AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Teachers College Reading and Writing Project Study

January 2021

Abstract

This brief presents study findings on the association between school adoption of the Teachers College Reading and Writing Project (TCRWP) approach and state English language arts (ELA) test scores. The TCRWP approach, a curriculum and professional development for teaching reading and writing, is widely used across the country and around the world. Analyses for this brief used publicly available school-level data from New York City, New York, public schools and schools in four districts in Greater Atlanta, Georgia. A comparative interrupted time series analysis examined changes in ELA scores for Grades 3–5 for a sample of TCRWP schools, following their adoption of the approach, compared to similar schools that did not adopt the approach. Consistent with prior literature on professional development for teaching reading and writing, we found no change in ELA scores 1 year after initial TCRWP implementation. Beginning in the 2nd year following TCRWP implementation, however, we observed statistically significant increases in ELA scores among TCRWP-implementing schools, as compared with the matched comparison schools. Between 5 and 7 years following adoption, ELA scores in TCRWP schools were higher by 0.22–0.38 standard deviations, suggesting cumulative effects of use of the TCRWP approach.

Introduction

The Teachers College Reading and Writing Project (TCRWP) is a widely used approach to literacy instruction for students in elementary and middle schools. Throughout its history, the two primary goals of TCRWP have been to improve students' reading and writing and to help children become lifelong, confident readers and writers who display agency and independence through work with teachers, collaboration with peers, and independent work.

Although the approach is used widely, it has never been subjected to a rigorous evaluation in which the reading achievement of schools implementing TCRWP is compared with that of non-TCRWP schools. Researchers at the American Institutes for Research (AIR) conducted such a study and looked specifically at TCRWP schools in New York City (NYC), New York, and in four districts in Greater Atlanta, Georgia. Specifically, we investigated the extent to which the literacy achievement of schools whose teachers were implementing the TCRWP approach differed from that of similar schools whose teachers were not. The study used a quasi-experimental design—specifically, a comparative interrupted time series (CITS) approach—to analyze changes in test scores following TCRWP implementation. Propensity score matching was used to identify a sample of comparison schools resembling the TCRWP schools on prior English language arts (ELA) achievement and key demographics. In addition to the examination of school-level literacy achievement, a small-scale implementation study was conducted in eight of the TCRWP schools in Georgia.

The study's original intent was to examine school-level achievement through the spring 2020 state testing cycle. COVID-19-related school closures and the cancellation of spring 2020 state testing negated this plan.

This paper focuses on the results of analyses of school-level data on ELA test scores from the 2010–11 through the 2018–19 school years for NYC schools and from 2014–15 through 2018–19 for Georgia schools.[1] The main finding from these analyses is that although there is no significant difference in ELA achievement for TCRWP and comparison schools in the first year after the program is introduced, in subsequent years, TCRWP schools show achievement that is not only higher, but statistically significantly higher, relative to schools in the comparison group. The differences in test scores between TCRWP schools and comparison schools grow larger the longer the schools have been implementing the program. After describing the study's methods and findings, this paper concludes with some speculation about why the test scores in TCRWP schools trend upward, relying for this speculation on research on instructional practice and, to a lesser extent, on data gathered from teachers in a small-scale implementation study conducted in eight of the TCRWP schools in Georgia.

[1] See Hallberg et al., 2018, and Jacob et al., 2014, for discussions of use of aggregate data.

Background

The TCRWP approach is widely used across the country and around the world. TCRWP staff host workshops and seminars and offer a wide range of print and online resources for professional development. Curriculum guides[2]—the *Units of Study in Reading* and the *Units of Study in Writing*—and numerous other books about the approach are readily available.[3] The *Up the Ladder*[4] series provides guidance to help teachers introduce the approach to upper elementary students who have not previously experienced a similar instructional method and for students who are not performing at grade level. Digital and video resources are also available.

When schools adopt TCRWP, participating teachers receive grade-specific *Units of Study* in reading and writing as comprehensive, yearlong curriculum guides. They also may receive TCRWP Classroom Libraries,[5] which feature informational and narrative texts at varying reading and complexity levels to give students opportunities to read independently, move up levels of complexity, and advance their content knowledge. Teachers use books from this collection and from their own classroom libraries when conducting *read alouds*, during which students are encouraged to discuss and analyze what they hear. *Mentor texts* are also available to provide models of good genre-specific writing. In addition to reading and writing achievement, a goal for students in TCRWP classrooms is development of a *literacy self-concept*, that is, a sense of themselves as engaged and active readers and writers.

Teachers have many opportunities for in-person and virtual support and professional development as part of their participation in TCRWP, including on-site support from Teachers College staff developers. Teachers may also seek out professional development on their own, for example, signing on to the "Office Hours With Lucy Calkins."[6] Other opportunities, such as attendance at weeklong summer institutes[7] or local "home-grown" workshops, may be sponsored by schools or districts. These offerings, for which attendance is usually voluntary, seek to improve teachers' understanding of particular instructional practices, including differentiation, and of the classroom management strategies needed to encourage students to work independently and collaboratively as teachers work with small groups or individuals.

The framework for the TCRWP approach is the workshop, a series of interactions between teachers and students and among students. The workshop framework provides the structure for teaching the *Units of Study* in reading and writing. Each session in each unit is intended for 1 day, and each day's workshop is supposed to last 50 to 60 minutes. The workshop begins with a teacher-led minilesson that provides explicit instruction to the whole class. Minilessons typically last approximately 10 minutes, and, although the content of minilessons changes from day to day, the structure remains the same. Minilessons begin with

---

[2] The *Units of Study* are grade-level-specific curriculum guides in reading and writing, written by Lucy Calkins and colleagues at the Teachers College Reading and Writing Project. They are published by Heinemann; see https://www.unitsofstudy.com/.

[3] *Units of Study in Phonics* for kindergarten to Grade 2 are also available but were not included in the resources provided to teachers in the AIR study.

[4] See: https://www.unitsofstudy.com/uptheladder/.

[5] See: https://www.unitsofstudy.com/classroomlibraries/.

[6] See: https://webinars.heinemann.com/calkins-office-hours.

[7] See: https://readingandwritingproject.org/services/institutes.

the teacher connecting the day's teaching with the ongoing work that students have been doing and introducing the teaching point for the day. Next, the teacher demonstrates the step-by-step way to accomplish or use the skill or strategy that is to be the focus of the teaching point. After the demonstration, students try to do what the teacher demonstrated, typically with a partner. The minilesson ends with students dispersing to their own independent work. During the independent reading or writing time that follows the minilesson, the teacher confers with students and leads small groups. Partway through the independent work time, the teacher may stand and deliver a mid-workshop teaching point. The workshop typically ends with an opportunity for students to share what they are working on or what they have learned.

Study Methods

Data and Sample

The key data sources for this study are publicly available school-level data on ELA test scores and student demographics for public schools from NYC and four districts in Georgia. For NYC schools, we use data from school years 2010–11 through 2018–19, and, for Georgia schools, we use data from school years 2014–15 through 2018–19. The ELA test scores were average scores on the New York State ELA Test and the Georgia Milestones End-of-Grade ELA tests for students in Grades 3, 4, and 5. To ensure comparability of scores between grades and across years, we standardized the scores within state, year, and grade, using student-level standard deviations calculated assuming an intracluster correlation (ICC) of 0.20.[8] For the main analysis, the outcome measure is the school average of standardized scores across Grades 3, 4, and 5. The school-level demographic data include information on percentage of students eligible for free or reduced-price lunch (FRPL), percentage of students by race/ethnicity and gender, percentage of students eligible for special education services, and percentage of students who are English learners (ELs).

Treatment schools are defined as schools that adopted the grade-appropriate *Units of Study in Reading* and *Units of Study in Writing* as their literacy curricula. Teachers in these schools also received professional development services from TCRWP staff.[9] The duration of the professional development typically ranged from 5 to 10 days per year and was tailored to the specific needs of each participating school.[10] The support typically involved TCRWP staff working with groups of teachers or individual teachers on implementing the curriculum, conducting model lessons, or coteaching with individual teachers.[11] The treatment sample consists of 40 schools from NYC that began implementing TCRWP from

---

[8] Student-level standard deviations (SDs) calculated based on the formula $SD_{school-level} = SD_{student-level} * \sqrt{ICC}$. For details about baseline equivalence criteria and choice of ICC, see the *What Works Clearinghouse (WWC) Procedures Handbook* (https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf).

[9] To be considered a treatment school, schools had to have continued to receive professional development services from TCRWP staff over the course of the study period (2012–2020) *and* continued to use the *Units of Study*. Therefore, schools that continued to use the *Units of Study* but did not continue to receive professional development services from TCRWP staff were not included as treatment schools.

[10] AIR researchers did not observe any of the professional development or obtain documentation of the dosage teachers received.

[11] The 5 to 10 days of services were generally divided among teachers in Grades 3–5. Typically, TCRWP is adopted as a whole-school approach, so schools typically receive services for K–2 teachers as well. However, not all TCRWP schools that participated in the current study received services for K–2 teachers.

2012–13 through 2018–19 and 11 schools from Georgia that began implementing TCRWP from 2016–17 through 2018–19. Table 1 shows the number of treatment schools by state and implementation cohort—that is, the year when schools adopted the *Units of Study* as their language arts curriculum—and the number of corresponding matched comparison schools. We describe the matching methodology in detail below.

Table 1. TCRWP Schools and Matched Comparison Schools

Implementation Year	TCRWP Schools	Matched Comparison Schools	Total
NYC			
2012–13	7	28	35
2013–14	4	16	20
2015–16	7	28	35
2016–17	6	20	26
2017–18	9	36	45
2018–19	7	23	30
Georgia			
2016–17	2	3	5
2017–18	2	5	7
2018–19	7	19	26
Total	51	178	229

Note. Based on a power analysis conducted by AIR, Teachers College staff identified a sample of 51 treatment schools for inclusion in the study. There is no row for 2014–15 because none of the identified/sampled treatment schools began implementing TCRWP that year.

Matching Method

A matched comparison group was constructed using nearest-neighbor matching on propensity scores derived from test scores and demographics from 1 year prior to implementation. Cohorts were matched sequentially, such that comparison schools matched to treatment schools in any given implementation cohort were dropped when matching subsequent cohorts.[12]

To provide greater statistical power for the analysis, we selected multiple comparison schools for each treatment school. Given the high number of potential comparison schools in NYC and the additional statistical power provided by having a larger comparison sample, we matched each NYC treatment school

[12] We matched schools sequentially beginning from the earliest cohort in NYC and latest cohort in Georgia. Because the majority of treatment schools in Georgia adopted TCRWP in 2018–19 and the school districts in which Georgia treatment schools are located are relatively small, to ensure best matches for the largest number of schools, we implemented sequential matching beginning with the most recent implementation cohort (2018–19). For NYC, the number of matched schools was higher, as well as more balanced in terms of pre-implementation characteristics, when matching sequentially beginning with the earliest cohort.

with up to four comparison schools. In Georgia, where the school districts are smaller, we identified up to three comparison schools for each treatment school, and, to enhance the quality and comparability of the matches, we matched schools within the same district whenever possible.[13] Column 3 in Table 1 shows the number of comparison schools matched within each cohort by state. As noted earlier, there were a total of 51 treatment schools; the selected comparison sample consists of 178 schools. Some cohorts have fewer comparison schools matched to each treatment school (e.g., four treatment schools from NYC that implemented TCRWP in 2018–19 were matched to fewer than four comparison schools) because of the sequential matching.

As shown in Table 2, the treatment and comparison schools were well balanced—that is, equivalent at baseline—on pre-implementation characteristics. For ELA test scores, the baseline effect size between treatment and comparison schools is less than 0.05, which satisfies WWC standards for baseline equivalence.[14] For the demographic characteristics, the baseline effect sizes ranged from 0.03 to 0.12.

Table 2. Characteristics of Schools and Baseline Effect Sizes

Baseline Covariates (prior school year)	Treatment Schools (*n* = 51)	Matched Comparison Schools (*n* = 178)	Standardized Mean Difference (baseline effect size)[a]
ELA test scores (standardized)	-0.006 (0.416)	-0.007 (0.413)	-0.034
Percentage of students eligible for FRPL	68.84 (23.80)	68.13 (25.98)	0.028
Percentage of female students	48.61 (2.44)	48.70 (3.06)	-0.028
Percentage of White students	17.52 (20.76)	18.16 (23.65)	-0.029
Percentage of students eligible for special education services	19.35 (7.38)	18.87 (7.21)	0.065
Percentage of students who are ELs	17.36 (15.34)	15.61 (13.37)	0.122

Note. Standard deviations are in parentheses.

[a] Standardized mean difference = $\frac{Mean\ in\ Treated - Mean\ in\ Untreated}{Pooled\ Standard\ Deviation}$.

[13] Two Georgia treatment schools were matched with comparison schools in other districts.

[14] According to the WWC standards, baseline effect sizes with absolute values between 0.05 and 0.25 satisfy baseline equivalence as long as appropriate statistical adjustments (i.e., inclusion of control variables) are made in the subsequent impact analyses. For details, see https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_procedures_handbook_v4_draft.pdf. In our CITS model, we controlled for all school-level characteristics reported in Table 2.

Analytic Approach

To examine whether TCRWP implementation was associated with improved ELA scores, we conducted a CITS analysis with the following specification:

$$Y_{s,t} = \alpha_0 + \alpha_1 PY_1 + \alpha_2 PY_2 + \alpha_3 PY_3 + \alpha_4 PY_{4+} + \alpha_5 PY_5 + \alpha_6 PY_6 + \alpha_7 PY_7 + \alpha_8 X_{st} + School_s + Year_t + e_{st} \quad Eq\ (1)$$

In this specification,

$Y_{s,t}$, the outcome variable, is the mean standardized test score of School s in Year *t*.

$PY_1, PY_2, \ldots, PY_7$ are indicators for post-implementation Years 1, 2, … 7, respectively. Effectively, they take a value of 0 for all schools in pre-implementation years and a value of 1 for treatment schools in the corresponding post-implementation years.

X_{st} is a set of time-varying school characteristics, including % White students, % female students, % students who are ELs, % students eligible for FRPL, and % students eligible for special education services.

$School_s$ are school fixed effects, that is, indicator (0/1) variables for each school, which control for all time-invariant factors that could lead to differences in outcomes between schools. For example, test scores could differ between schools because of inherent fixed factors, such as location of the school.

$Year_t$ are year fixed effects, which account for changes in test scores over time that would affect treatment and comparison schools equally, for example, the test being harder or easier in a particular year for a particular grade.

Results

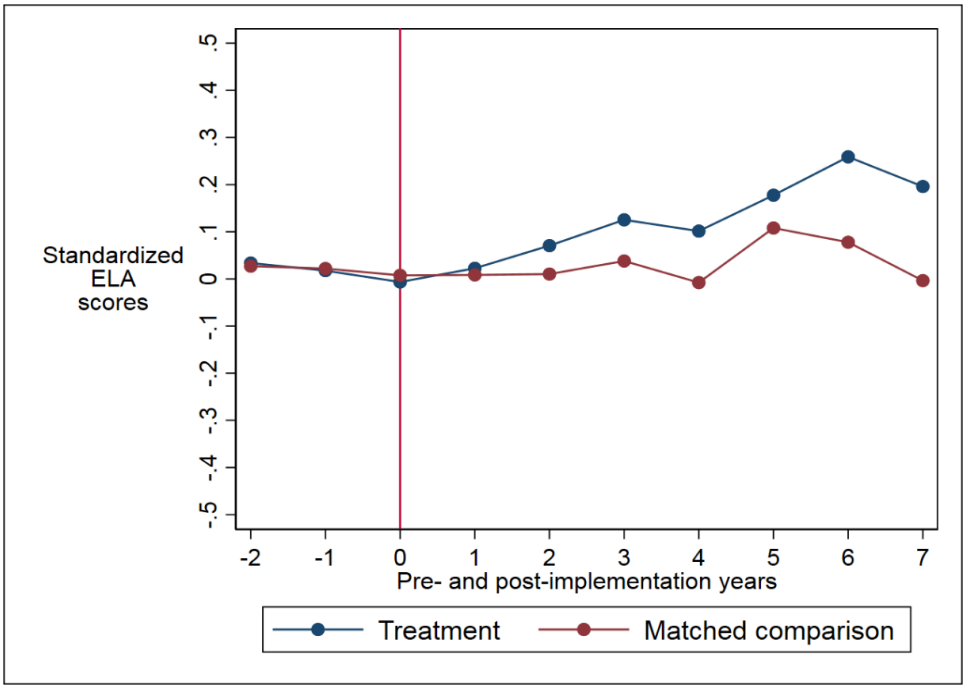
Main Results

The first phase of the analysis was to construct a graphical representation of changes in reading achievement of students in TCRWP classrooms compared with that of students in comparison classrooms during the years covered by this study. Figure 1 shows the unadjusted (i.e., not controlling for differences in school-level characteristics) pre- and post-trends of test scores for treatment schools and for matched comparison schools before and after TCRWP was adopted by treatment schools. The vertical line at 0 indicates the last test administration before treatment schools started implementing TCRWP. For example, for the 2018–19 cohort, 0 represents test scores from the 2018 state test administrations.

As the figure shows, there are no apparent differences in test scores between treatment and matched comparison schools in the 3 years prior to TCRWP implementation (i.e., years -2, -1, 0). Relatedly, the scores for both groups display the same trend throughout the 3 years, in that the trends are both relatively constant. However, beginning in the first TCRWP implementation year, we observe that treatment schools' test scores increased, while comparison schools' scores remained relatively constant. In later years, the two groups' scores follow relatively similar trends, but in all post-implementation years, treatment schools' scores remain higher than those scores of the comparison group in post-implementation years, and, in general, the differences widen over time. By the 5th and 6th years after the implementation

of TCRWP, treatment schools' scores have become higher than those scores of the comparison schools by 0.21 standard deviations.

Figure 1. Pre- and Post-Trends in Standardized ELA Test Scores for Treatment and Matched Comparison Schools, Descriptive Data



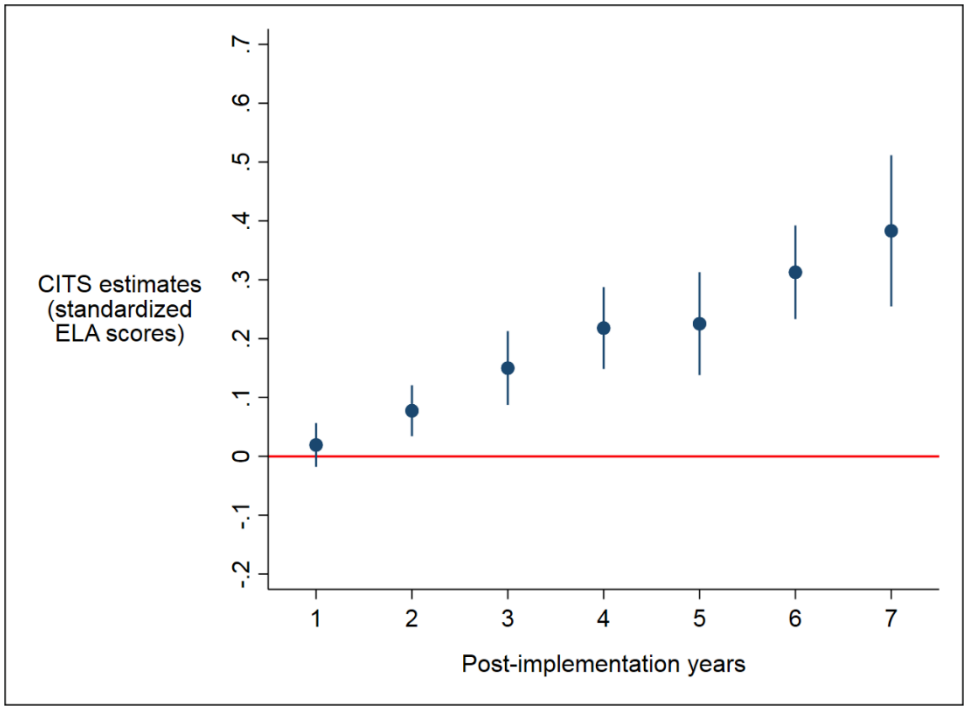
Note. Data points represent standardized average test scores from the 51 treatment schools (blue) and the 178 matched comparison schools (red). All schools in the sample have test score data for at least 1 year after implementation. However, the results for later years are based on fewer schools (e.g., seven schools that began implementing TCRWP in 2012–13 have seven years of post-implementation data).

The unadjusted scores presented in Figure 1 do not control for time-varying school-level characteristics, such as the percentage of students eligible for FRPL and the percentage of students who are ELs. Such characteristics have the potential to affect test scores and therefore can also affect observed differences in test scores between treatment and comparison schools. For this reason, we estimated a CITS model, which statistically takes such characteristics into account, to verify the descriptive evidence.

Figure 2 and Table 3 present results from the CITS analysis. In Figure 2, the data points and vertical bars represent coefficients from the CITS model (also presented in Table 3) and the corresponding confidence intervals. The horizontal red line at zero indicates no relationship between TCRWP implementation and ELA scores; points that have confidence intervals that do not cross the red line indicate a statistically significant relationship. As the figure shows, TCRWP implementation is associated with statistically significant positive effects on ELA test scores beginning in Year 2. Specifically, we see an effect of 0.08 and 0.15 standard deviations in standardized test scores 2 and 3 years after implementation, respectively. In addition, the difference between treatment schools and comparison schools becomes larger over time, with the treatment schools having test scores higher by 0.22 standard deviations relative to the matched comparison group as of the 4th year of implementation and approaching half of a standard deviation

(0.38 standard deviations) 7 years into implementation. WWC considers an effect size of 0.25 or larger as substantively important, a threshold these results approach in the 4th and 5th years and exceed in the 6th and 7th years.

Figure 2. Point Estimates and Confidence Intervals of Effect of TCRWP on Standardized Scores



Note. Data points represent coefficients from the CITS model, and the vertical bars represent the corresponding 95% confidence intervals. The horizontal red line at zero indicates no relationship between TCRWP implementation and state ELA scores; points that have confidence intervals that do not cross the red line indicate a statistically significant relationship. As previously noted, all schools in the sample have test score data for at least 1 year after implementation. However, the results for later years are based on fewer schools (e.g., seven schools that began implementing TCRWP in 2012–13 have seven years of post-implementation data).

Table 3. Estimates of Effect of TCRWP on Standardized Scores

Post-Implementation Year	Coefficients (standard errors)
1	0.019 (0.019)
2	0.077*** (0.022)
3	0.150*** (0.032)
4	0.218*** (0.035)
5	0.225*** (0.045)
6	0.313*** (0.041)
7	0.383*** (0.065)
N (school-year observations)	2,138[a]

Note. Robust standard errors, clustered at school level, in parentheses. All regressions include controls for % White students, % female students, % students who are ELs, % students eligible for FRPL, and % students eligible for special education services, as well as school and year fixed effects.

[a] Even though our analysis includes only 229 schools (51 treatment schools and 178 comparison schools), the CITS analysis has 2,138 observations. This is because each observation is not a school but rather *a school in a particular year*. For NYC schools, we have data for 10 years, from school years 2009–10 through 2018–19. Thus, for NYC schools, we have 191 schools and a total of (191*10 = 1,910 school-year observations). For each Georgia school (*N* = 38), we have data for 6 years, from the 2013–14 year through the 2018–19 year; thus, we have a total of (38*6 = 228) observations for Georgia. Therefore, we have 1,910 + 228 = 2,138 total observations.

+ $p < 0.10$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Subgroup Analyses

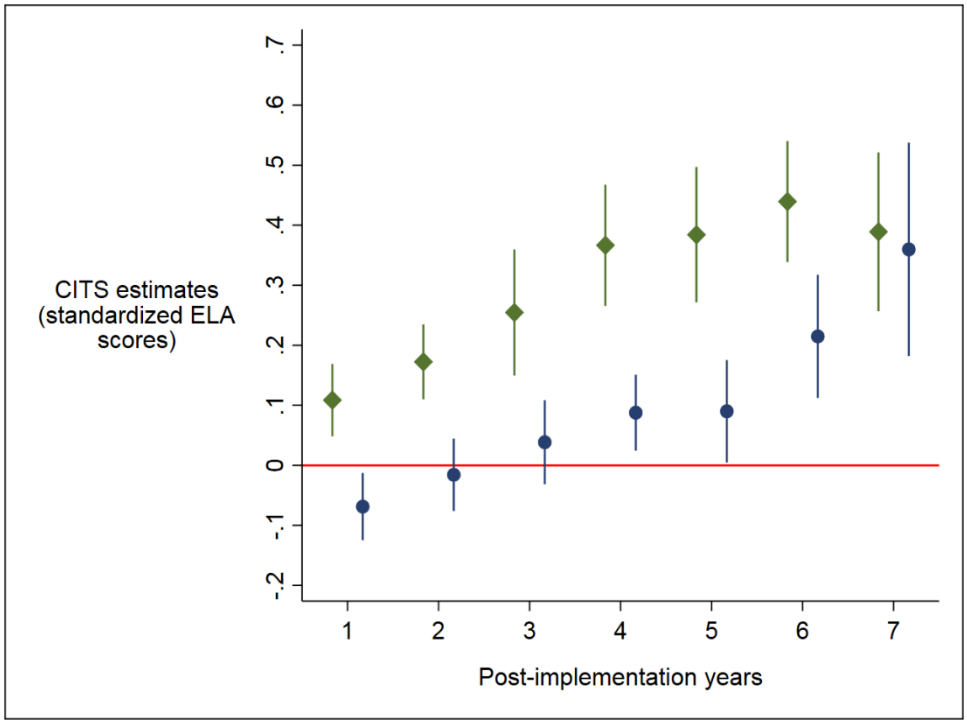
The overall results presented in the previous subsection indicate that, beginning in year 2, TCRWP implementation is associated with statistically significant positive effects on ELA test scores and that these effects increase as schools implement the approach for a longer time. However, the effectiveness of the program may vary for different subgroups of students. We tested whether the effects of TCRWP vary for schools serving high and low percentages of students who are (a) learning English and (b)impacted by poverty as indicated by eligibility for FRPL. We focus on these two characteristics of the student population because students who are learning English and students who qualify for FRPL often

face challenges in the classroom and have lower reading achievement (Belfi et al., 2016; Goddard et al., 2015; Kanno & Cromley, 2015; Ortiz & Robertson, 2018; Raudenbush, 2004).

We present results separately for subsamples of schools classified as having a low percentage of students who are ELs or a high percentage of students who are ELs. To classify treatment schools for this analysis, we compared the percentage of students who are ELs in each school in the year prior to implementation with the median percentage of students who are ELs among all treatment schools that year (i.e., by cohort). Treatment schools that had higher than the median percentage of students who are ELs were classified as *high % EL*, and remaining treatment schools were classified as *low % EL*. We estimated the same CITS specification described earlier using the subsample of treatment schools in each of the two categories and their corresponding matched comparison schools. Because of the way the schools were matched, however, it is not necessarily the case that the comparison schools for a treatment school with a high (or low) percentage of EL students would have fallen into the same category as the treatment school.

Figure 3 and Table 4 present estimates from the CITS model and corresponding confidence intervals for the subgroup analyses concerning students who are ELs. In comparing treatment schools with a low percentage of students who are ELs with their matched comparison schools (Table 4, Column 1), we found that TCRWP had a statistically significant positive association with improved test scores beginning in the 1st year of implementation, including every year thereafter, up to an effect of 0.44 standard deviations in the 6th year after implementation. On the other hand, Column 2 shows that treatment schools with a high percentage of students who are ELs had significantly lower test scores relative to their matched comparison schools in the 1st year after implementation and were statistically equivalent to the comparison schools in the 2nd and 3rd years. However, beginning in the 4th year, TCRWP implementation was correlated with improved test scores in these high-EL treatment schools relative to their comparison schools (0.09 standard deviations in Years 4 and 5, growing to 0.36 standard deviations in Year 7). A possible explanation for the negative and null earlier-year results in these high-EL schools is that students who are ELs often face multiple challenges in the classroom and with ELA assessments (Kanno & Cromley, 2015; Ortiz & Robertson, 2018). Overall, though, the results for these subgroup analyses reinforce the finding that longer implementation of the approach may help improve aggregate test scores, even in schools with a higher percentage of students who are ELs.

Figure 3. Point Estimates and Confidence Intervals of Effect of TCRWP on Standardized Scores, by Treatment School Percentage of EL Students



Note. Estimates from CITS model for the subsample of schools with low % students who are ELs (green) and high % students who are ELs, respectively (navy). For each cohort, treatment schools that had higher than the median percentage of students who are ELs in the year prior to TCRWP implementation were classified as *high % EL*, and remaining treatment schools were classified as *low % EL*. Data points represent coefficients from the CITS model, and the vertical bars represent the corresponding 95% confidence intervals. The horizontal red line at zero indicates no relationship between TCRWP implementation and state ELA scores; points that have confidence intervals that do not cross the red line indicate a statistically significant relationship. As previously noted, all schools in the sample have test score data for at least 1 year after implementation. However, the results for later years are based on fewer schools.

Table 4. Comparison of Estimates of Effect of TCRWP on Standardized Scores, by Treatment School Percentage of EL Students

Post-Implementation Year	Low % EL	High % EL
1	0.109***	-0.069*
	(0.031)	(0.029)
2	0.172***	-0.016
	(0.032)	(0.031)
3	0.255***	0.038
	(0.053)	(0.036)
4	0.367***	0.088**
	(0.051)	(0.032)
5	0.384***	0.090*
	(0.057)	(0.043)
6	0.439***	0.215***
	(0.051)	(0.052)
7	0.389***	0.360***
	(0.067)	(0.091)
N (school-year observations)	1,086	1,052

Note. Robust standard errors, clustered at school level, are in parentheses. All regressions include controls for % White students, % female students, % students who are ELs, % students eligible for FRPL, and % students eligible for special education services, as well as school and year fixed effects. For each cohort, treatment schools that had higher than the median percentage of students who are ELs in the year prior to TCRWP implementation were classified as *high % EL*, and remaining treatment schools were classified as *low % EL*.

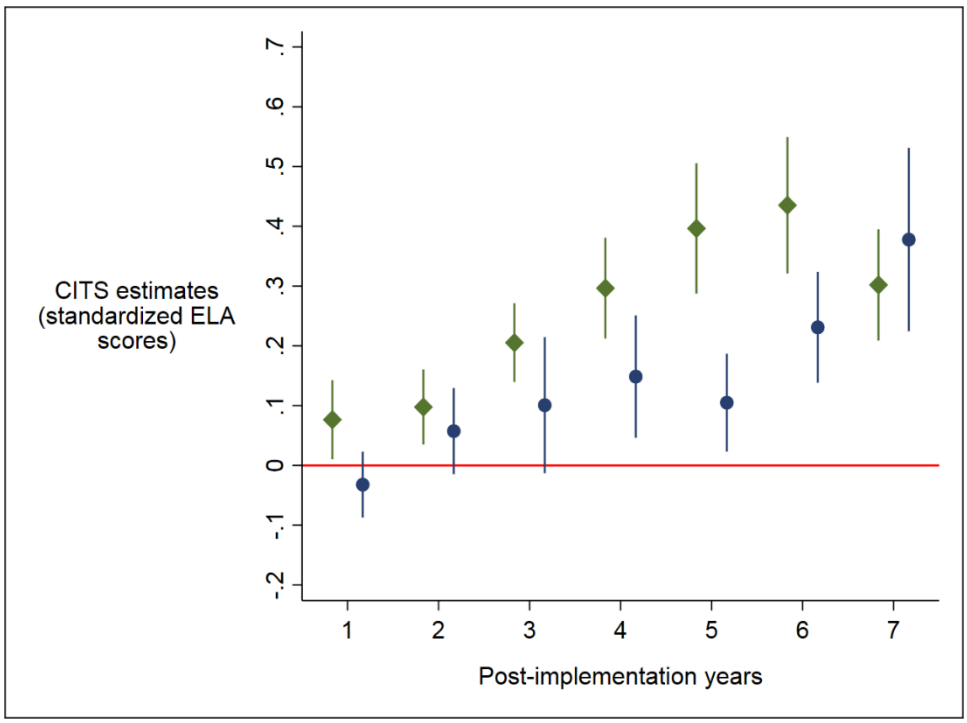
+ $p < 0.10$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Similar to the subgroup analyses for the school percentage of EL students, we present results separately for subsamples of schools classified as having a low percentage of students eligible for FRPL and a high percentage of students eligible for FRPL. Treatment schools that had higher than the median percentage of students eligible for FRPL were classified as *high % FRPL*, and remaining treatment schools were classified as *low % FRPL*. We then estimated the same CITS specification, described earlier, using the subsample of treatment schools in each of the two categories and their corresponding matched comparison schools.

Figure 4 and Table 5 present estimates from the CITS model and corresponding confidence intervals for these subgroup analyses. In treatment schools with a low percentage of students eligible for FRPL, TCRWP implementation was associated with small but statistically significant differences in test scores (0.08 and 0.1 standard deviations), relative to the matched comparison schools in Years 1 and 2. TCRWP implementation was associated with significantly improved test scores among treatment schools, relative to the comparison schools from Year 4 onward, with effects exceeding 0.25 standard deviations in each year. For treatment schools that served a high percentage of students eligible for FRPL, there was no association between TCRWP implementation and improved ELA scores in the first 2 years after

implementation. Beginning the 3rd year, TCRWP is associated with increased ELA scores, with largest effects in Years 6 and 7 (0.23 and 0.38 standard deviations).

Figure 4. Point Estimates and Confidence Intervals of Effect of TCRWP on Standardized Scores, by Treatment School Percentage of Students Eligible for FRPL



Note. Estimates from CITS model for the subsample of schools with low % students eligible for FRPL (green) and high % students eligible for FRPL, respectively (navy). For each cohort, treatment schools that had higher than the median percentage of students eligible for FRPL in the year prior to TCRWP implementation were classified as *high % FRPL*, and remaining treatment schools were classified as *low % FRPL*. Data points represent coefficients from the CITS model, and the vertical bars represent the corresponding 95% confidence intervals. The horizontal red line at zero indicates no relationship between TCRWP implementation and state ELA scores; points that have confidence intervals that do not cross the red line indicate a statistically significant relationship. As previously noted, all schools in the sample have test score data for at least 1 year after implementation. However, the results for later years are based on fewer schools.

Table 5. Comparison of Estimates of Effect of TCRWP on Standardized Scores, by Treatment School Percentage of Students Eligible for FRPL

Post-Implementation Year	Low % FRPL	High % FRPL
1	0.076*	-0.032
	(0.034)	(0.028)
2	0.098**	0.057
	(0.032)	(0.037)
3	0.205***	0.101+
	(0.034)	(0.058)
4	0.296***	0.149**
	(0.043)	(0.052)
5	0.396***	0.105*
	(0.056)	(0.042)
6	0.435***	0.231***
	(0.058)	(0.047)
7	0.302***	0.378***
	(0.047)	(0.078)
N (school-year observations)	1,052	1,086

Note. Robust standard errors, clustered at school level, are in parentheses. All regressions include controls for % White students, % female students, % students who are ELs, % students eligible for FRPL, and % students eligible for special education services, as well as school and year fixed effects. For each cohort, treatment schools that had higher than the median percentage of students eligible for FRPL in the year prior to TCRWP implementation were classified as *high % FRPL*, and remaining treatment schools were classified as *low % FRPL*.

+ $p < 0.10$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

We also investigated whether the effects differ by grade level (Table 6). We found positive, statistically significant differences in Grade 3 test scores between treatment schools and matched comparison schools beginning in the 1st year of implementation, with the differences becoming larger over time, particularly in the 4th year and beyond. In Grades 4 and 5, TCRWP was not associated with higher test scores in the first few years of implementation, but positive differences appear 3 years after implementation for Grade 5 and 5 years after implementation for Grade 4. For all three grades, the positive effects are substantial—at least a quarter of a standard deviation—6 and 7 years after implementation.

Table 6. Comparison of Estimates of Effect of TCRWP on Standardized Scores, by Grade

Post-Implementation Year	Standardized Scores (Grade 3)	Standardized Scores (Grade 4)	Standardized Scores (Grade 5)
1	0.060* (0.026)	-0.079* (0.039)	-0.012 (0.026)
2	0.109*** (0.026)	-0.012 (0.046)	0.032 (0.031)
3	0.176*** (0.040)	0.028 (0.053)	0.093** (0.032)
4	0.252*** (0.040)	0.065 (0.042)	0.150*** (0.036)
5	0.255*** (0.054)	0.117*** (0.033)	0.180*** (0.047)
6	0.352*** (0.046)	0.258*** (0.073)	0.305*** (0.050)
7	0.403*** (0.075)	0.419*** (0.046)	0.348*** (0.068)
N (school-year observations)	2,138	2,138	2,138

Note. Robust standard errors, clustered at school level, are in parentheses. All regressions include controls for % White students, % female students, % students who are ELs, % students eligible for FRPL, and % students eligible for special education services, as well as school and year fixed effects.
+ $p < 0.10$. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Discussion

The statistical analyses presented in Table 3 reinforce the finding depicted graphically in Figure 1, that is, that upward trends in the reading achievement of treatment schools became evident starting in the 2nd year of TCRWP use and continued to rise. The careful matching of treatment and comparison schools and the establishment of baseline equivalence among the two groups of schools indicate that the differences in treatment and comparison schools' achievement are real and *most likely* reflect the introduction of the TCRWP approach (use of the *Units of Study* and receipt of TCRWP professional development).

The analysis of multiple years of student achievement data makes this study unique compared with many other studies of the impact of new literacy and mathematics programs for use in elementary grades (for examples, see Garet et al., 2008; Garet et al., 2016; and Gersten et al., 2010). Typically, in these other studies, actual program implementation was preceded by lengthy content-focused professional development designed to orient teachers to the new program and enhance their professional knowledge. In some of these other studies, content-focused coaching and additional large-group professional development sessions were also part of the delivery model. The intensity of the professional development and support attest to the challenges practitioners in most fields often experience as they integrate new

approaches into their existing professional knowledge (Fixsen et al., 2005; Joyce & Showers, 2002). Evaluations of literacy and mathematics programs have typically studied impact after teachers receive extensive professional develop and support but have implemented the new programs for only 1 year. Often, these studies have found that participating in the intervention's professional development and use of program resources had an initial positive effect on teachers' instructional practices or scores on a test of content-specific professional knowledge, but there was usually no accompanying effect on their students' achievement. Garet et al. (2016) summarized the results of three large, experimental studies of content-focused teacher professional development funded by the Institute of Education Sciences and found that, even though extensive professional development and support yielded improvements in treatment teachers' practices, these gains did not seem to transfer to increases in student achievement, as compared with achievement of students in comparison schools. A common element across the interventions examined in these studies was the total amount—or dosage—of support that teachers received and the extent to which the professional development adhered to a predetermined plan or scope and sequence. A large meta-analysis of studies of reading interventions also shows that positive changes to student achievement rarely appear after 1 year of implementation of a new program and frequently take multiple years to become evident (Basma & Savage, 2018).

In comparison to the interventions examined in these other studies, individual TCRWP teachers have more options and likely more choice in terms of the types and dosage of professional development they receive. Although schools that adopt the TCRWP approach determine the amount of site-based professional development that TCRWP staff developers will provide each year, teachers and school administrators can also access a range of different professional development opportunities to strengthen their understanding of the program and enhance teachers' instructional practices. Overall, however, the total dosage of professional development that individual TCRWP teachers might receive, on average, is likely less than the total dosage reported in other studies.[15]

The question, therefore, arises as to what *other* factors, besides professional development, would likely contribute to the apparent success of the TCRWP approach. These factors may include the quality of the professional development; specific aspects of the curriculum or the approach; or the agency teachers develop as they choose among in-person, virtual, or print-based resources provided by TCRWP. Indeed, the very diversity of the opportunities and resources available to TCRWP schools and teachers may contribute to the success of the TCRWP approach. Teachers, school-based coaches, and administrators can select resources that they deem to have potential for meeting their immediate needs, such as implementing components of the model or differentiating instruction for diverse populations. Having these choices and being able to select the professional development and resources that they think will be most useful in their unique setting, might give teachers a sense of agency over their own learning and increase their personal investment in mastering the approach.

[15] For this study, we did not collect documentation of individual teachers' professional development dosage other than through self-report survey questions.

Another possible factor is the extent to which TCRWP encourages teachers to collaborate as they seek to understand not just the instructional practices but also the nuances of a new approach and the ways in which they can build on their existing professional knowledge as they master the new curriculum (Coburn, 2005; Goddard et al., 2000).[16] In the implementation portion of the study, participants in teacher focus groups spoke about the extent to which they worked with each other to interpret the *Units of Study* and implement TCRWP instruction in their local context. It is possible that positive effects are attributable, at least in part, to the high level of teacher collaboration fostered by TCRWP.

The emphasis on *student collaboration* in the TCRWP model is another possible factor. Time for student collaboration is built into the TCRWP model, from the Turn-and-Talk routines during the minilesson to multiple peer interactions throughout the literacy block. According to the Center for Teaching Innovation (2020), students who collaborate teach each other, model skills for one another, correct each other's misconceptions, and practice emerging skills together. Students learn to assess both themselves and their peers and to give each other feedback on their learning. These collaborative interactions, coupled with students' abilities to select the books they read and the topics of their writing, have the potential to increase their agency and self-efficacy—that is, their sense that they can play an active role in their own learning and achievement (Shunk & Bursuck, 2016).

Limitations

This study examined differences in test scores between TCRWP-adopting schools and matched comparison schools and effects of TCRWP use from 1 through 7 years after its initial implementation. The study employed a retrospective CITS design. A CITS is one of the strongest quasi-experimental research designs to use when a randomized controlled trial (RCT) is not possible (Shadish et al., 2002), but only an RCT trial allows for determination of the causal impact of a program. Studies that cannot randomize participants to treatment and control conditions may suffer from selection issues in that individuals (or schools) that are more likely to succeed can self-select into the program. Schools and districts must purchase TCRWP resources and services and can choose whether and when to implement the program. Such decision-making prerogative can lead to a selection issue when estimating the impact of TCRWP.

In addition, the treatment sample was a nonrandom, retrospectively selected subset of TCRWP schools that began implementing the program at some point between the 2012–13 and 2018–19 school years *and was still implementing it* in 2018–19.[17] For example, if a school that implemented the program in 2013–14 did not find the program effective after the first few years and stopped implementing it prior to 2018–19, that school would have been excluded from the treatment sample. Because our study does not

[16] Coburn and her colleagues did extensive research on how teachers collaborated among themselves and with their administrators to understand and implement the commercial programs and procedures required by the federal Reading First program; Coburn's term for this process is *sensemaking*. Goddard et al. write of the process of building *collective self-efficacy* as teachers analyze each other's successful teaching experiences and the barriers they have overcome in working toward mastery.

[17] Our analyses relied on publicly available aggregate data, and no data were available to contextualize why treatment schools continued or discontinued implementation.

include schools that may have discontinued use of the program because they were finding it ineffective, our results may overstate the effect of TCRWP on test scores.[18]

Moreover, as mentioned in notes for relevant tables and figures, the magnitude of effects from the CITS analysis for later years should be interpreted with particular caution because they are based on a small number of schools. Most notably, the effect for 7 years post implementation is identified from only seven schools.

On the other hand, TCRWP is a widely adopted program across NYC and Georgia schools. Therefore, there might be treatment spillovers, meaning that the matched comparison schools could have been exposed to some elements of the treatment. For example, teachers in NYC schools can access and participate in professional development offered by TCRWP on their own, even if the TCRWP approach is not adopted by their schools. In the presence of spillovers, our estimates might be biased downward.

Thus, for multiple reasons, limitations of the study could bias the results in either direction. This does not, of course, mean that the sources of bias perfectly counteract one another—one source could be stronger than another.

Overall, however, our results indicate that TCRWP has a positive effect on ELA achievement starting in the 2nd year of implementation and that, for schools that continue with the program, the effects grow larger over time.

Conclusion

Overall, results indicate that TCRWP implementation is associated with improvements in ELA achievement starting in the second year of implementation, and in schools that opt to continue with the approach long term, the magnitude of the effects grow larger over time. However, further research is needed to explore the factors that help or hinder implementation beyond the first two years so that we understand why the effects grow larger over time.

Additional research conducted prospectively rather than retrospectively would allow researchers to follow implementation in real time as implementation progresses. Studying why schools do and do not continue to use the *Units of Study* or choose to continue to receive professional development from TCRWP staff over time would contribute to the understanding of the program's effects and of schools' uptake of TCRWP methodology.

[18] To test if the treatment schools differed on observable characteristics from TCRWP schools not included in the treatment sample, we compared the mean and spread of test scores and demographics between the two groups. We obtained information on implementation years for all schools that affiliated with TCRWP. We then compared test scores and student demographic characteristics between treatment schools and nontreatment TCRWP schools using data from the 2010–11 school year for NYC and 2015–16 school year for Georgia, the year before any of the treatment schools started implementing the program. We also compared test scores and student demographic characteristics between the two groups in each implementation cohort, using data from 1 year prior to implementation for each cohort. The two groups differed on characteristics such as test scores, percentage of students eligible for FRPL, and percentage of students eligible for special education services, suggesting that results from the CITS analysis may indeed be biased, although the potential direction of the bias is unclear.

References

Basma, B., & Savage, R. (2018). Teacher professional development and student literacy growth: A systematic review and meta-analysis. *Educational Psychology Review 30*, 457–481.

Belfi, B., Haelermans, C., & Fraine, B. D. (2016). The long-term differential achievement effects of school socioeconomic composition in primary education: A propensity score matching approach. *British Journal of Educational Psychology, 86*(4), 501–525. https://doi.org/10.1111/bjep.12120

Center for Teaching Innovation. (2020). *Collaborative learning*. https://teaching.cornell.edu/teaching-resources/engaging-students/collaborative-learning#:~:text=Collaborative%20Learning-,Engaging%20Students,or%20find%20solutions%20to%20problems

Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy, 19*(3), 476–509. https://doi.org/10.1177/0895904805276143

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation research: A synthesis of the literature.* University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pdf/20084030.pdf

Garet, M. S., Heppen, J., Walters, K., Smith, T., & Yang, R. (2016). *Does content-focused teacher professional development work? Findings from three Institute of Education Sciences studies* (NCEE Evaluation Brief). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/20174010/pdf/20174010.pdf

Gersten R., Dimino, J., Jayanthi, M., Kim, J. S., & Santoro, L. E. (2010). Teacher Study Group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal, 47*(3), 694–739.

Goddard, Y., Goddard, R., & Kim, M. (2015). School instructional climate and student achievement: An examination of group norms for differentiated instruction. *American Journal of Education, 122*(1), 111–131. https://doi.org/10.1086/683293

Goddard, R., Hoy, W., & Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, *37*(2), 479-507.

Hallberg, K., Williams, R., Swanland, A., & Eno, J. (2018). Short comparative interrupted time series using aggregate school level data in education research. *Educational Researcher, 47*(5), 295–306.

Jacob, R., Goddard, R., & Kim, E. S. (2014). Assessing the use of aggregate data in the evaluation of school-based interventions: Implications for evaluation research and state policy regarding public use data. *Educational Evaluation and Policy Analysis, 36*(1), 44–66.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Association for Supervision and Curriculum Development.

Kanno, Y., & Cromley, J. G. (2015). English language learners' pathways to four-year colleges. *Teachers College Record*, 117(12), 1–44.

Ortiz, A. A., & Robertson, P. M. (2018). Preparing teachers to serve English learners with language- and/or literacy-related difficulties and disabilities. *Teacher Education and Special Education, 4*(3), 176–187. https://doi.org/10.1177/0888406418757035

Raudenbush, S. W. (2004). Schooling, statistics, and poverty: Can we measure school improvement? William H. Angoff Memorial Lecture Series, Educational Testing Service. https://files.eric.ed.gov/fulltext/ED486417.pdf

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Houghton Mifflin.

Shunk, S. H., & Bursuck, W. D. (2016). Self-efficacy, agency, and volition: Student beliefs and reading motivation. In Peter Afflerbach (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 54–67). Routledge.

Acknowledgements

The authors of this Research Brief thank our fellow AIR research team members: Angelica Herrera, Emily Baumann, Elaine Liebesman, Mary Bridget Burns, and Sarah Mae Olivar. We also thank Candace Hester and Feng Liu, who provided guidance and conducted reviews of the data analysis and reports at various points throughout the course of the study, and Deborah Holtzman and Michael Garet, whose thoughtful reviews ensured the quality of the work and of the reporting.

AMERICAN INSTITUTES FOR RESEARCH®

1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3239
202.403.5000

About the American Institutes for Research

Established in 1946, with headquarters in Arlington, VA, the American Institutes for Research (AIR) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance, both domestically and internationally, in the areas of education, health, and the workforce. **For more information, visit www.air.org.**

13132_12/20