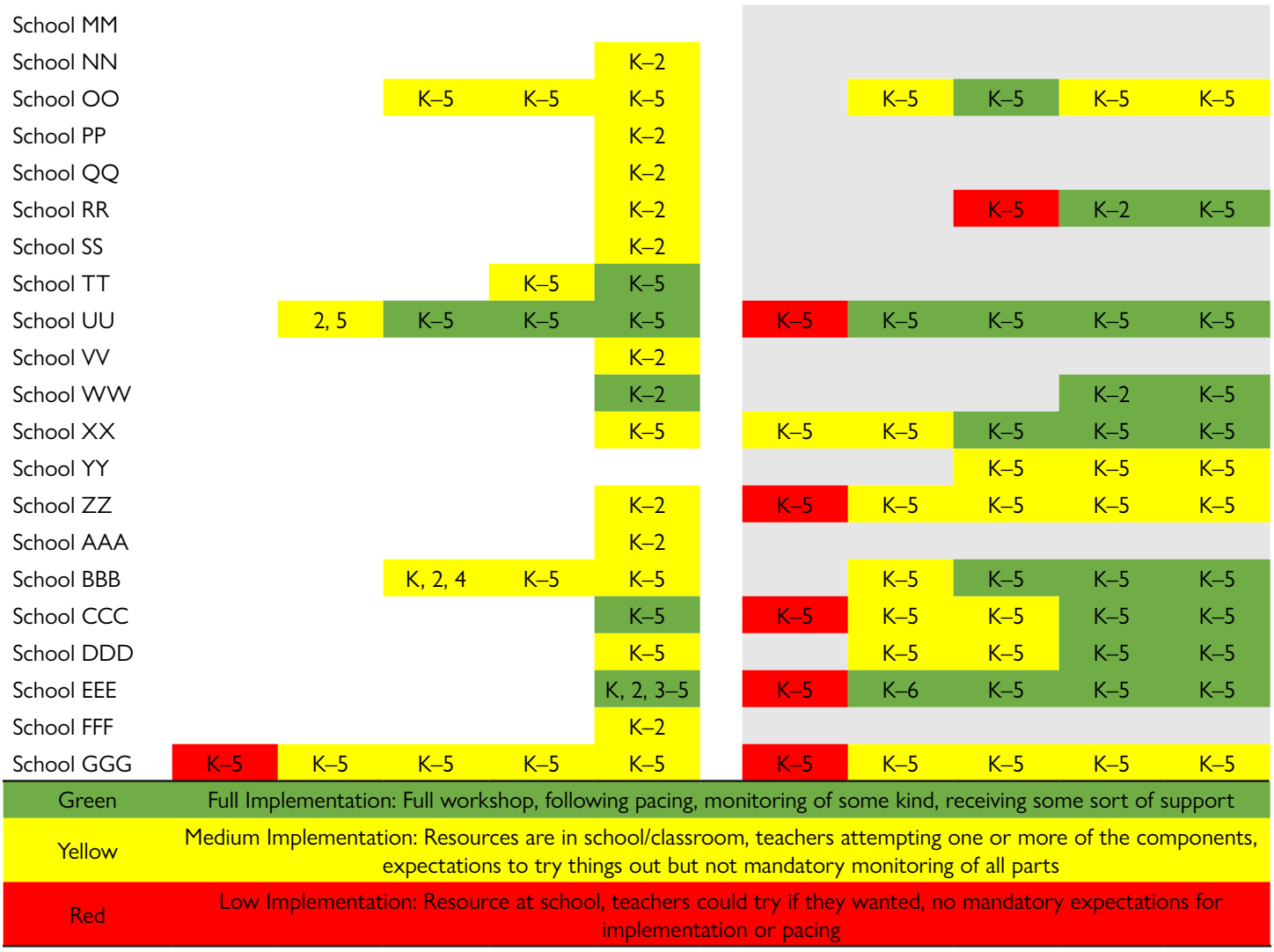
A panel of reading experts recently criticized the Lucy Calkins Program,[1] and some teachers and administrators within the district have questioned the program's efficacy and cost. In this report, we seek to provide evidence on the impact of the Lucy Calkins Program on student achievement in reading/language arts.[2]

Data and Methodology

The main analysis sample consists of students taking summative assessments in Grades 3–5 in traditional public schools in a metro-Atlanta school district from SY 2012–13 to SY 2018–19. The sample includes over 75,000 student-year observations. Our secondary analysis sample, which captures students taking formative assessments in Grades K–3, is limited to just over 12,000 student-year observations over the period SY 2017–18 to SY 2018–19. Consequently, estimates from the secondary analysis are less precise, making it more difficult to determine if effects are truly different from zero.

Table 1. Lucy Calkins Program Implementation by School, Year, and Program Intensity

School	Reading					Writing				
	2015	2016	2017	2018	2019	2015	2016	2017	2018	2019
School A										
School B			5	3–5	K–5	K–5	K–5	K–5	K–5	K–5
School C				K–5	K–5				K–5	K–5
School D				3–5	3–5			K–5	K–5	K–5
School E					K–2					
School F			2	K–5	K–5	K–5	K–5	K–5	K–5	K–5
School G					K–2					
School H				K–2	K–5				K–2	K–2
School I				K–5	K–5				K–5	K–5
School J				K–5	K–5	K–5	K–5	K–5	K–5	K–5
School K					K–2					
School L										
School M	1, 4	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5
School N					K–5	K–3	K–4	K–5	K–5	K–5
School O										
School P					K–5					K–5
School Q				K–2	K–5					K–2
School R		K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5
School S					K–2					
School T			K–5	K–5	K–5	K–5	K–7	K–5	K–5	K–5
School U					K–2					
School V										
School W				K–5	K–5				K–5	K–5
School X			K–5	K–5	K–5			K–5	K–5	K–5
School Y					K–5	K–5	K–5	K–5	K–5	K–5
School Z					K–2					
School AA								K–5	K–5	K–5
School BB										
School CC	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5
School DD									1	1
School EE						K–5	K–5	K–5	K–5	K–5
School FF		K–5	K–5	K–5	K–5		K–5	K–5	K–5	K–5
School GG					K–5					
School HH		K	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5
School II										
School JJ					2–3	K–5	K–5	K–5	K–5	K–5
School KK	5	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5
School LL		K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5	K–5



In our main analysis, we utilize scores on end-of-grade summative assessments (Georgia's CRCT and Milestones exams) for English language arts (ELA) as our measure of student achievement.[3] Given that end-of-grade assessments are administered to students in Grades 3–5, we can only measure annual learning gains for students in Grades 4 and 5. Test scores are "normalized" so that a value of zero represents the statewide mean for each grade and year; the units of measure are statewide standard deviations from the mean or "effect sizes." Our main analysis estimates the effect of Lucy Calkins Program use on individual-level ELA achievement scores, holding constant prior-year scores in ELA and math, student demographics, and a measure of prior-year student behavior.

Because schools chose whether to adopt the Lucy Calkins Program, student exposure to the program differed in a non-random fashion that could conflate program effects with other attributes of the adopting schools. To avoid potential bias, we compare student test scores in a school and grade that had implemented the Lucy Calkins Program to students with similar characteristics and prior-year scores from the same school in different grades and years who were not exposed to the program. We also control for districtwide variation in student performance over time. As shown in Table 1, there is more within-school variation in Lucy Calkins Program implementation over time for reading than writing. Because of this, within-school estimates of program impact are more precise for reading than for writing.

Because statewide achievement test scores are not available in grade levels below Grade 3, we gauge the impact of exposure to the Lucy Calkins Program in the early-elementary grades in two ways. First, we follow the same approach for Grades 4–5 but employ the FastBridge formative assessment scores in reading as the outcome measure.[4] Given that the FastBridge exam is nationally-normed, the test scores are expressed in standard deviations from the national average of scores at each grade level. Data from the National Assessment of Educational Progress (NAEP) for Grade 4 reading indicate that Georgia NAEP scores across the distribution are quite close to the national scores.[5] Thus, the normalized FastBridge scores should be comparable to the normalized scores for the Milestones and CRCT state assessments.

Second, we use Milestones/CRCT data to estimate a model of Grade 3 achievement levels as a function of exposure to the program in each grade from kindergarten to Grade 3. As with the analysis of test-score gains, we control for observable student characteristics. Thus, we compare the Grade 3 scores of students with similar characteristics who attended the same school in Grade 3 but who have differing levels of exposure to the Lucy Calkins Program due to when they were enrolled in Grades K–3. This analysis of test-score levels, however, does not directly control for prior learning.

Research Questions

We address three research questions:

1. What is the average impact of utilizing the Lucy Calkins Program on student achievement (relative to other reading/writing programs)?

2. How does the impact of the Lucy Calkins Program on student achievement vary with the level of implementation?
3. To what extent does the impact of the Lucy Calkins Program vary by student characteristics, such as eligibility for free or reduced-price meals and English language learner status?

Finding 1: Average Impacts on Achievement Gains

The direction of the effects of the Lucy Calkins Reading and Writing Programs on student achievement vary by grade but are not statistically significant.

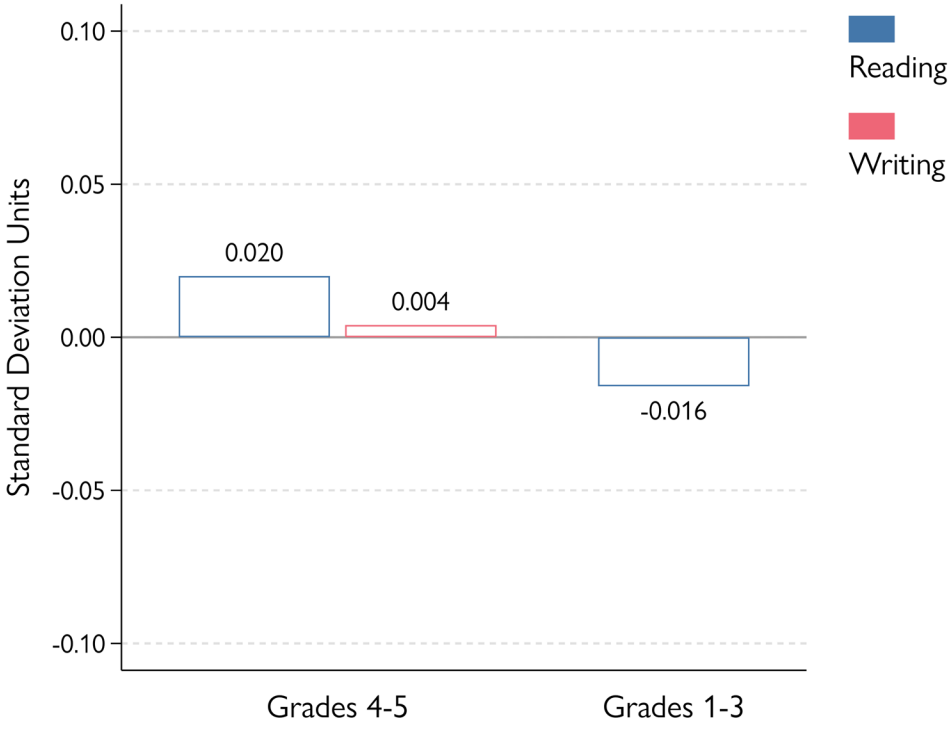
The estimated average impacts of the Lucy Calkins Reading and Writing Programs on annual student achievement gains in Grades 4 and 5 are positive but small and cannot be confidently distinguished from zero (no effect). In Grades 1–3, use of the reading program is estimated to have a small negative effect on student achievement, but this estimated effect cannot confidently be distinguished from zero.

The left side of Figure 1 shows the impact of the implementation of the Lucy Calkins Reading and Writing Programs on student ELA achievement gains in Grades 4 and 5, regardless of implementation level (i.e., intensity). The estimated impact of the reading program on annual student achievement gains is 2% of a standard deviation in statewide ELA test scores. This is equivalent to a student moving from the statewide average score (i.e., the 50th percentile) to the 50.8 percentile (or about one-quarter of the typical difference in gains from having a first-year teacher versus a teacher with three to five years of experience). The estimated impact of the writing program is 75% smaller, less than one-half of 1% of a standard deviation. These estimates are relatively noisy, and we cannot say with confidence that the estimated effects are not zero.

The right side of Figure 1 shows the Lucy Calkins Reading Program's estimated impact on reading score gains in Grades 1–3. The estimated impact of the reading program is less than a 2% reduction in annual student achievement gains in reading, though the effect cannot confidently be distinguished from zero.[6]

Figure 2 presents the estimated impacts of the Lucy Calkins Reading Program on ELA achievement levels in Grade 3. These estimates should be viewed with

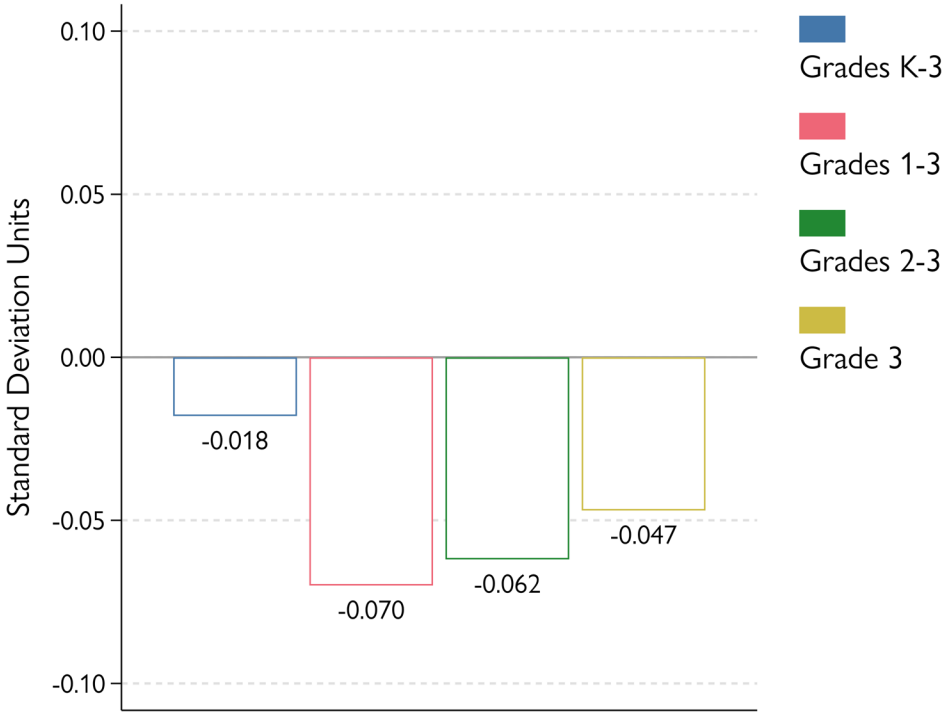
Figure 1. Estimated Impact of Lucy Calkins Reading and Writing Program Implementation on Student Annual Achievement Gains by Grade Group—Within-School Comparison (Standard Deviation Units)



Notes. Outlined bars are not statistically significant at the 95% confidence level.

some caution as there are no controls for prior achievement in the model. Consistent with the FastBridge analysis of achievement gains in Grades 1–3, the estimated impacts are all negative. The estimated impact for students exposed each year from kindergarten through Grade 3 is small: less than 2% of a standard deviation. In contrast, results for students who were first exposed to the Lucy Calkins Reading Program in Grades 1, 2, or 3 yield larger negative impacts on achievement in Grade 3. This suggests that switching programs mid-stream may have harmful effects on student achievement in the early-elementary grades.

Figure 2. Estimated Impact of Lucy Calkins Reading Program Implementation on Student Achievement Levels in Grade 3 by Grade Levels of Exposure—Within-School Comparison (Standard Deviation Units)



Notes. Outlined bars are not statistically significant at the 95% confidence level.

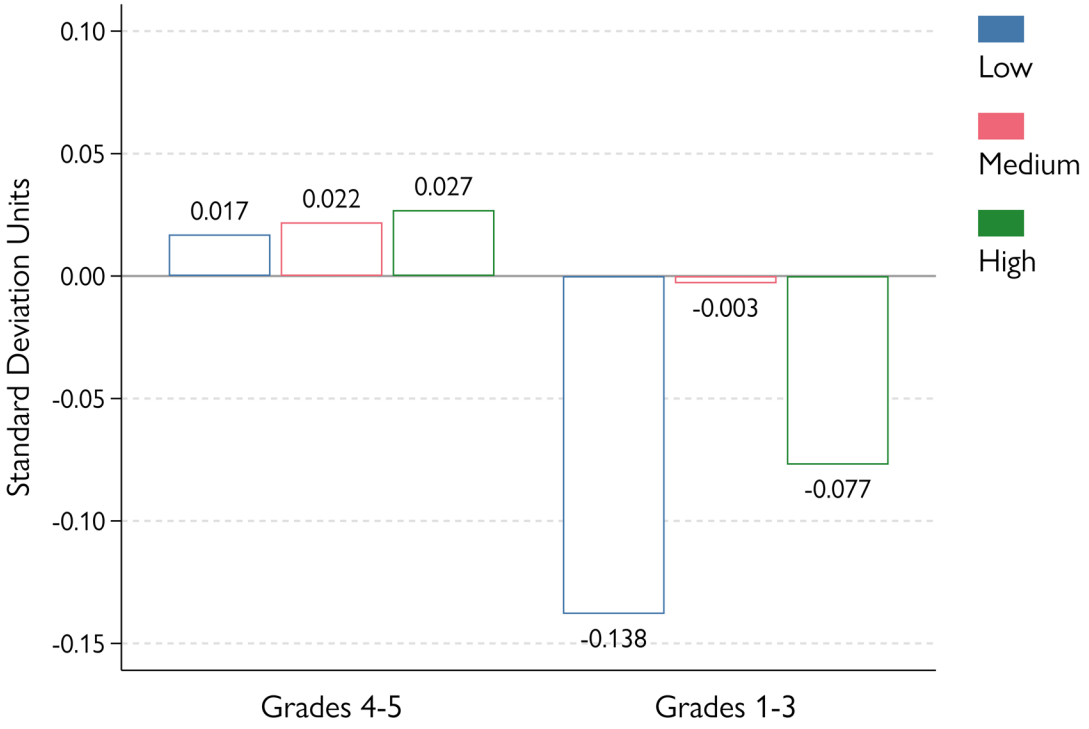
Finding 2: Variation in Impacts by Implementation Level

Effects vary with the level of implementation, but the impacts are inconsistent across subjects and by grade and are not statistically significant.

Impacts of the Reading Program in Grades 4 and 5 appear to increase with the level of implementation, though we cannot say with confidence that the impacts differ significantly across implementation levels. There is no clear relationship between program impacts and implementation level at the lower elementary level (Grades 1–3). Likewise, the impacts of the writing program do not appear to increase with the degree of implementation.

As illustrated in Table 1, there was considerable variation in program implementation within schools over time. Typically, schools started at a low level of implementation (denoted by red shading in Table 1) or a medium level

Figure 3. Estimated Impact of Lucy Calkins Reading Program Implementation on Student Annual Achievement Gains by Intensity of Implementation and Grade Group—Within-School Comparison (Standard Deviation Units)

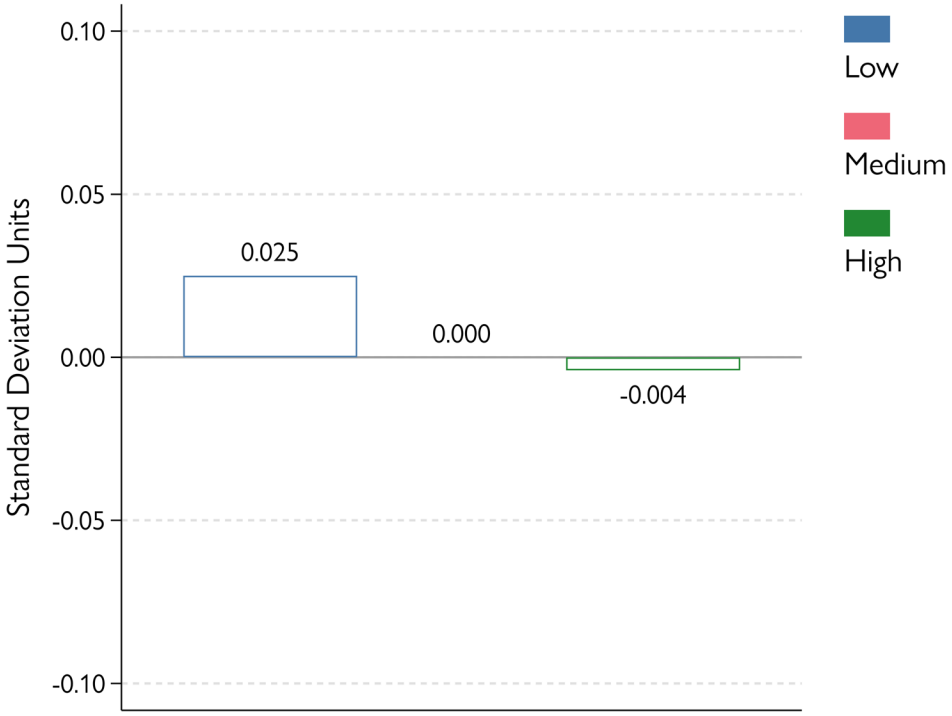


Notes. Outlined bars are not statistically significant at the 95% confidence level.

of implementation (shaded in yellow) and then increased their implementation level over time. Low implementation means there are resources at the school, and implementation is up to the teacher (with no expectations for implementation). Full implementation means there is monitoring of some sort, and teachers receive some support.

To determine if estimated program impacts vary with the level of implementation, we re-estimated our models to allow for differential impacts across implementation levels. As illustrated in Figure 3, impacts of the Reading Program in the upper-elementary grades increase with the level of implementation, though we cannot say with confidence that the impacts differ significantly across implementation levels due to variability in the estimates. In the lower-elementary grades, the impacts vary widely by implementation level with no clear pattern. Due to uncertainty in the estimates, we cannot rule out the possibility that there is no difference in early-elementary impacts across implementation levels.

Figure 4. Estimated Impact of Lucy Calkins Writing Program Implementation on Student Annual Achievement Gains in Grades 4 and 5 by Intensity of Implementation—Within-School Comparison (Standard Deviation Units)



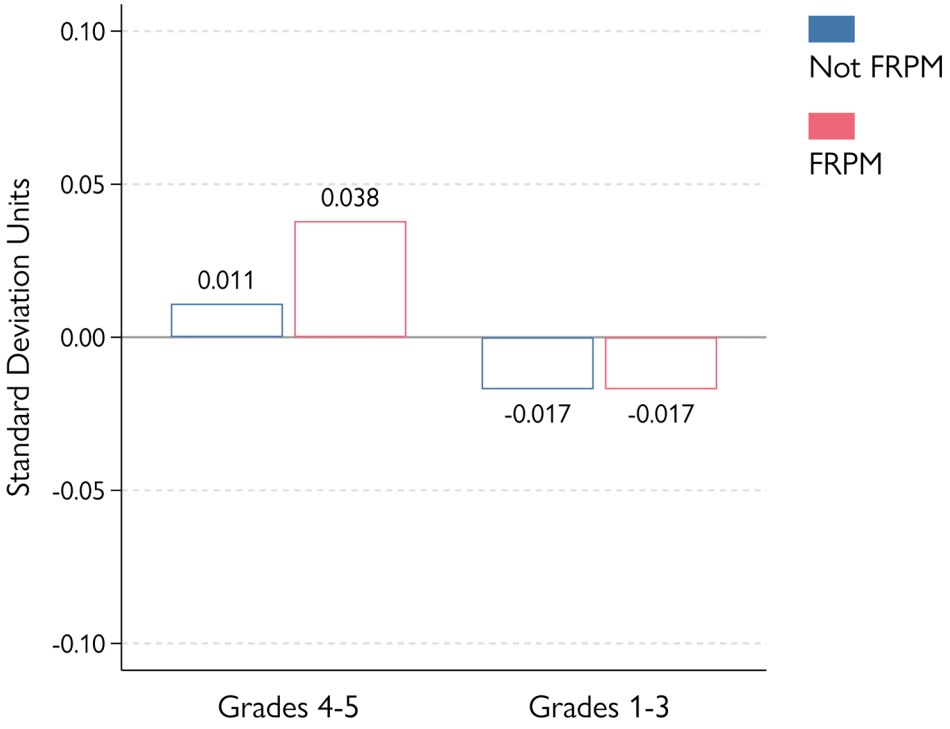
Notes. Outlined bars are not statistically significant at the 95% confidence level.

As shown in Figure 4, the impacts of the Writing Program on achievement do not increase with the degree of implementation. The estimated effect for low implementation is positive, while impacts for medium or high implementation are virtually zero. Due to the variability in the estimates, we cannot say with confidence that impacts at any of the three implementation levels are not zero.

Finding 3: Variation in Impacts by Student Characteristics

Impacts of the Reading Program in Grades 4 and 5 appear to be higher for students experiencing poverty than for students from more affluent families and higher for English language learners than for students proficient in English.

Figure 5. Estimated Impact of Lucy Calkins Reading Program Implementation on Student Annual Achievement Gains by Free/Reduced-Price Lunch Status—Within-School Comparison (Standard Deviation Units)

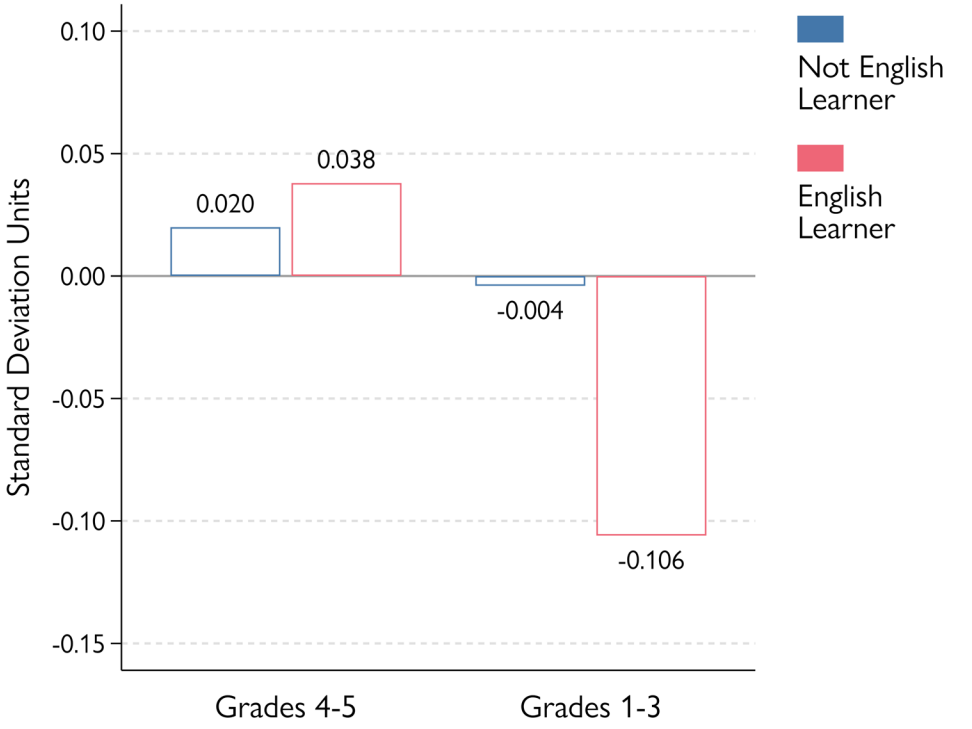


Notes. Outlined bars are not statistically significant at the 95% confidence level.

To determine if different types of students experienced differential benefits from exposure to the Lucy Calkins Reading and Writing Programs, we analyzed program impacts for students who were eligible for free or reduced-price meals—a proxy measure for poverty—and English language learners. Results for the Reading Program are presented in Figures 5 and 6, and results for the Writing Program are provided in Figure 7.

While the differences are not statistically significant, the estimates from the analysis in Grades 4 and 5 suggest that students experiencing poverty and English language learners gain more from exposure to the Lucy Calkins Programs than do students from more financially-advantaged families or students who are proficient in English. In Grades 1–3, we do not observe a difference in achievement gains for students experiencing poverty. English language learners are estimated to experience a reduction in achievement gains of 0.106 standard deviation units, which is markedly different from the achievement gains among students who are not English language learners. Due

Figure 6. Estimated Impact of Lucy Calkins Reading Program Implementation on Student Annual Achievement Gains by English Language Learner Status—Within-School Comparison (Standard Deviation Units)



Notes. Outlined bars are not statistically significant at the 95% confidence level.

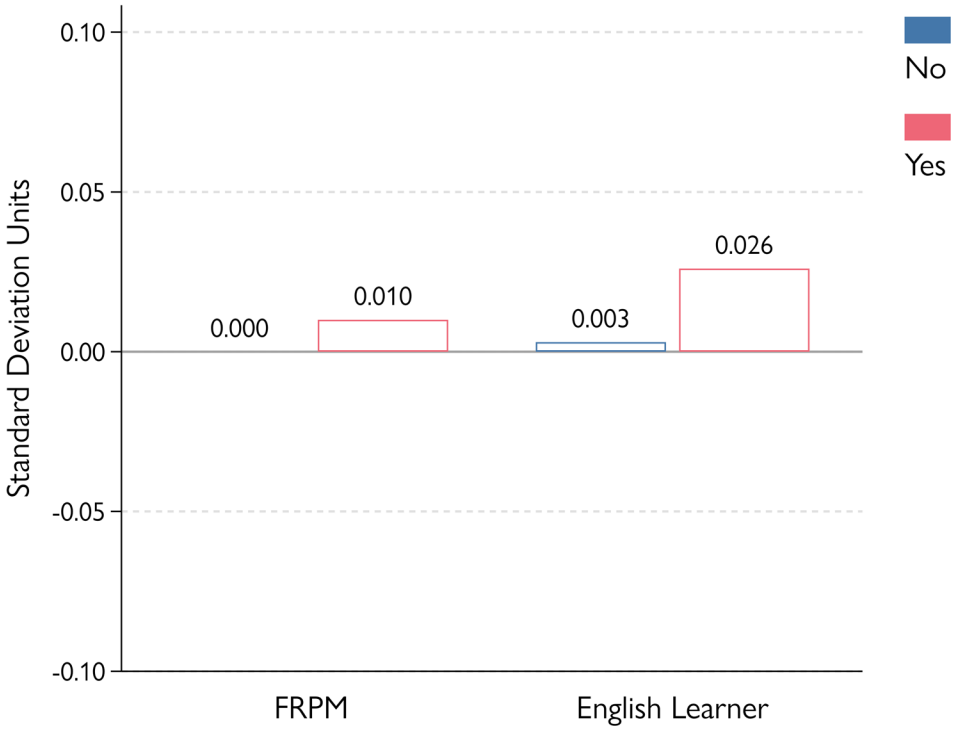
to the small sample of English language learners, however, we cannot rule out that the true difference is zero.

Conclusions

In a school district in metro Atlanta, the Lucy Calkins Programs in reading and writing have, at most, very small average effects on student achievement in ELA relative to the "business as usual" approach of using other reading and writing programs. There is at least suggestive evidence, however, that students experiencing poverty and English language learners may benefit more from the Lucy Calkins Program than other more advantaged students in Grades 4–5.

Our ability to measure program impacts in Grades 1–3 is limited by the fact that state assessments are not administered in the lower-elementary grades, and the sample of students who take formative assessments is relatively small. Given the available data, we find no evidence that the Lucy Calkins Reading

Figure 7. Estimated Impact of Lucy Calkins Writing Program Implementation on Student Annual Achievement Gains by Free/Reduced-Price Lunch Status and by English LanguageLearner Status—Within-School Comparison (Standard Deviation Units)



Notes. Outlined bars are not statistically significant at the 95% confidence level.

Program has a positive effect on reading achievement gains in Grades 1–3. There is also some suggestive evidence that switching to or from the Lucy Calkins Reading Program (from some other program) in Grades K–3 has a negative effect on Grade 3 reading scores. This suggests that if a different program is selected, it would be worthwhile to consider a phased approach whereby the new program is initially offered to a kindergarten cohort and then expanded to additional cohorts over time.

Endnotes

1. Adams, Marilyn J., Lily W. Filmore, Claude Goldenberg, Jane Oakhill, David D. Paige, Timothy Rasinski, and Timothy Shanahan. 2020. "Comparing Reading Research to Program Design: An Examination of Teachers College Units of Study." Student Achievement Partners. Retrieved from achievethecore.org/page/3240/comparing-reading-research-to-program-design-an-examination-of-teachers-college-units-of-study

Hanford, Emily. 2020. "Influential Literacy Expert Lucy Calkins is Changing Her Views." American Public Media Reports. October 16, 2020. Retrieved from apmreports.org/story/2020/10/16/influential-literacy-expert-lucy-calkins-is-changing-her-views

2. We initially planned to study how the efficacy of the Lucy Calkins Program varied with the extent of program-specific professional development for teachers. Unfortunately, the available data on professional development was not sufficient to support such an analysis.

3. The Georgia Criterion-Referenced Competency Tests (CRCT) program was used until SY 2013–14. The Georgia Milestones Assessment System replaced the CRCT in SY 2014–15.

4. There is no FastBridge Language Arts (writing and grammar) exam. Consequently, we did not examine the impact of the writing program in Grades 1–3. Another formative assessment—the iReady exam—was used on a very limited basis in elementary grades in the district prior to SY 2018–19. It eventually replaced the FastBridge assessment in SY 2019–20.

5. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013, 2015, 2017, and 2019 Reading Assessments.

6. To check comparability, we also estimated effects of the Lucy Calkins Reading Program in Grades 4 and 5 using FastBridge scores. The results were similar to those from the main analysis: The estimated average impact was a statistically significant 0.09.

About the Authors

Béla Figge

Béla Figge is a fifth-year doctoral student in economics at Georgia State University and a graduate research assistant with the Georgia Policy Labs. He received his bachelor's degree in economics from Bennington College and pursued his graduate studies in economics at the University of Mannheim in Germany. His research centers on policy evaluation.

Tim R. Sass

Tim Sass is a Distinguished University Professor in the department of economics at Georgia State University and the W.J. Usery Chair of the American Workplace in the Andrew Young School of Policy Studies. He is also the faculty director of the Metro Atlanta Policy Lab for Education (MAPLE). His research interests include the teacher labor supply, the measurement of teacher quality, and school choice. His work has been published in numerous academic journals and has been supported by several federal and philanthropic grants. He has acted as a consultant to school systems across the country. He is also a senior researcher at the Center for Analysis of Longitudinal Data in Education Research (CALDER).

About the Georgia Policy Labs

The Georgia Policy Labs is an interdisciplinary research center that drives policy and programmatic decisions that lift children, students, and families—especially those experiencing vulnerabilities. We produce evidence and actionable insights to realize the safety, capability, and economic security of *every* child, young adult, and family in Georgia by leveraging the power of data. We work alongside our school district and state agency partners to magnify their research capabilities and focus on their greatest areas of need. Our work reveals how policies and programs can be modified so that every child, student, and family can thrive.

Housed in the Andrew Young School of Policy Studies at Georgia State University, we have three components: the Metro Atlanta Policy Lab for Education (metro-Atlanta K-12 public education), the Child & Family Policy Lab (supporting children, families, and students through a cross-agency approach), and the Career & Technical Education Policy Exchange (a multi-state consortium exploring high-school based career and technical education).

Learn more at gpl.gsu.edu.