IES Institute of Education Sciences

APPENDIXES
September 2021

Regional Educational Laboratory Midwest
At American Institutes for Research

Using Enhanced Coaching of Teachers to Improve Reading Achievement in Grades PreK–2 in Chicago Public Schools

See https://go.usa.gov/xFzBr for the full report.

Appendix A. About the study

The P–2 Balanced Literacy Initiative is a multiyear initiative that began in the 2016/17 school year to improve literacy instruction for students in grades preK–2 in Chicago Public Schools. The initiative involves collaboration among the district's Department of Literacy, the Office of Early Childhood Education, the Office of Diverse Learner Supports and Services, the Office of Language and Cultural Education, and the Department of Student Assessment. Although the balanced literacy approach is not new to teachers, the goal of the initiative is to spread best practices to more classrooms through professional development and coaching. The initiative offers sustained professional development to teachers, with a focus on differentiated, small group instruction and independent reading.

Balanced literacy approach

The term *balanced literacy* is often used to refer to literacy programs that involve mixed environments with both phonics and whole language instruction (Snow, 2020). Many districts across the nation, including Chicago Public Schools, currently use a balanced literacy approach. Opponents of balanced literacy argue that the approach lacks explicit, systematic phonics instruction (Castles et al., 2018). Explicit, systematic phonics instruction involves direct instruction of phonics components with opportunities to apply skills in decodable text formats, with a well-specified scope and a sequence that builds on skills in order. But balanced literacy is not a well-defined or specific curriculum, so approaches under its umbrella vary in their emphasis and enactment of phonics instruction and other foundational skill building (Snow, 2020). Literacy instruction that is most effective employs evidence-based practices described in the *What Works Clearinghouse Practice Guide* (Foorman et al., 2016) and related research (National Reading Panel, 2000; Shanahan et al., 2010).

Exposing children to systematic phonics instruction has significant positive impacts on children's ability to read (National Reading Panel, 2000). A high-quality phonics instruction approach incorporates the systematic and explicit introduction of a sequential set of phonics elements (National Reading Panel, 2000). However, the ability to decode words is necessary but not sufficient for reading comprehension. To succeed as readers, students need more than phonics instruction. They need both foundational reading skills and reading comprehension skills (Foorman et al., 2016). Comprehension for a child who can decode depends on several factors, such as word-level skills, vocabulary knowledge, oral language skills, broad conceptual knowledge, knowledge and abilities, thinking and reasoning skills, and motivation, the last three of which can be explicitly taught and all of which can be supported through the right learning opportunities (Shanahan et al., 2010). The What Works Clearinghouse's most

recent guidance, based on the National Reading Panel report and research published after its release, recommends that phonics instruction be part of an integrated approach to reading instruction (Foorman et al., 2016). An integrated approach includes instruction in all five areas recognized in the National Reading Panel report: phonics, phonological awareness, fluency, vocabulary, and comprehension. The What Works Clearinghouse practice guide recommends instruction in oral language skills to prepare students to read and communicate, and it recommends opportunities to read connected text (Shanahan et al., 2010).

Chicago Public Schools' P–2 Balanced Literacy Initiative

Chicago Public Schools' P–2 Balanced Literacy Initiative is characterized by explicit skill- and meaning-based instruction with independent literature and language exploration. A typical reading block in grades K–2 includes 20–30 minutes on systematic instruction in word study and phonics (phonemic awareness, phonics, reading practice, spelling, high-frequency and sight word instruction, and oral language) to build decoding skills and oral vocabulary, followed by 30 minutes of interactive read aloud and independent reading, 30–60 minutes of independent practice, small group instruction, and individual conferences, and ending with a 5-minute whole group share. An additional writing block provides students the opportunity to develop writing skills through whole group practice, small group and individual conferences, and independent practice. The literacy block's elements include the five areas recommended by the National Reading Panel report (National Reading Panel, 2000). Teachers are trained to listen to students read, ask comprehension questions, and check for the application of reading strategies during small group and individual conferences. Teachers provide instruction on decoding skills, comprehension, vocabulary development, and word study. Teachers are also asked to document student progress. A self-assessment tool helps teachers track their implementation of these conferencing components. The What Works Clearinghouse recommends one-on-one and small group instruction to build reading accuracy, fluency, and comprehension and to support struggling readers (Foorman et al., 2016).

The P–2 Balanced Literacy Initiative professional development for all schools in the initiative consists of professional development sessions for teachers and lead administrators, cross-network collaboration,[1] self-assessments of practice, and district assessments. Demonstration classrooms serve as models, and demonstration classroom teachers serve leadership roles such as co-facilitating professional development sessions and modeling components of balanced literacy through classroom visits.

In a set of priority schools identified in 2018/19, instructional support coaches provided additional school-based support to teachers through more intensive coaching at least twice a month, and lead administrators completed paired observations with coaches three times a year. Teachers in priority schools also received library infusion sets—collections of interdisciplinary books tailored to the cultural, linguistic, and socioeconomic diversity of the district—at the end of the 2018/19 school year.

Each year since 2016/17, networks have provided three one-day P–2 Balanced Literacy Initiative professional development sessions on focus topics chosen through a needs assessment and stakeholder engagement. In 2016/17 and 2017/18 the focus areas were classroom environment, independent literacy tasks, and guided reading (for K–2 teachers) and small group instruction (for preK teachers). In 2018/19 the three professional development sessions focused on independent reading and conferring. Teachers were trained on strategies for establishing a community of readers, engaging students with culturally relevant texts, and conducting independent reading minilessons. Teachers were guided on how to organize and evaluate classroom libraries, schedule and implement daily independent reading, assess children's response to independent reading, and set end-of-year goals for independent reading.

[1] Chicago Public Schools' 13 geographic networks for elementary schools provide oversight and support to schools within their region.

The professional development offered to priority schools was consistent with recommendations by Darling-Hammond et al. (2017), who reviewed 35 methodologically rigorous studies of effective professional development. In that review the features of effective professional development that changed teacher practices and student learning outcomes included sustained duration, coaching and expert support, feedback and reflection, and collaboration. Those practices can encourage community building that can induce changes beyond the classroom to entire grade levels, schools, or districts. Research has identified duration and intensity as two important elements of successful professional development (Garet et al., 2001; Justice & McGinty, 2012). Coaching and expert feedback help teachers reflect on and make changes to their practice (Darling-Hammond et al., 2017), and can contribute to the duration and intensity of professional development by increasing the time that teachers engage with experts on professional learning and focus on implementing practice. Other research on language- and literacy-focused professional development suggests that the number of components of professional development is strongly associated with instructional quality and interactions in the classroom (Markussen-Brown et al., 2017). The professional development offered to the P–2 Balanced Literacy Initiative's priority schools involves greater duration and intensity through coaching and additional components than does the professional development offered to the initiative's nonpriority schools.

References

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development.* Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/Effective_Teacher_Professional_Development_REPORT.pdf.

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE No. 2016-4008). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved July 2, 2021, from https://ies.ed.gov/ncee/wwc/PracticeGuide/21.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945. https://doi.org/10.3102/00028312038004915.

Justice, L. M., & McGinty, A. S. (2012). Early literacy intensity and its relation to child outcomes. In C. Howes, B. Hamre, & R. Pianta (Eds.), *Effective early childhood professional development* (pp. 89–110). Brookes Publishing.

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, *38*(1), 97–115. https://doi.org/10.1016/j.ecresq.2016.07.002.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication no. 00-4769). National Institute of Child Health & Development. https://www1.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE No. 2010-4038). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved July 2, 2021, from https://ies.ed.gov/ncee/wwc/PracticeGuide/14.

Snow, P. (2020). Balanced literacy or systematic reading instruction? *Perspectives on Language and Literacy, Winter*, 35–39. http://www.onlinedigeditions.com/publication/?i=655062&article_id=3634779&view=articleBrowser&ver=html5.

Appendix B. Methods

This appendix provides information about the data sources, data cleaning and preparation, missing data, and the methods used to conduct the analyses. Information also is provided about interpreting the results.

Data sources

The study team compiled data from two sources to address research questions 1 and 2.

Data for addressing research question 1. For research question 1 the study team obtained students' administrative data (school enrollment and demographic characteristics) and assessment data (proficiency ratings on the Text Reading and Comprehension [TRC] assessment and scale scores on the Measures of Academic Progress for Primary Grades [MAP] reading assessment) through a data-sharing agreement with Chicago Public Schools. Data on the characteristics of the sample of schools included in the study were compiled from publicly available data on the Illinois State Board of Education's state report card website (https://www.isbe.net/reportcard) and publicly available data from Chicago Public Schools Accountability Reports (https://www.cps.edu/about/district-data/metrics/accountability-reports; table B1).

Table B1. Data used to answer research questions

Data element	Years of data used in analysis	Source	Research questions
Student covariates			
Gender (1 = female, 0 = male)	2018/19	Chicago Public Schools student administrative records data	1
Grade (1 = grade 1, 0 = kindergarten)			
Eligible for the National School Lunch Program (1 = yes, 0 = no)			
Receives special education services (1 = yes, 0 = no)			
English learner student (1 = yes, 0 = no)			
American Indian/Alaska Native (1 = yes, 0 = no)			
Asian (1 = yes, 0 = no)			
Black (1 = yes, 0 = no)			
Hispanic (1 = yes, 0 = no)			
Pacific Islander (1 = yes, 0 = no)			
White (1 = yes, 0 = no)			
More than one race/ethnicity (1 = yes, 0 = no)			
Proportion of days enrolled in Chicago Public Schools that the student was present (continuous, grand mean centered)			
Student program enrollment			
School identification number	2018/19	Chicago Public Schools student administrative records data	1
Student assessment data			
Kindergarten and grade 1: Proficiency on the Text Reading and Comprehension assessment in fall 2018 and spring 2019 (1 = Proficient or Above Proficient, 0 = Far Below Proficient or Below Proficient)	2018/19	Chicago Public Schools student assessment records data	1
Grade 2: Scale scores on the Measures of Academic Progress for Primary Grades assessment (continuous, standardized; mean = 0, standard deviation = 1)			
Teacher program enrollment			
School identification number	2018/19	Chicago Public Schools teacher administrative data	2

Data element	Years of data used in analysis	Source	Research questions
School status			
School priority status (1 = priority, 0 = nonpriority)	2018/19	Chicago Public Schools implementation data, P–2 Balanced Literacy Initiative 2018/19	1,2
School covariates			
Proportion of students who are female (continuous, grand mean centered)	Aggregated from student data for 2018/19	Chicago Public Schools student administrative records data	1
Proportion of students eligible for the National School Lunch Program (continuous, grand mean centered)			
Proportion of students who received special education services (continuous, grand mean centered)			
Proportion of students who are English learner students (continuous, grand mean centered)			
Proportion of students who are Black (continuous, grand mean centered)			
Proportion of students who are Hispanic (continuous, grand mean centered)			
Proportion of students who are White (continuous, grand mean centered)			
Proportion of students who are other race/ethnicity (continuous, grand mean centered)			
Proportion of teachers with a professional licensure certification (continuous, grand mean centered)	Aggregated from teacher data for 2018/19	Chicago Public Schools teacher administrative records data	
Average number of years of teaching experience in the district (continuous, grand mean centered)			
School offers prekindergarten (1 = yes, 0 = no)	2017/18	Illinois State Board of Education's state report cards	
Average daily attendance (continuous, grand mean centered)	Average of school years 2016/17 and 2017/18	Illinois State Board of Education's state report cards	
School Quality Rating Policy (Level 3 = 0, Level 2+ = 1, Level 2 = 2, Level 1+ =3, Level 1 = 4)[a]	Average of school years 2016/17 and 2017/18	Chicago Public Schools Accountability data	
Implementation data			
Total number of professional development sessions teacher attended (0–3)	2018/19	Chicago Public Schools implementation data, P–2 Balanced Literacy Initiative	2
Total number of professional development sessions administrator attended (0–3)			

a. Chicago Public Schools provides the School Quality Rating Policy to each school. Each school receives a School Quality Rating and an Accountability Status every year. The School Quality Rating Policy helps communicate to school stakeholders the academic success of individual schools and the district as a whole; provides a framework for school goal setting; and guides Chicago Board of Education decisionmaking related to school support and intervention. The rating for elementary schools is based on several indicators of success, including student test score performance, student academic growth, closing of achievement gaps, school culture and climate, attendance, and data quality. Level 1+ is the highest performance, Level 1 is high performance, Level 2+ is average performance, Level 2 is below average performance, and Level 3 is the lowest performance. Schools that receive a School Quality Rating Policy rating of Level 2+ are in good standing; schools that receive a Level 2 rating need provisional support (Chicago Public Schools, n.d.).

Source: Authors' compilation.

Students' proficiency ratings on the TRC assessment were used in kindergarten and grade 1, and scale scores on the MAP reading assessment were used in grade 2.

TRC assessment in kindergarten and grade 1. Chicago Public Schools requires that all schools administer a formative assessment three times per year, and it allows schools to choose from a variety of formative assessment providers. The TRC assessment is one of the formative assessments that schools can use. Classroom teachers administer the assessment in the first month of school (between September and early October), at midyear (between December and January), and at the end of the school year (between May and early June). Teachers can administer the assessment more than three times each year to track student progress as needed. The assessment is an early reading formative assessment designed to measure reading accuracy (ability to read without errors), self-correction (ability to recognize and correct errors without prompting), instructional reading level (text level at which the student should be instructed), and reading comprehension (ability to read for understanding and to answer both literacy and inferential text-based questions) for students in grades K–2 (Chicago Public Schools, 2019). To complete the assessment, teachers take running notes about mistakes and fluency while a student is reading and then assess comprehension using leveled text. The test takes 30 minutes per student to administer. For English learner students whose primary language of instruction is Spanish, teachers can administer the TRC in Spanish to assess Spanish reading development. Data from the technical manual suggest high internal consistency, as measured by marginal reliability and fair to good interrater reliability (Amplify Education, 2019). Results from the assessment are reported using the following categories: Far Below Proficient, Below Proficient, Proficient, and Above Proficient. A rating of Proficient indicates that a student is performing at grade level. For the purposes of addressing research question 1, the categories were collapsed into a dichotomous variable: 0 is Far Below Proficient or Below Proficient and 1 is Proficient or Above Proficient.

MAP reading assessment in grade 2. All schools were required to administer the MAP reading assessment, developed by the Northwest Evaluation Association, to grade 2 students in spring 2018/19 (between May and early June). Schools could administer a different assessment in fall and winter, or they could also administer this formative assessment in the first month of school (between September and early October), and at midyear (between December and January). Because the district requires all schools to administer this assessment to grade 2 students at the end of the school year, the study team used MAP data only for grade 2 students. The valid range for MAP scores is between 100 and 350.

Data for addressing research question 2. For research question 2 the study team obtained implementation data on the P–2 Balanced Literacy Initiative for the 2018/19 school year from Chicago Public Schools, including attendance records for professional development sessions for administrators in initiative schools and teachers eligible to participate in the initiative.

Data cleaning and preparation

TRC assessment. The TRC data included student proficiency ratings for the three benchmark assessment periods, beginning, middle, and end of year for 2018/19. Data were provided at the test level. To prepare the data for analysis, the study team did the following:

- Replaced missing proficiency ratings for students using the corresponding proficiency level for the student's TRC level score and grade, based on the proficiency benchmark guide (Amplify Education, Inc., 2014).
- Because some students had more than one assessment proficiency rating for a benchmark assessment period (beginning and end of year), dropped all but the first proficiency rating for each benchmark period to avoid using a beginning-of-year rating that might have been collected after the start of the intervention.
- Reshaped the test-level data from long to wide so that each student had only one row in the data.

MAP assessment data. The data received from Chicago Public Schools included student scale scores for all MAP assessments for the 2018/19 school year, provided at the test level. To prepare the data for analysis, the study team did the following:

- Dropped scores that were not for reading assessments.
- Dropped the cases that were not in grade 2.
- Reshaped the test-level data from long to wide so that each student had only one row in the data.

Chicago Public Schools administrative records. Chicago Public Schools provided administrative records for students enrolled in grades K–2 in a P–2 Balanced Literacy Initiative school, including information on students' demographic characteristics and school enrollment. To prepare these data for analysis, the study team did the following:

- Created a measure from the beginning-of-year and end-of-year flags for whether the student was ever eligible for the National School Lunch Program in 2018/19.
- Created a second school-level dataset that described the school composition by summarizing the frequency or mean for each student demographic variable, including the proportion of students within the school who were female, ever eligible for the National School Lunch Program, ever received special education services, English learner students, and Black students, Hispanic students, White students, or students of other races/ethnicities.

Chicago Public Schools also provided administrative records for teachers of students in grades K–2 in a P–2 Balanced Literacy Initiative school, including information on teachers' certification and experience in 2018/19. To prepare these data for analysis, the study team did the following:

- Created a measure of whether the teacher was certified from the teacher accomplishment variable using the category that indicated whether the teacher had a professional educator license.
- Created a school-level dataset that described teacher composition by summarizing the mean years in Chicago Public Schools and the proportion of teachers who were certified.

Chicago Public Schools accountability reports. The study team downloaded the Chicago Public Schools accountability reports for the 2016/17–2017/18 school years, which included school-level demographic and academic data. To prepare these data for analysis the study team created an average rating for the School Quality Rating Policy by averaging the scores for the 2016/17 and 2017/18 school years.

Illinois State Board of Education. The study team downloaded the Illinois School Report Card data from the Illinois State Board of Education for the 2016/17–2017/18 school years, which included school-level demographic and academic data. To prepare these data for analysis, the study team did the following:

- Generated an indicator for whether the school offered a preK program from the measure indicating the grade levels served at each school in 2017/18.
- Created an average daily attendance rate for each school by averaging the scores for the 2016/17 and 2017/18 school years.

Implementation data on the P–2 Balanced Literacy Initiative. The P–2 Balanced Literacy Initiative implementation data included records of whether an administrator or teacher attended each of the initiative's three professional development sessions in 2018/19 and 2019/20. To prepare the data for analysis, the study team did the following:

- Dropped any records containing data for the 2019/20 school year.
- Generated the total number of sessions that each teacher attended.

- Generated the total number of sessions that at least one administrator from the school attended.

After cleaning files from each source, the study team merged individual files to construct analysis files to address each research question.

Research question 1 analysis files:

- A student-level dataset, created by merging the student administrative records with the TRC and MAP assessment data.
- A school-level dataset, created by merging the school-level demographic summaries from the student administrative records and teacher administrative records with the state report card school data and merging in the priority school indicator from the P–2 Balanced Literacy Initiative implementation records.

Research question 2 analysis file:

- An implementation dataset, created by merging the school-level indicator for priority status with the professional development attendance data from the P–2 Balanced Literacy Initiative dataset.

Sample. The study included two analysis samples for research question 1: 8,642 kindergarten and grade 1 students in 85 schools and 5,882 grade 2 students in 115 schools.[2] The sample for research question 2 included 115 schools (26 priority schools and 89 nonpriority schools) and 964 teachers.

Missing data

Before conducting analyses, the study team assessed the amount of data missing for each variable.

Response rates: Research question 1 TRC analysis sample. Of the 8,642 kindergarten and grade 1 students in the analysis sample for the TRC assessment, 92.9 percent (8,032 students) had scores available on the TRC assessment (table B2). All students in this analysis sample had indicators of whether they were enrolled in a priority school during 2018/19. The response rate for the research question 1 TRC analysis sample was 92.9 percent.

Table B2. Kindergarten and grade 1 Text Reading and Comprehension assessment response rates

Type of data the study attempted to collect	Number of students from whom study attempted to collect data	Number of students with data	Percent of students with nonmissing data
Any kindergarten or grade 1 district administrative data from a school that administered the TRC assessment	8,642	8,642	100.0
Key variable: kindergarten or grade 1 spring TRC score	8,642	8,032	92.9
Key variable: whether enrolled in a priority school	8,642	8,642	100.0
Both key variables for addressing the research question	8,642	8,032	92.9

TRC is the Text Reading and Comprehension assessment.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools.

Covariate missing data. Missing data rates of covariate measures were small for the fall TRC assessment: 9.9 percent of the students in this analysis sample were missing fall TRC assessment data (table B3). Approximately 7.8 percent of the students were missing race/ethnicity data, 1.6 percent of the students were missing data regarding their eligibility for the National School Lunch Program, and 0.2 percent of the students were missing attendance data. No students in the analysis sample were missing data related to gender, whether they were

[2] Kindergarten and grade 1 students were combined because they were administered the same assessment. Grade 2 students were administered a different assessment and were therefore analyzed separately.

English learner students, or whether they received special education services. None of the school background characteristics data was missing for this analysis sample, with one exception. One school did not receive a rating for the School Quality Rating Policy; the study team imputed that indicator using a rating from the following 2019/20 school year (no ratings were available for the three previous years for that school).

Table B3. Number of valid cases and percentage missing for each variable for the Text Reading and Comprehension reading assessment analysis sample

Variable	Number of valid cases	Percent missing
Spring TRC reading score	8,032	7.1
Fall TRC reading score	7,788	9.9
Gender	8,642	0.0
Grade	8,642	0.0
Received special education services	8,642	0.0
Eligible for the National School Lunch Program	8,501	1.6
English learner student	8,642	0.0
Race/ethnicity	8,642	7.8
Member days for which students are present[a]	8,622	0.2
School proportion of students who are female	8,642	0.0
School proportion of students who are eligible for the National School Lunch Program	8,642	0.0
School proportion of students who received special education services	8,642	0.0
School proportion of students who are who are English learner students	8,642	0.0
School proportion of students who are Black	8,642	0.0
School proportion of students who are Hispanic	8,642	0.0
School proportion of students who are all other races/ethnicities	8,642	0.0
School proportion of students who are White	8,642	0.0
School proportion of teachers who are certified	8,642	0.0
School average of teachers' years in Chicago Public Schools	8,642	0.0
School offers prekindergarten	8,642	0.0
School average daily attendance	8,642	0.0
School Quality Rating Policy score	8,642	0.0

TRC is the Text Reading and Comprehension assessment.
a. Number of days for which students are enrolled in Chicago Public Schools.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools and publicly available 2016/17 and 2017/18 school year data from the Illinois State Board of Education and the Chicago Public Schools Accountability Reports.

The study team used multiple imputation to create 10 multiply imputed datasets. All available student and school data were used to impute the scores using a logistic regression model, and all 10 multiply imputed datasets were analyzed simultaneously. The regression coefficients and variance components presented are pooled estimates across the 10 datasets.

For comparison, the study team also conducted analyses using the subset of students with data on all variables. Both analyses found no relationship between priority status and reading proficiency rates.

Response rates: Research question 1 MAP reading assessment analysis sample. Of the 5,882 grade 2 students in the analysis sample for the MAP assessment, 88.5 percent (5,205 students) had scores available on the grade 2

spring assessment (table B4). All students in this analysis sample had indicators of whether they were enrolled in a priority school during 2018/19. The response rate for the research question 1 analysis sample was 88.5 percent.

Table B4. Response rates for grade 2 Measures of Academic Progress for Primary Grades reading assessment

Type of data the study attempted to compile	Number of students from whom study attempted to collect data	Number of students with data	Percent of students with nonmissing data
Any grade 2 district administrative data	5,882	5,882	100.0
Key variable: grade 2 spring MAP assessment score	5,882	5,205	88.5
Key variable: whether student was enrolled in a priority school	5,882	5,882	100.0
Both key variables for addressing the research question	**5,882**	**5,205**	**88.5**

MAP is Measures of Academic Progress for Primary Grades assessment.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools.

Nonresponse bias analysis: MAP grade 2 reading assessment scores. Because the response rate for research question 1 is close to 85 percent (88.5 percent), the study team conducted a nonresponse bias analysis (table B5). The covariates displayed in the table include only those that had a correlation of .20 or higher with the grade 2 MAP scores and that had less than 5 percent missing data. The fall grade 2 MAP reading assessment scores had more than 5 percent missing data and thus are not presented here, though they are strongly related to the spring MAP reading assessment scores (see the next subsection for a description of the missing data for all covariates). Results from these analyses and the small differences (0.06 standard deviation or smaller) in covariates between the original analysis sample and the analysis sample with data on all key variables suggest that the sample with spring MAP scores is not atypical of the original analysis sample.

Table B5. Nonresponse bias analysis of grade 2 Measures of Academic Progress for Primary Grades reading assessment scores

Characteristic	Mean for students with spring MAP scores	Mean (standard deviation) for original analysis sample	Difference in standard deviation units	Correlation with spring MAP scores
Receives special education services	12.0	14.1 (34.9)	0.06	−0.24
Eligible for the National School Lunch Program	92.1	92.7 (26.0)	0.02	−0.31
Number of students	**5,205**	**5,882**		

MAP is Measures of Academic Progress for Primary Grades.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools.

Covariate missing data: Research question 1 MAP analysis sample. A substantial percentage of students—54.9 percent of this analysis sample—were missing fall grade 2 MAP assessment scores (table B6). Fewer than 1 percent of the students were missing data about whether they were eligible for the National School Lunch Program, and 5.5 percent of the students were missing race/ethnicity data. None of the students was missing data on gender, attendance rates, or indicators for whether they were English learner students or received special education services. Similarly, none of the school background characteristics data was missing, with one exception. One school did not receive a rating for the School Quality Rating Policy in the 2016/17 and 2017/18 school years; the study team imputed that indicator using a rating from the 2019/20 school year (no ratings were available for the three previous years for that school).

Table B6. Number of valid cases and percentage missing for each variable for the Measures of Academic Progress reading assessment analysis sample

Variable	Number of valid cases	Percent missing
Spring MAP reading score	5,205	11.5
Fall MAP reading score	2,650	54.9
Gender	5,882	0.0
Received special education services	5,882	0.0
Eligible for the National School Lunch Program	5,825	1.0
English learner student	5,882	0.0
Race/ethnicity	5,588	5.5
Percent of member days for which student is present[a]	5,882	0.0
School proportion of students who are female	5,882	0.0
School proportion of students who are eligible for the National School Lunch Program	5,882	0.0
School proportion of students who received special education services	5,882	0.0
School proportion of students who are who are English learner students	5,882	0.0
School proportion of students who are Black	5,882	0.0
School proportion of students who are Hispanic	5,882	0.0
School proportion of students who are all other races/ethnicities	5,882	0.0
School proportion of students who are White	5,882	0.0
School proportion of teachers who are certified	5,882	0.0
School average of teachers' years in district	5,882	0.0
School offers prekindergarten	5,882	0.0
School average daily attendance	5,882	0.0
School Quality Rating Policy score	5,882	0.0

MAP is Measures of Academic Progress for Primary Grades assessment.
a. Number of days for which students are enrolled in Chicago Public Schools.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools and publicly available 2016/17 and 2017/18 school year data from the Illinois State Board of Education and the Chicago Public Schools Accountability Reports.

Because the data for the fall MAP reading assessment were missing at such high rates, the study team used multiple imputation to create 10 multiply imputed datasets. All available student and school data were used to impute the scores using a linear regression model, and all 10 multiply imputed datasets were analyzed simultaneously. The regression coefficients and variance components presented are pooled estimates across the 10 datasets.

For comparison, the study team also conducted analyses using the subset of students who had both beginning- and end-of-year scores (see table C4 in appendix C). The team also conducted a third set of analyses that did not adjust for fall MAP scores, to retain the subset of students with no fall scores in the analyses. All three analyses found no relationship between priority status and reading achievement.

Data analysis for research question 1

To answer research question 1, the study team used a series of two-level hierarchical regression models with inverse probability weighting to compare reading achievement of students in priority schools with that of students in nonpriority schools, after adjusting for the variability in student and school covariates. Because schools were not randomly assigned to receive the initiative and because students were not randomly assigned to schools, it is

possible that preexisting student and school differences confounded any observed differences in students' end-of-year reading achievement. For example, differences in student reading achievement at the end of the year might have resulted from the extra professional development that priority schools received or from students in priority schools performing lower at the start of the program and thus having more room to improve. To minimize preexisting differences, inverse probability weights[3] were generated and subsequently applied at the school level in the analyses conducted to compare reading achievement for students in priority and nonpriority schools.

Generating inverse probability weights from propensity scores. The study team first generated propensity scores by estimating the conditional probability that a school was a priority school given a set of school-level predictors. These predictors included school characteristics from the Illinois State Board of Education's state report cards (average daily attendance and whether the school offered preK), Chicago Public Schools accountability data (School Quality Rating Policy rating), and all available student characteristics aggregated to the school level (such as proportion female, proportion eligible for the National School Lunch Program, and proportion receiving special education services). The study team used a logistic regression model to generate the school propensity scores, $\hat{p}$, which were converted to inverse probability weights: $1/\hat{p}$ for priority schools and $1/(1 - \hat{p})$ for nonpriority schools.

Assessing baseline equivalence. Before applying the weights in the analysis models for addressing the research question, the study team assessed the baseline balance on all potential confounders by calculating the absolute standardized mean difference for each covariate in the unweighted and the weighted samples. Absolute standardized mean difference values of less than .20 indicate that balance has been achieved (Abdia et al., 2017; Austin, 2011; Burgette et al., 2015; McCaffrey et al., 2013; McCaffrey et al., 2004). However, as in recommendations from previous research, the covariates used in generating the propensity score weights also were included in the outcome analyses to estimate the difference in end-of-year reading scores (Austin, 2011).

For all covariates the absolute weighted standardized mean difference was less than 0.20 (table B7).

[3] Rather than matching schools on their propensity scores, which could result in unmatched schools, the study team used an inverse probability weighting approach that allowed all schools in the sample to be retained (Austin, 2011; Lunceford & Davidian, 2004).

Table B7. Establishing covariate balance in unweighted and weighted samples

Characteristic	Nonpriority schools		Priority schools		ASMD balance	
	Mean	Standard deviation	Mean	Standard deviation	Unweighted	Weighted
School characteristic						
PreK available (0 = no, 1 = yes)	92.13	27.07	96.15	19.61	−0.16	−0.14
Average SQRP rating[a,b]	2.49	0.96	2.52	1.02	−0.01	−0.09
Average percentage daily attendance[a]	95.03	1.10	94.75	1.13	0.25	0.12
Average percentage female[c]	48.61	4.52	49.27	4.44	−0.15	0.01
Average percentage eligible for the National School Lunch Program[c]	92.74	15.26	94.91	6.23	−0.16	−0.18
Average percentage in special education[c]	14.50	8.05	16.10	5.87	−0.21	−0.13
Average percentage English learner students[c]	23.92	25.2	22.82	25.2	0.04	−0.04
Average percentage Black[c]	47.17	38.54	51.60	37.53	−0.12	−0.01
Average percentage Hispanic[c]	35.58	37.33	33.69	35.93	0.05	−0.03
Average percentage White[c]	4.52	9.96	2.96	5.30	0.17	0.16
Average percentage all other races/ethnicities[c]	3.21	9.16	3.36	8.87	−0.02	0.01
School average of teachers' years in Chicago Public Schools	15.82	4.41	15.72	4.58	0.02	0.01
Proportion of teachers certified	94.54	10.10	95.28	7.45	−0.08	−0.14
Number of schools	89		26			
Student characteristic						
Fall TRC score	0.52	0.50	0.54	0.50	−0.05	−0.07
Fall MAP score	167.12	15.97	166.00	15.03	0.07	0.07
Female	49.13	49.99	49.42	50.00	−0.01	< 0.01
Eligible for the National School Lunch Program	91.28	28.21	95.00	21.81	−0.14	−0.13
Received special education services	12.97	33.60	14.51	35.23	−0.05	0.00
English learner student	29.36	45.54	30.74	46.15	−0.03	−0.05
Black	33.47	47.19	38.67	48.71	−0.11	−0.08
Hispanic	49.80	50.00	44.54	49.71	0.11	0.08
White	6.08	23.90	4.26	20.21	0.08	0.07
Multiracial	10.65	30.85	12.53	33.11	−0.06	−0.05
Member days for which students are present[d]	94.42	6.94	94.37	6.57	0.01	0.01
Number of students	11,147		3,377			

ASMD is absolute standardized mean difference, which was calculated by dividing the mean difference by the pooled standard deviation equal to the square root of the average of the squared standard deviations from the priority and nonpriority schools. MAP is Measures of Academic Progress for Primary Grades assessment. SQRP is School Quality Rating Policy. TRC is Text Reading and Comprehension assessment.

a. Averaged across two school years, 2016/17 and 2017/18.

b. Recoded so that Level 3 is 0, Level 2+ is 1, Level 2 = 2, Level 1+ is 3, and Level 1 is 4.

c. Aggregated from student data.

d. Number of days for which students are enrolled in Chicago Public Schools.

Source: Authors' analysis of data provided by Chicago Public Schools and publicly available 2016/17 and 2017/18 school year data from the Illinois State Board of Education and the Chicago Public Schools Accountability Reports.

Estimating the difference in end-of-year reading achievement. After establishing baseline equivalence on the covariates, the study team formulated a series of two-level hierarchical regression models with inverse probability weights applied at the school level. Because different assessments were analyzed at kindergarten and grade 1 and at grade 2, the analysis procedures are described separately.

Outcomes analysis at kindergarten and grade 1. In a series of statistical models the study team regressed students' end-of-year proficiency on the TRC (0 is Far Below Proficient or Below Proficient and 1 is Proficient or Above Proficient) on an array of *Q* student covariates at level 1 and *S* school covariates at level 2 using two-level hierarchical logistic models. The level 2 models also included an indicator of a school's priority status (0 is nonpriority, 1 is priority). The logit associated with the indicator of a school's priority status was the primary coefficient of interest.

The initial model included only the array of *Q* student covariates and the indicator of a school's priority status. In subsequent models, the array of *S* school covariates was added. The final models included cross-level interactions to examine whether the predicted difference between the priority and nonpriority schools on the outcome variable varied by student characteristics and school-level interactions between priority status and school characteristics. The findings presented in the main report are based on the second model, which included the array of *Q* student covariates and *S* school covariates.

In the initial model the binary outcome, η_{ij}, for student *i* in school *j* was regressed on an array of *Q* student covariates, X_{qij}, and a binary indicator of schools' priority status, $Priority_j$ (0 is nonpriority, 1 is priority), using the following mixed model:

$$Logit(\Pr(\eta_{ij} = 1)) = \gamma_{00} + \gamma_{01} Priority_j + \sum_{q=1}^{Q} \gamma_{qij} X_{qij} + \sum_{q=1}^{Q} u_{qj} X_{qij} + u_{0j} .$$

At the school level (level 2), the model for predicting the level 1 intercept included the indicator of a school receiving the priority school resources, $Priority_j$ (0 is nonpriority, 1 is priority) as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Priority_j + u_{0j}$$

$$\beta_{qj} = \gamma_{q0} + u_{qj} .$$

The level 1 intercept, γ_{00}, was allowed to vary across schools, with the variability captured by the level 2 residual term, u_{0j}, and was interpreted as the logit of being in the reference group, 0 (0 is Far Below Proficient or Below Proficient), for a student in a nonpriority school that has a value of 0 on all student and school covariates. The coding and centering of the student and school covariates (see table B1) allow the intercept to be meaningfully interpreted in this way. The significance of the variability in the level 1 slopes for each of the *Q* student covariates was evaluated, and if found to be significant, a random parameter for that slope was added at level 2, u_{qj}. In deciding whether to include these random parameters, the study team examined the model fit using the Akaike information criterion, which is appropriate for binary outcomes for models with and without the random component, and selected the best fitting model (Raudenbush & Bryk, 2002). The coefficient associated with being a priority school, γ_{01}, was interpreted as the predicted change in the log of the odds of attaining proficiency or above on the TRC assessment for students in priority schools compared with students in nonpriority schools, partialing out the variability because of the *Q* student covariates.

In the second model an array of *S* school-level covariates, W_{sj}, was added, along with the student covariates and the $Priority_j$ indicator, to predict variability in the level 1 intercept. As in the first model, random parameters, u_{qj}, were included for all randomly varying level 1 slopes. The model took the following form:

$$Logit(\Pr(\eta_{ij} = 1)) = \gamma_{00} + \gamma_{01} Priority_j + \sum_{q=1}^{Q} \gamma_{qj} X_{qij} + \sum_{s=1}^{S} \gamma_{0s} W_{sj} + \sum_{q=1}^{Q} u_{qj} X_{qij} + u_{0j} .$$

Again, the coefficient associated with being a priority school, γ_{01}, was the primary focus. This coefficient was interpreted as the predicted change in the log of the odds of attaining proficiency or above on the TRC assessment for students in priority schools compared with students in nonpriority schools, conditional on the *Q* student covariates and the *S* school covariates.

In additional models the $Priority_j$ indicator was included in the models for the level 1 slopes that varied significantly across schools. This allowed the study team to examine whether the predicted difference between priority and nonpriority schools on the outcome variable varied by student characteristics. In addition, *N* interaction terms between priority status and the school characteristics that were significantly associated with the outcome were included in the model. The model was as follows:

$$Logit(\Pr(\eta_{ij} = 1)) = \gamma_{00} + \gamma_{01}Priority_j + \sum_{q=1}^{Q}\gamma_{qj}X_{qij} + \sum_{s=1}^{S}\gamma_{0s}W_{sj} + \gamma_{qq}Priority_j * \sum_{q=1}^{Q}X_{qij} + \gamma_{0N}Priority_j * \sum_{N=1}^{N}W_{sj} + \sum_{q=1}^{Q}u_{qj}X_{qij} + u_{0j}\ .$$

Represented by the γ_{0N} terms in the model, the magnitude and direction of these coefficients allowed the study team to examine whether the significant school covariates in the model moderated the differences between outcomes for students in priority and nonpriority schools. The γ_{qq} coefficients in this equation represent the cross-level interactions between the *Q* student characteristics at level 1 and the $Priority_j$ indicator at level 2. The magnitude and direction of these coefficients allowed the study team to explore whether the student covariates in the model moderated the differences between proficiency rates on the TRC assessment for students in priority and nonpriority schools.

When presenting the results, the study team transformed the logits estimated in these models to odds and graphed the predicted probabilities, π_{ij}, of attaining proficiency conditional on schools' receipt of the priority resources and the arrays of *Q* student and *S* school covariates using the following equation:

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} = \frac{odds_{(ij)}}{1+odds_{(ij)}}.$$

An odds ratio of 1 for the $Priority_j$ variable indicated no association between attaining proficiency and priority status, whereas an odds ratio greater than 1 indicated a greater probability of being proficient for students in the priority schools.

Outcomes analysis at grade 2. Similar to the approach adopted for kindergarten and grade 1, the study team formulated a series of statistical models in which students' standardized end-of-year scores on the MAP reading assessment (mean = 0, standard deviation = 1) were regressed on the indicator of schools' priority status (0 is nonpriority, 1 is priority), an array of *Q* student covariates at level 1, and an array of *S* school covariates at level 2. In these models a regression coefficient is interpreted as the predicted standard deviation change in MAP scores for every one-unit change in the predictor, conditional on the other variables in the model. The regression coefficient associated with schools' priority status was the primary coefficient of interest.

As in the analyses conducted for the TRC assessment, the study team formulated three primary models. The first model included the array of *Q* student covariates and the indicator of schools' priority status; the second model included the *Q* student covariates, S school covariates, and the indicator of schools' priority status; and the final model included cross-level interaction terms. The findings presented in the main report are based on the second model.

In the first model, MAP scores for student *i* in school *j*, Y_{ij}, were regressed on the *Q* student covariates at level 1 and the indicator of schools' priority status, $Priority_j$, at level 2:

$$Y_{ij} = \gamma_{00} + \gamma_{01}Priority_j + \sum_{q=1}^{Q}\gamma_{qj}X_{qij} + \sum_{q=1}^{Q}u_{qj}X_{qij} + +u_{0j} + r_{ij}.$$

The level 1 intercept, γ_{00}, was allowed to vary across schools, with the variability captured by the level 2 residual term, u_{0j}, and was interpreted as the predicted MAP score for an average student in a nonpriority school with average characteristics. As before, the significance of the variability in the level 1 slopes for each of the *Q* student covariates was evaluated, and a random parameter, u_{qj}, was added at level 2 if the model fit improved. The key regression coefficient in this model, γ_{01}, represents the predicted difference between the average MAP scores in the nonpriority schools and the priority schools, conditioned on the student covariates in the model.

In the subsequent models the array of *S* school covariates was added to the level 2 intercept equation as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{01}Priority_j + \sum_{q=1}^{Q}\gamma_{qj}X_{qij} + \sum_{s=1}^{S}\gamma_{0s}W_{sj} + \sum_{q=1}^{Q}u_{qj}X_{qij} + u_{0j} + r_{ij}.$$

In this model the regression coefficient associated with priority status, γ_{01}, represents the predicted difference between the average MAP scores in nonpriority schools and in priority schools, conditioned on the *Q* student and *S* school covariates in the model.

Following procedures used for the binary models for TRC outcome, the $Priority_j$ indicator was included in the models for the level 1 slopes that varied significantly across schools. This allowed the study team to examine whether the predicted difference between priority and nonpriority schools on the outcome variable varied by student characteristics. In addition, *N* interaction terms between priority status and the school characteristics that were significantly associated with the outcome were included in the model as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{01}Priority_j + \sum_{q=1}^{Q}\gamma_{qj}X_{qij} + \sum_{s=1}^{S}\gamma_{0s}W_{sj} + \gamma_{qq}Priority_j * \sum_{q=1}^{Q}X_{qij} + \gamma_{0N}Priority_j * \sum_{N=1}^{N}W_{sj} + \sum_{q=1}^{Q}u_{qj}X_{qij} + u_{0j} + r_{ij}\ .$$

The γ_{0N} terms in the model represent the magnitude and direction of interactions between schools' priority status and the significant school-level characteristics. The γ_{qq} coefficients in this equation represent the cross-level interactions between the *Q* student characteristics at level 1 and the $Priority_j$ indicator at level 2, allowing the study team to examine whether the student covariates in the model moderate the differences between average MAP reading assessment scores for students in priority and nonpriority schools.

Data analysis for research question 2

To answer research question 2, the study team calculated the number and percentage of teachers who participated in zero, one, two, or all three sessions of professional development within each school; calculated whether at least one principal or administrator participated in zero, one, two, or all three sessions of professional development; and calculated the percentage of priority schools that had at least 75 percent of teachers who participated in one or more sessions of professional development. The study team calculated descriptive statistics for teachers and schools separately by priority status to compare how participation varied between the two program types. The study team examined whether the number and percentage of teachers who participated in zero, one, two, or all three sessions of professional development varied by whether their school was a priority school using crosstabulations of level of professional development participation by priority status. Similarly, to compare differences in administrator attendance by program, the study team calculated the percentage of schools in which at least one administrator attended zero, one, two, or all three sessions of professional development for the priority and nonpriority schools. The study team calculated the frequencies and percentages of teachers in each school who participated in zero, one, or two or more sessions of professional development and compared the mean percentage of teachers at each level of participation for priority and nonpriority schools. The study team used *t*-tests and chi-square or Fisher's exact tests to test for statistically significant differences. The study reports meaningful group differences, defined as those that were 5 percentage points or greater. That threshold was determined in consultation with the stakeholder advisory group.

Interviews with district, network, and school leaders; instructional support coaches; and teachers

Interview sample. The study team asked district leaders to nominate two networks to participate in the interviews based on their perceptions of which networks had been successful at implementation. District and network leaders then identified two instructional support coaches in each network. District and network leaders and coaches helped the study team identify two priority schools in each network, and the study team reached out to principals and teachers in those schools. The study team conducted 30-minute semistructured interviews with seven district and network leaders, three instructional support coaches, one principal, and three teachers individually. So, the findings here are not representative of all educators' experiences in the district.

Interview protocols. The interviews included questions about professional development supports and the conditions for successful implementation of the initiative. See boxes B1, B2, and B3 for the interview protocols.

Box B1. Interview protocol for coaches

Hello, I am _____ with the Regional Educational Laboratory Midwest.

Thank you for deciding to participate in this interview. The purpose of this interview is to provide the district with feedback from coaches about the conditions they deem necessary for successful implementation of the P–2 Balanced Literacy Initiative. The information collected from the interviews will be reported back to the district, but no identifying information about your name or the schools you serve will be attached to the reports.

The interview will last approximately 30 minutes.

Your participation in this interview is completely voluntary, and any information you provide will be confidential. As a reminder, this conversation is being recorded, and after the transcription of our conversation, the digital file will be destroyed. A data file with the transcriptions of all interviews will be shared with CPS after removing all identifying information including your name and the schools you serve. All identifying information will be kept confidential. Do I have permission to record you? [*Note:* If the respondent wishes not to be recorded, take notes but do not proceed with recording. If the respondent consents to being recorded, please record the interview.] Do you have any questions before we begin?

1. Background
    a. How many years have you been working as a coach? How long did you teach prior to becoming a coach?
    b. What is your educational background in literacy instruction?
2. How would you describe the purpose of the P–2 Balanced Literacy Initiative?
3. In your opinion, what does successful implementation of the initiative look like in classrooms and schools?
    a. How can you tell if a teacher is successfully implementing the practices promoted through the professional development and coaching?
    b. How can you tell if a school as a whole is successfully implementing the practices promoted through the professional development and coaching?
4. What are the conditions necessary for the initiative to be implemented successfully?
    a. How do you know if a teacher is ready to implement the practices promoted by the initiative?
    b. How do you know if a grade band of teachers is ready to support implementation?
    c. How do you know if a principal is ready to support implementation?
    d. What materials and resources are needed to support successful implementation at the classroom level? Grade band level? School level?
    e. Do teachers or principals need a certain level of knowledge about content or pedagogy to successfully implement the initiative?

5. What conditions would make it easier to implement the initiative successfully?
6. Is there any additional information that you would like to share about your experience with being a coach for the P–2 Balanced Literacy Initiative?

This question concludes the interview. Thank you for your time. If you have any questions about this study, please feel free to contact the Principal Investigator for this study, Juliette Berg, PhD (jberg@air.org).

Box B2. Interview protocol for teachers

Hello, I am _____ with the Regional Educational Laboratory Midwest.

Thank you for deciding to participate in this interview. The purpose of this interview is to provide the district with feedback from teachers about the conditions they deem necessary for successful implementation of the P–2 Balanced Literacy Initiative. The information collected from the interviews will be reported back to the district, but no identifying information about your name or the schools you serve will be attached to the reports.

The interview will last approximately 30 minutes.

Your participation in this interview is completely voluntary, and any information you provide will be confidential. As a reminder, this conversation is being recorded, and after the transcription of our conversation, the digital file will be destroyed. A data file with the transcriptions of all interviews will be shared with CPS after removing all identifying information including your name and your school. All identifying information will be kept confidential. Do I have permission to record you? [*Note:* If the respondent wishes not to be recorded, take notes but do not proceed with recording. If the respondent consents to being recorded, please record the interview.] Do you have any questions before we begin?

1. Background
    a. How many years have you been teaching at your school? How long have you been a teacher?
2. What does successful implementation of the Balanced Literacy professional development and coaching mean to you?
3. What conditions support successful implementation? What do you need to implement this initiative successfully in your classroom?
    a. What types of supports have you received to help you implement what you learned through the Balanced Literacy professional development and coaching? What supports have been especially useful?
    b. How often have you met with an instructional coach? What have you found most useful about the coaching support?
    c. Have you gotten support from your principal to participate in the implementation and coaching and to implement what you have learned? What supports have been especially useful?
    d. What materials and resources have been useful to help you implement in your classroom?
4. What conditions would make it easier to implement the initiative successfully?
5. Is there any additional information that you would like to share about your experience with receiving professional development and coaching through the P–2 Balanced Literacy Initiative?

This question concludes the interview. Thank you for your time. If you have any questions about this study, please feel free to contact the Principal Investigator for this study, Juliette Berg, PhD (jberg@air.org).

Box B3. Interview protocol for district leaders

Hello, I am _____ with the Regional Educational Laboratory Midwest.

Thank you for deciding to participate in this interview. The purpose of this interview is to learn more about the history of the adoption of the P–2 Balanced Literacy Initiative in Chicago. The information collected from the interviews will be reported back to the district, but no identifying information about your name will be attached to reports that summarize what we learn from this interview.

The interview will last approximately 30 minutes.

Your participation in this interview is completely voluntary, and any information you provide will be confidential. As a reminder, this conversation is being recorded, and after the transcription of our conversation, the digital file will be destroyed. A data file with the transcriptions of all interviews will be shared with CPS after removing all identifying information. All identifying information will be kept confidential. Do I have permission to record you? [*Note:* If the respondent wishes not to be recorded, take notes but do not proceed with recording. If the respondent consents to being recorded, please record the interview.] Do you have any questions before we begin?

1. Background
    a. How many years have you been in your position at CPS?
2. How would you describe the purpose of the P–2 Balanced Literacy Initiative?
3. How did you and your colleagues make the decision to implement a Balanced Literacy professional approach?
    a. What sources did you draw on?
    b. Did you consider any other professional development models? If so, how did you make the decision?
    c. Has there been any consideration of implementing other literacy approaches since adopting the initiative?
4. What type of feedback have you received from school administrators about the initiative?

This question concludes the interview. Thank you for your time. If you have any questions about this study, please feel free to contact the Principal Investigator for this study, Juliette Berg, PhD (jberg@air.org).

Interview data analysis. The study team qualitatively analyzed interview transcripts by first identifying themes related to conditions necessary for successful implementation of the initiative. These themes included collaboration and engagement across roles, knowledge sharing among teachers, and coaching. Study team members identified and counted subthemes. The study team met to reconcile findings and determine final themes across individuals.

References

Abdia, Y., Kulasekera, K. B., Datta, S., Boakye, M., & Kong, M. (2017). Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study. *Biometrical Journal*, *59*(5), 967–985. https://doi.org/10.1002/bimj.201600094.

Amplify Education, Inc. (2014). *mCLASS: Reading 3D – Text Reading and Comprehension. Setting cut points for the Amplify Atlas book set.* Retrieved July 2, 2021 from https://s3.amazonaws.com/amplify-assets/pdf/mCLASS_Reading3D_TRC_CutPoints_A.pdf.

Amplify Education, Inc. (2019). *mCLASS Reading 3D Atlas Book Set technical manual,* 2nd ed. Retrieved July 2, 2021, from https://amplify.com/research/mclass-research.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786.

Burgette, L. F., McCaffrey, D. F., & Groffin, B. A. (2015). Propensity score estimation with boosted regression. In W. Pan & H. Bai (Eds.), *Propensity score analysis: Fundamentals and developments* (pp. 44–73). Guilford Press.

Chicago Public Schools. (n.d.). *School quality rating policy (SQRP) overview*. Retrieved July 2, 2021, from https://www.cps.edu/globalassets/cps-pages/about/district-data/metrics/school-quality-rating-policy-sqrp/sqrp_overview.pdf.

Chicago Public Schools. (2019). *District data: Racial/ethnic report—School year 2018–2019.* Retrieved July 2, 2021, from https://www.cps.edu/about/district-data/demographics/.

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effect: A comparative study. *Statistics in Medicine*, *23*(19), 2937–2960. https://www4.stat.ncsu.edu/~davidian/statinmed.pdf.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, *32*(19), 3388–3414. https://doi.org/10.1002/sim.5753.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, *9*(4), 403–425. https://doi.org/10.1037/1082-989X.9.4.403.

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

Appendix C. Supporting analyses

This appendix presents detailed findings for research questions 1 and 2. Tables C1–C4 show the findings for research question 1. Tables C5–C7 show the findings for research question 2.

Table C1. Predicting end-of-year Text Reading and Comprehension proficiency ratings and Measures of Academic Progress for Primary Grades reading assessment scores with school priority status and student and school covariates (research question 1), 2018/19

	TRC kindergarten and grade 1 (*n* = 7,971 students in 85 schools)			MAP grade 2 (*n* = 5,882 students in 115 schools)
Characteristic	Logit (standard error)	Odds ratio	.95 confidence interval	Standardized coefficient (standard error)
Student-level predictor				
Intercept	0.06 (0.3)	1.06	(0.58,1.93)	0.13 (0.11)
Fall assessment score	1.76 (0.08)***	5.80	(4.94,6.81)	0.64 (0.02)***
Grade (0 = kindergarten, 1 = grade 1)[a]	0.30 (0.12)*	1.35	(1.06,1.71)	—
Female	0.09 (0.06)	1.09	(0.97,1.23)	0.07 (0.03)*
Ever eligible for the National School Lunch Program	−0.53 (0.18)**	0.59	(0.41,0.85)	−0.11 (0.05)*
Received special education services	−1.52 (0.14)***	0.22	(0.16,0.29)	−0.38 (0.06)***
English learner student	−0.04 (0.13)	0.97	(0.74,1.25)	−0.10 (0.03)
White vs. Black	0.24 (0.2)	1.27	(0.85,1.90)	0.13 (0.06)*
American Indian/Alaska Native vs. Black	−0.50 (0.55)	0.60	(0.20,1.83)	0.32 (0.25)
Hispanic vs. Black	−0.18 (0.15)	0.84	(0.63,1.11)	0.08 (0.05)
Multiracial vs. Black	0.29 (0.32)	1.34	(0.71,2.53)	−0.01 (0.21)
Asian vs. Black	0.73 (0.36)*	2.07	(1.02,4.21)	−0.03 (0.06)
Pacific Islander/Hawaiian vs. Black	0.73 (0.54)	2.08	(0.73,5.98)	0.20 (0.35)
Percentage of member days for which student is present[b]	0.06 (0.01)***	1.06	(1.04,1.08)	0.01 (< 0.01)**
School-level predictor				
Priority status	−0.12 (0.11)	0.88	(0.71,1.10)	−0.06 (0.06)
PreK available (0 = no, 1 = yes)	−0.36 (0.24)	0.69	(0.43,1.12)	−0.03 (0.10)
Average SQRP rating[c,d]	0.08 (0.06)	1.08	(0.95,1.23)	0.01 (0.02)
Average percentage daily attendance[c]	0.07 (0.08)	1.08	(0.92,1.26)	0.03 (0.03)
Average proportion female[e]	0.12 (1.5)	1.13	(0.06,22.63)	−0.58 (0.53)
Average proportion eligible for the National School Lunch Program	−1.34 (1.61)	0.26	(0.01,6.50)	0.63 (0.61)
Average proportion receiving special education services[e]	0.22 (1.09)	1.25	(0.14,10.91)	0.75 (0.46)
Average proportion English learner students[e]	−0.92 (0.8)	0.40	(0.08,1.96)	−0.04 (0.27)
Average proportion Black[e]	0.74 (2.79)	2.09	(0.01,541.22)	1.16 (0.59)
Average proportion Hispanic[e]	2.24 (2.12)	9.37	(0.16,648.55)	0.92 (0.55)
Average proportion White[e]	3.34 (1.78)	28.1	(0.81,978.41)	2.22 (1.12)*
Average proportion all other races/ethnicities[e]	1.95 (2.41)	7.01	(0.06,858.38)	1.68 (0.63)*
Average years of experience in Chicago Public	0.00 (0.02)	1.00	(0.97,1.04)	0.01 (0.01)

Characteristic	TRC kindergarten and grade 1 (*n* = 7,971 students in 85 schools)			MAP grade 2 (*n* = 5,882 students in 115 schools)
	Logit (standard error)	Odds ratio	.95 confidence interval	Standardized coefficient (standard error)
Schools[a]				
Proportion of teachers certified[a]	0.28 (0.63)	1.32	(0.37,4.64)	0.03 (0.25)
Random effect (variance component)				
Variance in intercept		0.44***		0.04***
Variance in fall assessment score slope		0.06*		0.01***
Variance in grade[f]		0.51***		—
Variance in English as a second language status slope		0.31***		0.01**
Variance in special education status slope[g]		—		0.11***

* Significant at $p < .05$, ** significant at $p < .01$; *** significant at $p < .001$.
— Indicates that the parameter was not included in the model.
MAP is Measures of Academic Progress for the Primary Grades assessment. SQRP is School Quality Rating Policy. TRC is Text Reading and Comprehension assessment.
Note: Because the MAP results for grade 2 are from linear models and the TRC results for grades K and 1 are from logistic models, the model coefficients are not directly comparable.
a. Aggregated from teacher data.
b. Number of days for which students are enrolled in Chicago Public Schools.
c. Mean calculated for school years 2016/17 and 2017/18.
d. Recoded where Level 3 is 0, Level 2+ is 1, Level 2 is 2, Level 1+ is 3, and Level 1 is 4.
e. Aggregated from student data.
f. Grade status was not included in the MAP model because the model only includes grade 2, whereas the TRC model includes kindergarten and grade 1.
g. The slope in the TRC model was fixed, and no random component was included because the variability in the special education status slope was not significant.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools and publicly available 2016/17 and 2017/18 school year data from the Illinois State Board of Education and the Chicago Public Schools Accountability Reports.

Table C2. Predicting end-of-year Text Reading and Comprehension proficiency ratings with school priority status and student and school covariates (research question 1), 2018/19

	Model 1			Model 2			Model 3			Model 4			Model 5		
Characteristic	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI
Student-level predictor															
Intercept	−0.80*** (0.09)	0.45	(0.37, 0.54)	−0.34 (0.22)	0.72	(0.46, 1.11)	0.06 (0.3)	1.06	(0.58, 1.93)	0.11 (0.29)	1.12	(0.63, 1.98)	0.09 (0.28)	1.09	(0.62, 1.92)
Fall TRC scores	1.74*** (0.09)	5.68	(4.74, 6.80)	1.74*** (0.08)	5.72	(4.87, 6.70)	1.76*** (0.08)	5.80	(4.94, 6.81)	1.86*** (0.08)	6.45	(5.49, 7.57)	1.85*** (0.08)	6.37	(5.44, 7.48)
Grade (0 = kindergarten, 1 = grade 1)	—	—	—	0.30* (0.12)	1.35	(1.07, 1.72)	0.30* (0.12)	1.35	(1.06, 1.71)	0.22 (0.13)	1.25	(0.97, 1.61)	0.22 (0.13)	1.25	(0.97, 1.61)
Female	—	—	—	0.09 (0.06)	1.09	(0.97, 1.23)	0.09 (0.06)	1.09	(0.97, 1.23)	0.09 (0.06)	1.09	(0.97, 1.23)	0.09 (0.06)	1.10	(0.98, 1.23)
Ever eligible for the National School Lunch Program	—	—	—	−0.55** (0.19)	0.58	(0.40, 0.83)	−0.53** (0.18)	0.59	(0.41, 0.85)	−0.52** (0.18)	0.59	(0.41, 0.85)	−0.52** (0.18)	0.59	(0.41, 0.85)
Received special education services	—	—	—	−1.51*** (−0.14)	0.22	(0.17, 0.29)	−1.52*** (0.14)	0.22	(0.16, 0.29)	−1.53*** (0.14)	0.22	(0.16, 0.29)	−1.52*** (0.14)	0.22	(0.16, 0.29)
English learner student	—	—	—	−0.04 (0.13)	0.96	(0.74, 1.25)	−0.04 (0.13)	0.97	(0.74, 1.25)	−0.08 (0.13)	0.92	(0.71, 1.20)	<0.01 (0.13)	1.00	(0.77, 1.31)
White vs. Black	—	—	—	0.42* (0.2)	1.53	(1.03, 2.27)	0.24 (0.2)	1.27	(0.85, 1.90)	0.23 (0.2)	1.26	(0.84, 1.88)	0.23 (0.20)	1.25	(0.84, 1.88)
American Indian/ Alaska Native vs. Black	—	—	—	−0.27 (0.55)	0.77	(0.26, 2.30)	−0.50 (0.55)	0.60	(0.20, 1.83)	−0.50 (0.55)	0.61	(0.20, 1.84)	−0.50 (0.56)	0.61	(0.20, 1.85)
Hispanic vs. Black	—	—	—	0.06 (0.11)	1.06	(0.86, 1.31)	−0.18 (0.15)	0.84	(0.63, 1.11)	−0.18 (0.14)	0.84	(0.63, 1.11)	−0.18 (0.15)	0.84	(0.63, 1.12)
Multiracial vs. Black	—	—	—	0.45 (0.33)	1.56	(0.82, 2.99)	0.29 (0.32)	1.34	(0.71, 2.53)	0.30 (0.32)	1.35	(0.71, 2.54)	0.30 (0.32)	1.34	(0.71, 2.53)
Asian vs. Black	—	—	—	0.86** (0.31)	2.36	(1.28, 4.35)	0.73* (0.36)	2.07	(1.02, 4.21)	0.72* (0.36)	2.05	(1.01, 4.16)	0.72* (0.36)	2.06	(1.01, 4.21)
Pacific Islander/ Hawaiian vs. Black	—	—	—	0.90 (0.54)	2.45	(0.85, 7.13)	0.73 (0.54)	2.08	(0.73, 5.98)	0.76 (0.55)	2.14	(0.73, 6.23)	0.77 (0.55)	2.16	(0.73, 6.38)

Characteristic	Model 1			Model 2			Model 3			Model 4			Model 5		
	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI
Percentage of member days for which student is present[a]	—	—	—	0.06*** (0.01)	1.06	(1.04, 1.08)	0.06*** (0.01)	1.06	(1.04, 1.08)	0.06*** (0.01)	1.06	(1.04, 1.08)	0.06*** (0.01)	1.06	(1.04, 1.08)
Priority x fall TRC score	—	—	—	—	—	—	—	—	—	−0.20 (0.15)	0.82	(0.60, 1.11)	−0.18 (0.16)	0.84	(0.61, 1.14)
Priority x grade status	—	—	—	—	—	—	—	—	—	0.14 (0.23)	1.15	(0.73, 1.82)	0.13 (0.23)	1.14	(0.72, 1.80)
Priority x English learner status	—	—	—	—	—	—	—	—	—	0.07 (0.23)	1.07	(0.67, 1.71)	−0.10 (0.24)	0.90	(0.56, 1.46)
School-level predictor															
Priority status	−0.21 (0.16)	0.81	(0.59, 1.10)	−0.20 (0.14)	0.82	(0.62, 1.10)	−0.12 (0.11)	0.88	(0.71, 1.10)	−0.20 (0.22)	0.82	(0.53, 1.28)	0.08 (0.29)	1.08	(0.61, 1.93)
PreK available (0 = no, 1 = yes)	—	—	—	—	—	—	−0.36 (0.24)	0.69	(0.43, 1.12)	−0.37 (0.24)	0.69	(0.43, 1.10)	−0.37 (0.23)	0.69	(0.44, 1.09)
Average SQRP rating[b,c]	—	—	—	—	—	—	0.08 (0.06)	1.08	(0.95, 1.23)	0.07 (0.06)	1.08	(0.95, 1.22)	0.08 (0.06)	1.08	(0.95, 1.23)
Average percentage daily attendance[b]	—	—	—	—	—	—	0.07 (0.08)	1.08	(0.92, 1.26)	0.08 (0.08)	1.08	(0.92, 1.27)	0.10 (0.07)	1.11	(0.95, 1.28)
Average proportion female[d]	—	—	—	—	—	—	0.12 (1.5)	1.13	(0.06, 22.63)	0.22 (1.51)	1.24	(0.06, 25.16)	0.02 (1.54)	1.02	(0.05, 22.16)
Average proportion eligible for the National School Lunch Program	—	—	—	—	—	—	−1.34 (1.61)	0.26	(0.01, 6.50)	−1.26 (1.6)	0.28	(0.01, 6.91)	−1.30 (1.57)	0.27	(0.01, 6.22)
Average proportion receiving special education services[d]	—	—	—	—	—	—	0.22 (1.09)	1.25	(0.14, 10.91)	0.24 (1.09)	1.28	(0.14, 11.26)	0.39 (1.13)	1.48	(0.15, 14.19)

Characteristic	Model 1			Model 2			Model 3			Model 4			Model 5		
	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI
Average proportion English learner students[d]	—	—	—	—	—	—	−0.92 (0.8)	0.40	(0.08, 1.96)	−0.99 (0.78)	0.37	(0.08, 1.76)	−1.29 (0.78)	0.28	(0.06, 1.30)
Average proportion Black[d]	—	—	—	—	—	—	0.74 (2.79)	2.09	(0.01, 541.22)	0.68 (2.76)	1.97	(0.01, 480.47)	0.36 (2.85)	1.44	(0.01, 422.03)
Average proportion Hispanic[d]	—	—	—	—	—	—	2.24 (2.12)	9.37	(0.16, 648.55)	2.09 (2.08)	8.08	(0.13, 515.57)	1.57 (2.05)	4.80	(0.08, 289.81)
Average proportion White[d]	—	—	—	—	—	—	3.34 (1.78)	28.10	(0.81, 978.41)	3.24 (1.75)	25.59	(0.77, 845.29)	2.58 (1.74)	13.23	(0.41, 423.79)
Average proportion other race/ethnicity[d]	—	—	—	—	—	—	1.95 (2.41)	7.01	(0.06, 858.38)	1.83 (2.36)	6.24	(0.06, 690.65)	−2.03 (5.23)	0.13	(0.000, 4,513.54)
Average year of teaching in Chicago Public Schools[e]	—	—	—	—	—	—	0.00 (0.02)	1.00	(0.97, 1.04)	0.00 (0.02)	1.00	(0.97, 1.04)	<0.01 (0.02)	1.00	(0.97, 1.04)
Proportion of teachers certified[e]	—	—	—	—	—	—	0.28 (0.63)	1.32	(0.37, 4.64)	0.28 (0.63)	1.33	(0.38, 4.68)	0.27 (0.65)	1.31	(0.36, 4.76)
Average proportion White x Priority	—	—	—	—	—	—	—	—	—	—	—	—	0.05 (1.68)	1.05	(0.04, 29.86)
Average proportion Black x Priority	—	—	—	—	—	—	—	—	—	—	—	—	−0.71* (0.33)	0.49	(0.26, 0.95)
Average proportion other race/ethnicity x Priority	—	—	—	—	—	—	—	—	—	—	—	—	3.01 (4.49)	20.37	(0.00, 159,887.68)

Characteristic	Model 1			Model 2			Model 3			Model 4			Model 5		
	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI	Logit (SE)	Odds ratio	.95 CI
Random effect (variance component)		*Variance*			*Variance*			*Variance*			*Variance*			*Variance*	
Variance in intercept		0.34***			0.48***			0.44***			0.44***			0.67***	
Variance in fall TRC proficiency slope		0.17***			0.05*			0.06*			0.06*			0.26*	
Variance in grade		—			0.51***			0.51***			0.50***			0.71***	
Variance in English learner status slope		—			0.29***			0.31***			0.31***			0.54***	

* Significant at $p < .05$, ** significant at $p < .01$; *** significant at $p < .001$.
— Indicates that the parameter was not included in the model.
SE is standard error. SQRP is School Quality Rating Policy. TRC is Text Reading and Comprehension assessment.
Note: The sample included 7,971 students in 85 schools. Model 1 includes priority status and fall TRC scores only. Model 2 adds student covariates. Model 3 adds school covariates. Model 4 adds cross-level interactions between priority status and student covariates. Model 5 adds interactions between priority status and school covariates.
a. Number of days for which students are enrolled in Chicago Public Schools.
b. Mean calculated for school years 2016/17 and 2017/18.
c. Recoded where Level 3 is 0, Level 2+ is 1, Level 2 is 2, Level 1+ is 3, and Level 1 is 4. Level 3 indicates intensive support is needed. Level 2 indicates provisional support is needed. Levels 1+ and 2+ indicate good standing.
d. Aggregated from student data.
e. Aggregated from teacher data.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools and publicly available 2016/17 and 2017/18 school year data from the Illinois State Board of Education and the Chicago Public Schools Accountability Reports.

Table C3. Predicting end-of-year Measure of Academic Progress reading scores with school priority status and student and school covariates (research question 1), 2018/19

Characteristic	Model 1 Coefficient (SE)	Model 2 Coefficient (SE)	Model 3 Coefficient (SE)	Model 4 Coefficient (SE)	Model 5 Coefficient (SE)	Model 6 Coefficient (SE)
Student-level predictor						
Intercept	−0.02 (0.02)	0.12 (0.06)	0.13 (0.11)	0.12 (0.11)	0.13 (0.11)	0.46 (0.14)
Fall MAP score	0.71 (0.02)***	0.64 (0.02)***	0.64 (0.02)***	0.64 (0.02)***	0.64 (0.02)***	—
Female	—	0.07 (0.03)*	0.07 (0.03)*	0.07 (0.03)*	0.07 (0.03)**	0.18 (0.04)***
Ever eligible for the National School Lunch Program	—	−0.13 (0.05)*	−0.11 (0.05)*	−0.12 (0.05)*	−0.12 (0.05)**	−0.41 (0.06)***
Received special education services	—	−0.37 (0.06)***	−0.38 (0.06)***	−0.41 (0.05)***	−0.40 (0.05)***	−0.90 (0.05)***
English learner student	—	−0.11 (0.04)*	−0.10 (0.03)	−0.04 (0.05)	−0.03 (0.05)	−0.24 (0.07)***
White vs. Black	—	0.14 (0.07)	0.13 (0.06)*	0.13 (0.06)*	0.13 (0.06)*	0.23 (0.07)**
American Indian/Alaska Native vs. Black	—	0.30 (0.25)	0.32 (0.25)	0.32 (0.24)	0.32 (0.24)	0.34 (0.25)
Hispanic vs. Black	—	0.06 (0.05)	0.08 (0.05)	0.08 (0.06)	0.08 (0.06)	0.08 (0.06)
Multiracial vs. Black	—	−0.03 (0.19)	−0.01 (0.21)	−0.01 (0.20)	−0.01 (0.21)	0.22 (0.18)
Asian vs. Black	—	0.01 (0.06)	−0.03 (0.06)	−0.02 (0.06)	−0.02 (0.06)	0.17 (0.10)
Pacific Islander/Hawaiian vs. Black	—	0.20 (0.33)	0.20 (0.35)	0.18 (0.35)	0.18 (0.35)	0.35 (0.38)
Percentage of member days for which student is present[a]	—	0.01 (< 0.01)**	0.01 (< 0.01)**	0.01 (< 0.01)**	0.01 (< 0.01)**	0.02 (< 0.01)***
Priority x Fall MAP score	—	—	—	0.01 (0.05)	< 0.01 (0.05)	—
Priority x English learner status	—	—	—	−0.13 (0.09)	0.05 (0.14)	−0.20 (0.10)
Priority x special education status	—	—	—	0.06 (0.14)	−0.15 (0.10)	0.15 (0.13)
School-level predictor						
Priority status	−0.06 (0.05)	−0.05 (0.06)	−0.06 (0.06)	−0.04 (0.06)	−0.13 (0.11)	−0.17 (0.13)
PreK available (0 = no, 1 = yes)	—	—	−0.03 (0.10)	−0.03 (0.10)	−0.05 (0.10)	−0.09 (0.13)
Mean SQRP rating[b,c]	—	—	0.01 (0.02)	0.01 (0.02)	0.02 (0.02)	0.04 (0.03)
Mean percentage daily attendance[b]	—	—	0.03 (0.03)	0.03 (0.03)	0.02 (0.03)	0.03 (0.03)
Average proportion female[d]	—	—	−0.58 (0.53)	−0.62 (0.51)	−0.82 (0.50)	−0.85 (0.72)
Average proportion eligible for the National School Lunch Program[d]	—	—	0.63 (0.61)	0.57 (0.60)	0.27 (0.52)	−0.93 (0.52)

Average proportion receiving special education services[d]	—	—	0.75 (0.46)	0.71 (0.45)	0.43 (0.41)	1.01 (0.49)**
Average proportion English learner students[d]	—	—	−0.04 (0.27)	−0.02 (0.26)	0.01 (0.28)	−0.09 (0.33)
Average proportion Black[d]	—	—	1.16 (0.59)	1.23 (0.58)*	1.04 (0.54)	0.58 (0.60)
Average proportion Hispanic[d]	—	—	0.92 (0.55)	0.97 (0.55)	0.79 (0.53)	0.46 (0.54)
Average proportion White[d]	—	—	2.22 (1.12)*	2.24 (1.09)*	1.30 (0.91)	−0.11 (0.87)
Average proportion other races/ethnicities[d]	—	—	1.68 (0.63)*	1.73 (0.65)*	1.46 (0.55)*	1.63 (0.58)**
Average years of teaching in Chicago Public Schools[e]	—	—	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)*
Proportion of teachers certified[e]	—	—	0.03 (0.25)	0.03 (0.25)	0.16 (0.22)	0.20 (0.27)
Average proportion White x priority	—	—	—	—	2.51 (0.55)***	3.25 (0.65)***
Average proportion Black x priority	—	—	—	—	0.05 (0.16)	< 0.01 (0.17)
Average proportion other races/ethnicities x priority	—	—	—	—	−0.39 (0.37)	−0.51 (0.36)
Random effect (variance component)	*Variance*	*Variance*	*Variance*	*Variance*	*Variance*	*Variance*
Variance in intercept	0.03***	0.04***	0.04***	0.04***	0.03***	0.06***
Variance in Fall MAP score slope	0.01**	0.01***	0.01***	0.01***	0.01**	—
Variance in special education status slope		0.01**	0.11***	0.11***	0.11***	0.09**
Variance in English learner status slope		0.11***	0.01**	0.01*	0.01*	0.02**

* Significant at $p < .05$, ** significant at $p < .01$; *** significant at $p < .001$.

— Indicates that the parameter was not included in the model.

MAP is Measures of Academic Progress for Primary Grades Assessment. SE is standard error. SQRP is School Quality Rating Policy.

Note: The sample included 5,882 students in 115 schools. Model 1 includes priority status and fall MAP scores only. Model 2 adds student covariates. Model 3 adds school covariates. Model 4 adds cross-level interactions between priority status and student covariates. Model 5 adds interactions between priority status and school covariates.

a. Number of days for which students are enrolled in Chicago Public Schools.

b. Mean calculated for school years 2016/17 and 2017/18.

c. Recoded where Level 3 is 0, Level 2+ is 1, Level 2 is 2, Level 1+ is 3, and Level 1 is 4.

d. Aggregated from student data.

e. Aggregated from teacher data.

Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools and publicly available 2016/17 and 2017/18 school year data from the Illinois State Board of Education and the Chicago Public Schools Accountability Reports.

Table C4. Predicting end-of-year Text Reading and Comprehension proficiency ratings and Measures of Academic Progress for Primary Grades reading assessment scores with school priority status and student and school covariates, using only students who have beginning-of-year and end-of-year scores (research question 1)

	TRC kindergarten and grade 1 (*n* = 7,450 students in 85 schools)			MAP grade 2 (*n* = 2,497 students in 71 schools)
Characteristic	Logit (standard error)	Odds ratio	.95 confidence interval	Standardized coefficient (standard error)
Student predictor				
Intercept	−0.20 (0.28)	0.82	(0.47,1.43)	−0.04 (0.10)
Fall assessment score	1.86 (0.09)***	6.43	(5.36,7.72)	0.71 (0.02)
Grade (0 = kindergarten, 1 = grade 1)[a]	0.38 (0.13)**	1.46	(1.12,1.91)	—
Female	0.09 (0.05)	1.09	(0.99,1.21)	0.05 (0.03)
Ever eligible for the National School Lunch Program	−0.41 (0.12)**	0.66	(0.52,0.84)	−0.11 (0.05)*
Receives special education services	−1.67 (0.13)***	0.19	(0.14,0.24)	−0.37 (0.10)***
English learning student	0.05 (0.13)	1.05	(0.82,1.35)	−0.07 (0.04)*
White vs. Black	0.32 (0.24)	1.37	(0.86,2.19)	0.12 (0.06)
American Indian/Alaska Native vs. Black	−0.57 (0.56)	0.57	(0.19,1.72)	0.23 (0.21)
Hispanic vs. Black	−0.16 (0.18)	0.85	(0.59,1.22)	0.09 (0.07)
Multiracial vs. Black	0.60 (0.38)	1.82	(0.86,3.83)	0.03 (0.23)
Asian vs. Black	0.74 (0.18)***	2.09	(1.46,2.99)	−0.03 (0.07)
Pacific Islander/Hawaiian vs. Black	0.72 (0.57)	2.05	(0.68,6.23)	−0.06 (0.16)
Percentage of member days for which student is present[b]	0.06 (0.01)***	1.06	(1.05,1.08)	0.01 (0.00)
School predictor				
Priority status	−0.18 (0.13)	0.84	(0.65,1.08)	−0.03 (0.05)
PreK available (0 = no, 1 = yes)	−0.27 (0.24)	0.77	(0.48,1.23)	0.17 (0.09)
Average SQRP rating[c,d]	0.05 (0.07)	1.05	(0.92,1.20)	0.00 (0.02)
Average percentage daily attendance[c]	0.09 (0.09)	1.09	(0.91,1.31)	0.01 (0.03)
Average proportion female[e]	−0.38 (1.73)	0.68	(0.02,21.72)	−1.53 (0.67)
Average proportion eligible for the National School Lunch Program	−1.92 (1.82)	0.15	(0.00,5.54)	0.06 (0.49)
Average proportion receiving special education services[e]	0.08 (1.26)	1.09	(0.09,13.39)	0.23 (0.42)
Average proportion English learner students[e]	−0.68 (0.85)	0.51	(0.09,2.76)	0.34 (0.18)
Average proportion Black[e]	1.13 (3.05)	3.08	(0.01,1347.22)	1.79 (1.01)
Average proportion Hispanic[e]	2.93 (2.31)	18.76	(0.19,1891.76)	1.17 (0.69)
Average proportion White[e]	3.84 (1.93)	46.48	(0.99,2179.96)	0.80 (0.65)
Average proportion all other races/ethnicities[e]	0.88 (2.36)	2.42	(0.02,270.64)	1.10 (0.69)
Average years of teaching experience in Chicago Public Schools[a]	0.01 (0.02)	1.01	(0.97,1.05)	0.01 (0.01)
Proportion of teachers certified[a]	0.06 (0.68)	1.06	(0.27,4.14)	0.26 (0.27)

Characteristic	TRC kindergarten and grade 1 (*n* = 7,450 students in 85 schools)			MAP grade 2 (*n* = 2,497 students in 71 schools)
	Logit (standard error)	Odds ratio	.95 confidence interval	Standardized coefficient (standard error)
Random effect (variance component)				
Variance in intercept		0.49***		0.05***
Variance in fall assessment score slope		0.09*		0.01***
Variance in grade[f]		0.65***		—
Variance in English learner status slope		0.32**		0.02
Variance in special education status slope[g]		—		0.24**

* Significant at $p < .05$, ** significant at $p < .01$; *** significant at $p < .001$.
— Indicates that the parameter was not included in the model.
MAP is Measures of Academic Progress for the Primary Grades assessment. SQRP is School Quality Rating Policy. TRC is the Text Reading and Comprehension assessment.
a. Aggregated from teacher data.
b. Number of days for which students are enrolled in Chicago Public Schools.
c. Measure is averaged across two school years, 2016/17 and 2017/18.
d. Recoded where Level 3 is 0, Level 2+ is 1, Level 2 is 2, Level 1+ is 3, and Level 1 is 4.
e. Aggregated from student data.
f. Grade status was not included in the MAP model because the model only includes grade 2, whereas the TRC model includes kindergarten and grade 1.
g. The slope in the TRC model was fixed, and no random component was included because the variability in the special education status slope was not significant.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools and publicly available 2016/17 and 2017/18 school year data from the Illinois State Board of Education and the Chicago Public Schools Accountability Reports.

Table C5. Number and percentage of teachers participating, by number of professional development sessions attended and school priority status (research question 2), 2018/19

Number of professional development sessions attended	Priority schools (*n* = 26)		Nonpriority schools (*n* = 89)			
	Number of teachers	Percent of teachers	Number of teachers	Percent of teachers	X^2 test statistic	*p*-value
Did not participate**	18	8.0	150	20.3	7.89	0.00
1 session	67	29.9	189	25.5	1.68	0.19
2 sessions	54	24.1	207	28.0	1.30	0.25
3 sessions**	85	38.0	194	26.2	11.50	0.00

** statistically significant at $p < .01$ (based on Pearson chi-square tests with 1 degree of freedom).
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools.

Table C6. School-level teacher participation rates, by number of professional development sessions attended and school priority status (research question 2), 2018/19

Participation rate	Priority schools (*n* = 26)		Nonpriority schools (*n* = 89)			
	Number of schools	Percent of schools	Number of schools	Percent of schools	X^2 test statistic	*p*-value
Fewer than 75 percent of teachers in the school attended one or more sessions	2	7.7	21	23.6	3.18	0.08
At least 75 percent of teachers in the school attended one or more sessions	24	92.3	68	76.4		

Note: Statistical significance was determined using Pearson chi-square tests with 1 degree of freedom.
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools.

Table C7. Number of sessions in which at least one administrator from a school attended professional development, by school priority status (research question 2), 2018/19

Number of professional development sessions attended	Priority schools (*n* = 26)		Nonpriority schools (*n* = 89)		*p*-value
	Number of schools	Percent of schools	Number of schools	Percent of schools	
Did not participate	1	6.3	2	3.4	0.52
1 session	3	18.8	17	29.3	0.53
2 sessions	4	25.0	20	34.5	0.56
3 sessions	8	50.0	19	32.8	0.25

Note: Differences were not statistically different at $p < 0.05$ (based on Fisher's exact two-sided tests).
Source: Authors' analysis of 2018/19 school year data provided by Chicago Public Schools.