U.S. Department of Education
September 2019

Evaluation of Support for Using Student Data to Inform Teachers' Instruction

APPENDICES

Philip Gleason
Sarah Crissey
Greg Chojnacki
Marykate Zukiewicz
Tim Silva
Mathematica

Sarah Costelloe
Abt Associates

Fran O'Reilly
Evidence-Based Education Research & Evaluation, LLC

Erica Johnson
Project Officer
Institute of Education Sciences

IES INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

CONTENTS

APPENDIX A

SUPPLEMENTAL INFORMATION ON STUDY DESIGN, DATA, AND METHODS

This appendix provides additional information on the study methodology presented in Chapter II. The sections that follow describe the details of our approach to conducting the study, including selecting the sample, implementing the study's research design, collecting data for the study, and conducting the analyses.

Study Sample

Study Districts and Schools

As described in Chapter II, we recruited 12 districts into the study in spring and summer 2014. The study aimed to include 8 to 10 schools per district, so recruitment efforts focused on districts with at least 8 schools containing fourth- and-fifth grade classrooms. Eligible districts were located in states that had adopted the Common Core State Standards. We selected districts that planned to administer well-aligned summative and interim assessments in 2014-2015 and 2015-2016, and did not have a DDI intervention in place or planned for these school years.

Districts and schools were eligible for study participation if they planned to administer both summative and interim assessments that were reasonably well aligned (and ideally created by the same test developer) during the 2014-2015 and 2015-2016 school years. These assessments also needed to be aligned with state content standards. By ensuring alignment of interim and summative assessments, we intended to maximize the usefulness of student assessment data available to help support teachers and administrators in implementing DDI.

We largely targeted states and districts in the Smarter Balanced Assessment Consortium due to the likelihood they would be using both summative and interim assessments highly aligned with each other and with the state's content standards throughout the intervention period. At the time we recruited them into the sample, 11 of the 12 sample districts had plans to administer Smarter Balanced summative assessments; seven of these districts also planned to use Smarter Balanced interim assessments. We included states outside the Smarter Balanced Assessment Consortium if we believed those conditions were present (summative and interim assessments aligned with each other and state content standards).[1]

For purposes of ensuring a sufficient contrast between treatment and control schools, the study targeted districts that had made minimal (or no) efforts to implement a DDI intervention. Because there is no single, commonly used definition of what features or activities a DDI intervention includes, we focused on the key elements of the DDI intervention planned for this study. In particular, we did not recruit districts with data coaches who worked directly with elementary school teachers to understand and use data to guide their instruction. Similarly, we avoided districts in which key school staff received regular professional development oriented toward helping teachers use data, or those having schools whose teachers participated regularly in professional learning communities specifically designed to help them use data.

Within study districts, recruitment efforts targeted schools that could benefit most from an intervention designed to raise student achievement—often schools that serve large percentages of

[1] We considered including districts in states that were in the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium but did not do so because of concerns that interim assessments aligned with PARCC summative assessments would not be fully developed by the time of the evaluation period.

low-income students. Thus, the study aimed to include schools with a high percentage of students receiving free or reduced-price meals, although there was no specific requirement as to the percentage of low-income students in study schools.

As with districts, we recruited schools within study districts that had not implemented (or had no plans to implement) key features of the study's DDI intervention. Along with the criteria mentioned above, the study avoided schools making efforts to implement Response to Intervention (RtI) programs, given that these programs use a data-driven approach. In addition, schools that had received School Improvement Grants (SIG) were not eligible for the study, given that DDI was an important part of required activities under some of the intervention models in this program. Finally, the study excluded charter schools, which often use data-driven approaches to instruction and, in some districts, are not included in district administrative records.

We targeted 103 districts for recruitment, and the sample recruited into the study included 12 districts and 104 schools. Ten of these districts contributed either 8 or 10 schools into the sample, with one district contributing 4 schools and one contributing 14.

Among the 91 districts not selected for the sample out of the original 103 districts that were targeted, we dropped approximately half (44 districts) because they lacked interest in the study and half (47 districts) because they did not meet all eligibility criteria. The most common reason that districts were found ineligible was being located in a state that did not plan to use summative and/or interim assessments that met our criteria. We dropped other districts because they already had an existing DDI program in place or did not have a sufficient number of schools that met eligibility criteria. We deemed some districts ineligible based on multiple criteria.

Our final sample included 102 schools in the 12 study districts. We dropped two study schools from the analytical sample after the start of the intervention period. One of these schools—the one assigned to the control group—originally served kindergarten through fifth grade but transitioned in summer 2015 to serving kindergarten through second grade. Because this school no longer served the grades that were the focus of the intervention, we dropped it, along with the treatment school to which it had been matched for purposes of random assignment.

Table A.1 displays the characteristics of students enrolled in study schools and those enrolled in all public school districts in the United States. Among students in study schools in 2013-2014, 53 percent were White and 15 percent were located in a large city, compared with 56 percent and 16 percent, respectively, among enrolled students in all U.S. districts. Although we do not have information on the percentage of English language learners in study schools, the percentage of such students was 9 percent in study districts, which was also the percentage of English language learners in all U.S. districts. About two-thirds of students in study schools were eligible for free or reduced-price meals, which is consistent with the recruiting strategy of targeted schools with large proportions of low-income students.

Table A.1. Comparison of study districts to all districts and largest districts in the United States[a]

	All districts in the U.S.	100 largest U.S. districts	Study districts	Study schools
Total enrollment (median)	1,159	70,677	18,505	452 (per school)
Percentage of students in large city	16%	47%	32%	15%
Free or reduced-price meal (percentage of students)	52%	60%	56%	64%
Student race and ethnicity (percentage of students)				
Percentage White	56%	36%	56%	53%
Percentage non-White	44%	64%	44%	47%
English language learners (percentage of students)	9%	13%	9%	n.a
Number of districts	13,079	100	12	102 schools (12 districts)

Source: 2013-2014 Common Core of Data.

[a]The last year of data available for comparisons across all districts in the United States was 2013-2014.

n.a. = Not available.

Samples of Principals, Teachers, and Students

The study aimed to include as many principals, teachers, and students from study schools in data collection efforts as possible. Accordingly, the principal sample included the principals of all 102 study schools in spring 2016.

There were 543 full-time fourth- and fifth-grade math and English/language arts teachers at study schools. Given resource constraints, we randomly sampled 501 teachers (roughly 5 teachers per school) from the larger population. We stratified the sample by school size to ensure that all eligible teachers from schools with 4 or fewer total fourth- and fifth-grade teachers would be included in the sample. During data collection, we identified some additional teachers who were ineligible—either on leave during 2015-2016 or not full-time, regular status teachers—thus, the final sample included 470 teachers.

The student sample included all fourth- and fifth-grade students in study schools as of spring 2016. Because of the possibility that the DDI intervention affected student mobility, and hence the composition of students in treatment schools, we conducted a sensitivity test using an alternative student sample. This sample included students enrolled in second and third grades in study schools as of spring 2014, before random assignment. This group of students comprised those who would be fourth- and fifth-graders in the treatment and control schools in spring 2016 if they maintained normal progress and did not leave their schools over this period.

The samples of principals, teachers, and students included in the analysis were based on the full samples described here. Figures A.1 through A.6 show the final analysis sample in each case after fully accounting for the study design and data collection efforts.

Figure A.1 Description of principal analysis sample

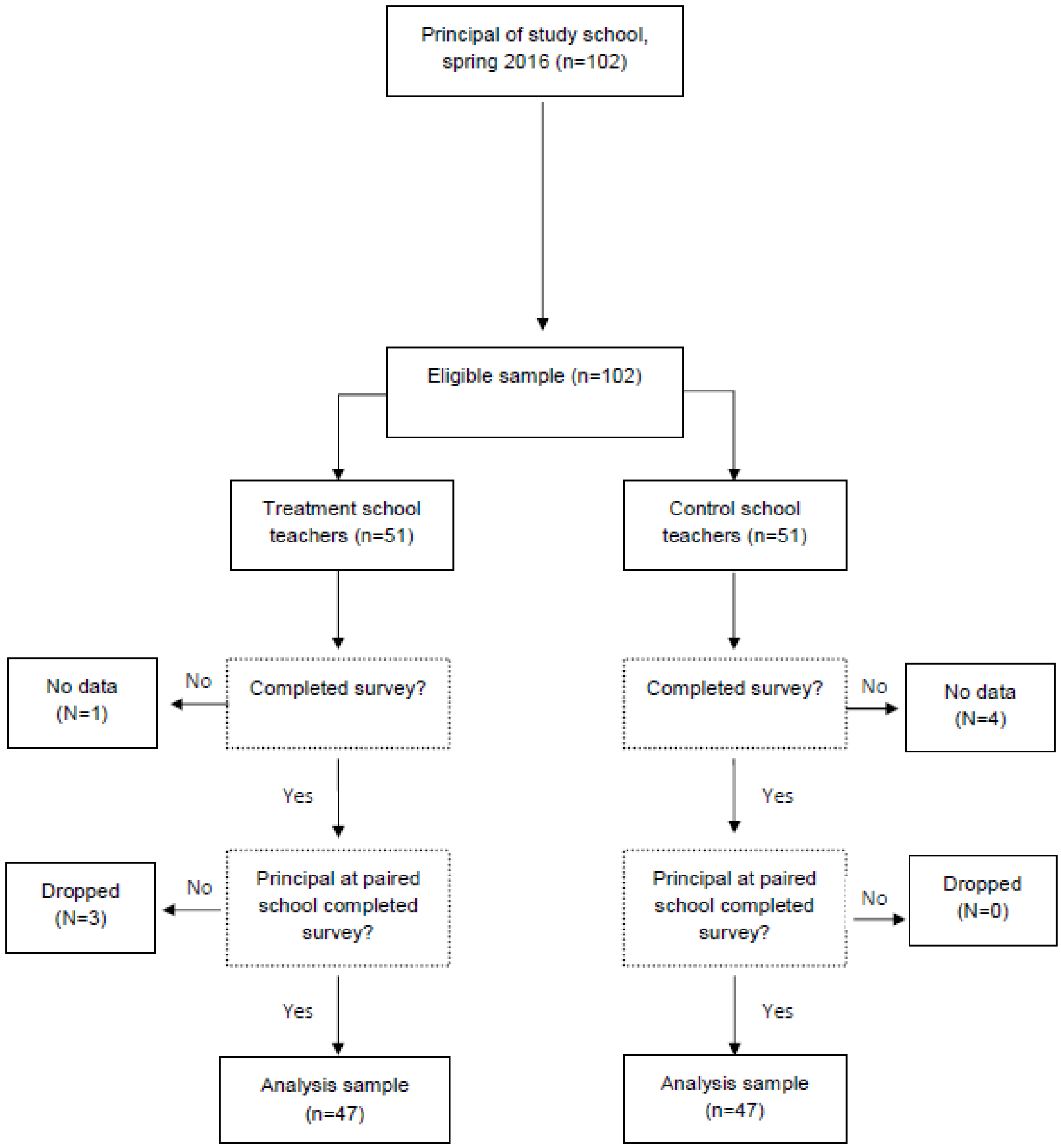


Figure A.2 Description of teacher analysis sample

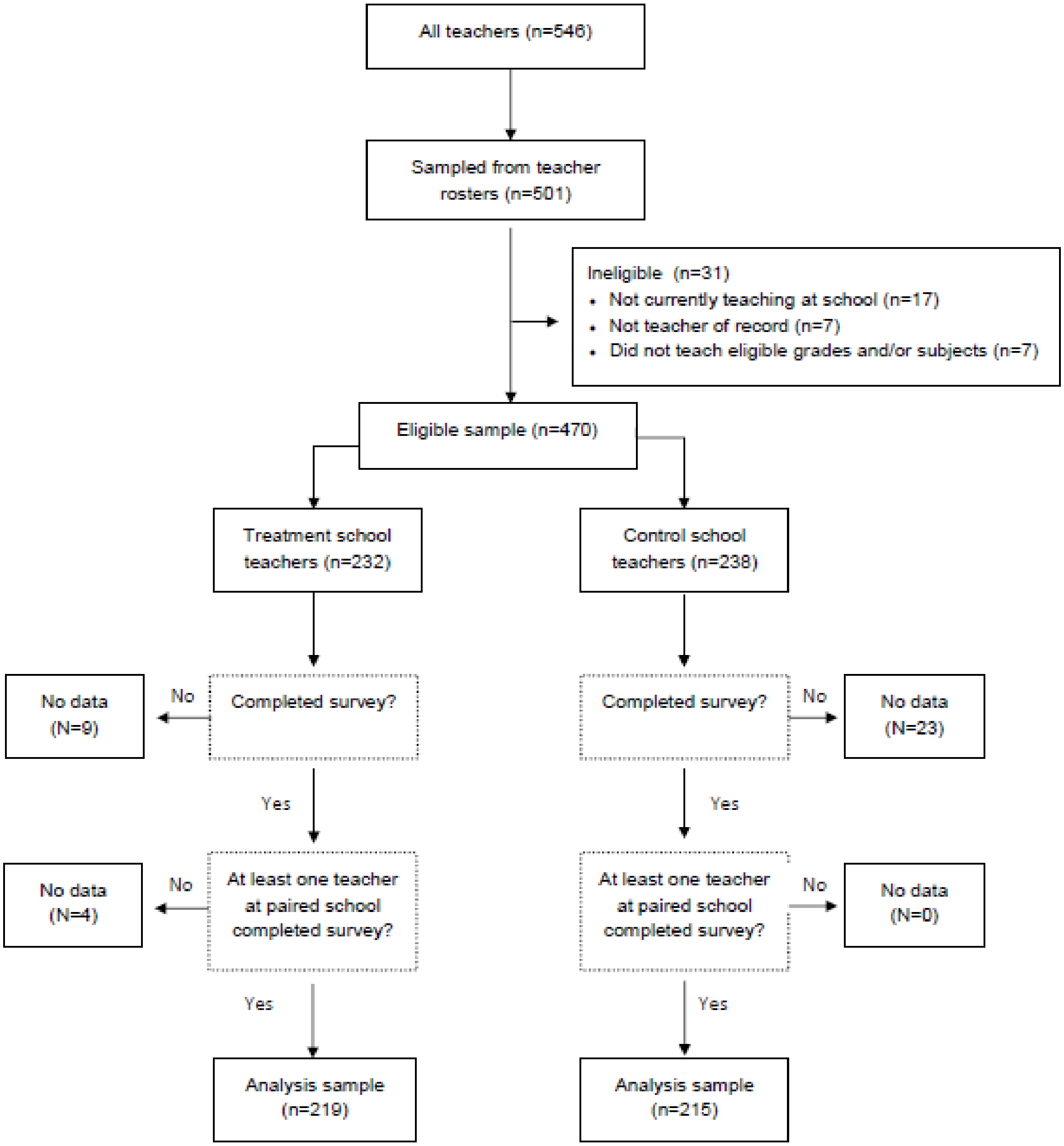
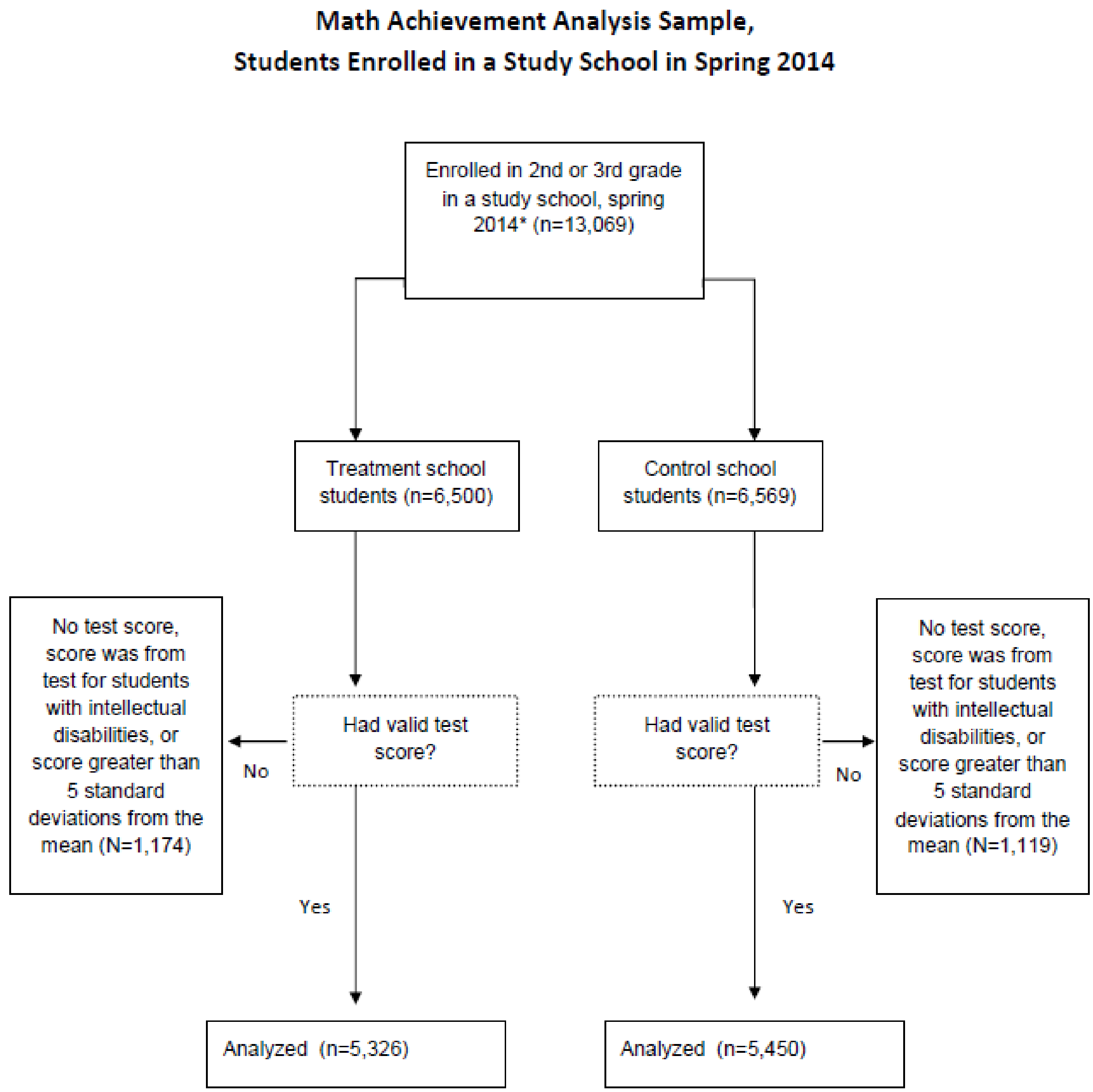
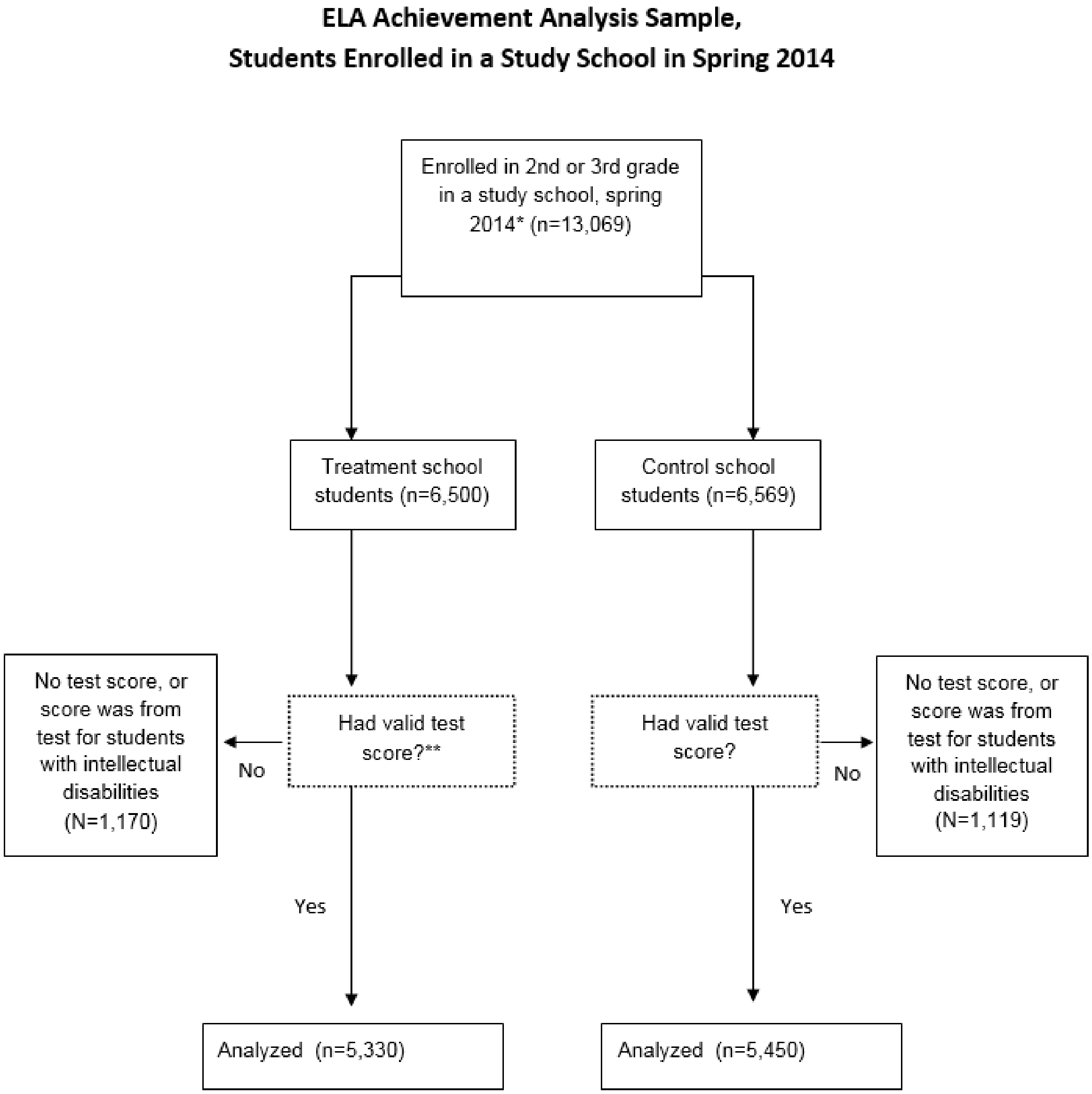


Figure A.3 Description of student analysis sample: Math achievement 2014



* For one of the twelve study districts, student records were only provided for students who were enrolled in the district both in spring 2016 and spring 2014. In other words, the 2014 enrollment sample for this district excluded students who left the school during the 2014-15 or 2015-16 school year.

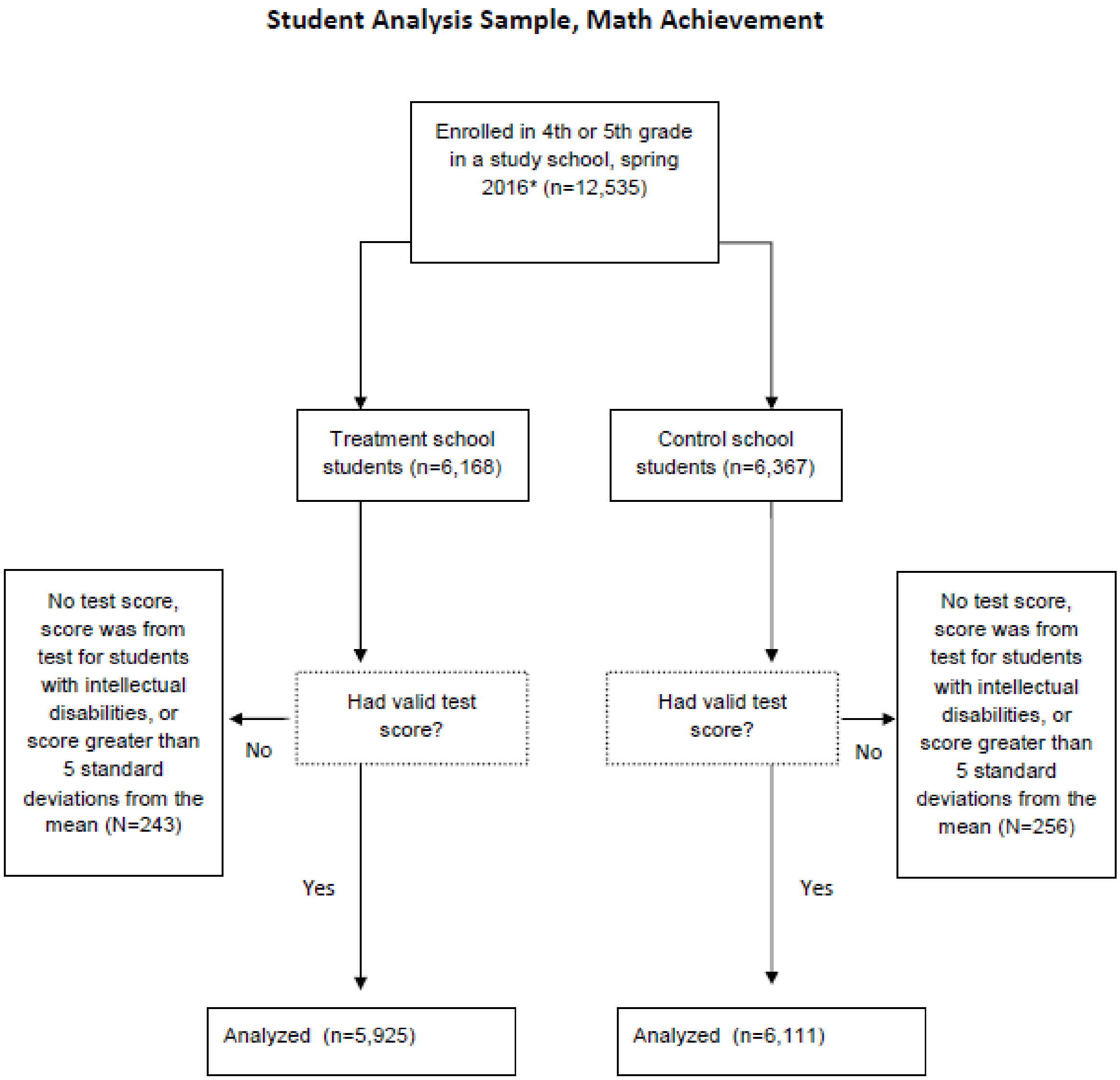
Figure A.4 Description of student analysis sample: ELA achievement 2014



* For one of the twelve study districts, student records were only provided for students who were enrolled in the district both in spring 2016 and spring 2014. In other words, the 2014 enrollment sample for this district excluded students who left the school during the 2014-15 or 2015-16 school year.

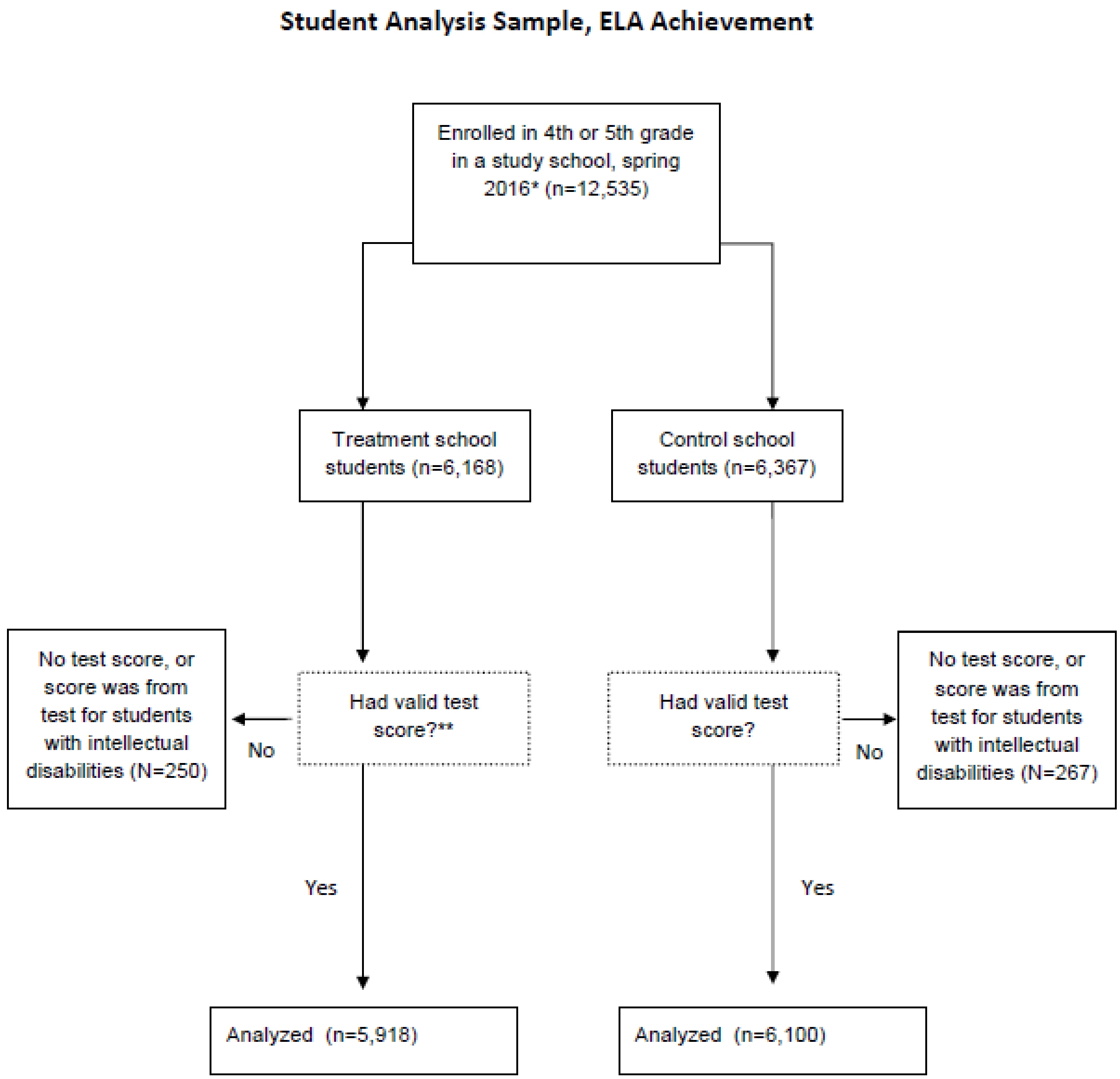
** There were no students in the sample with ELA test scores greater than 5 standard deviations from the mean.

Figure A.5 Description of student analysis sample: Math achievement 2016



* For one of the twelve study districts, student records were only provided for students who were enrolled in the district both in spring 2016 and spring 2014. In other words, the 2016 enrollment sample for this district excluded students who joined the school during the 2014-15 or 2015-16 school year.

Figure A.6 Description of student analysis sample: ELA Achievement 2016



* For one of the twelve study districts, student records were only provided for students who were enrolled in the district both in spring 2016 and spring 2014. In other words, the 2016 enrollment sample for this district excluded students who joined the school during the 2014-15 or 2015-16 school year.

** There were no students in the sample with ELA test scores greater than 5 standard deviations from the mean.

Data Coach Sample

All data coaches assigned to study schools as of the 2015-2016 school year were included in the coach sample. In the 2014-2015 school year, 41 coaches were assigned to the 51 study schools, including 10 who were assigned to two schools. Five of these data coaches did not return in the 2015-2016 school year, and four were replaced (including one case where a coach previously assigned to one school in the 2014-2015 school year was assigned to that school plus a second school in 2015-2016). One school did not replace its coach for the 2015-2016 school year. Coaches who were assigned to more than one school were asked to complete separate logs and interviews for each of the schools. Thus, the sample of data coaches included 39 coaches assigned to 50 schools in the 2015-2016 school year, including 11 coaches assigned to two schools.

Matched-Pair School-Level Random Assignment Design

As described in Chapter II, we used an experimental design to estimate the impacts of DDI on student achievement and other outcomes. After recruiting a sample of schools into the study, we matched pairs of similar schools within each study district. We then randomly assigned one school in each pair to the treatment group and the other school to the control group. The sections below describe statistical power of the design, the matching of schools, random assignment, and attrition of students from the sample's treatment and control groups.

Statistical Power

We designed the study to have enough statistical power to detect impacts from support for DDI of a size relevant to policymakers and practitioners. Table A.2 presents the minimum detectible effects (MDEs) on outcomes at the student, teacher, and principal level, for the study's intended or assumed sample of 104 schools, with 5 teachers per school and 112 students per school. The MDEs incorporate conservative assumptions about design effects due to the clustering of students or teachers in schools and precision gains from regression adjustments and stratified random assignment. These assumptions (listed in the notes of Table A.2) are based on estimates from recent large-scale studies in education with school-level random assignment. The study was designed to detect an impact on student achievement of 0.12 standard deviations and impacts on teacher and principal outcomes of 0.33 and 0.59 standard deviations, respectively. As shown in the table, MDEs for student or school subgroups are somewhat larger.

Table A.2. Minimum detectible effects in study design

	Sample size			Minimum detectable effect		
Number of schools	Number of students per school	Number of teachers per school	Number of principals per school	Student outcome	Teacher outcomes	Principal outcomes
Planned Statistical Power						
Balanced design						
104 (52 T, 52 C)	112	5	1	0.12	0.33	0.59
Student/teacher subgroup						
104 (52 T, 52 C)	56	3	1	0.14	--	--
School Subgroup						
52 (26 T, 26 C)	112	5	1	0.18	--	--
Realized Statistical Power						
Balanced design						

Table A.2. *(continued)*

Number of schools	Sample size: Number of students per school	Number of teachers per school	Number of principals per school	Minimum detectable effect: Student outcome	Teacher outcomes	Principal outcomes
102 (51 T, 51 C)	123	4.6	1	0.08	0.12	0.27
Student/teacher subgroup						
102 (51 T, 51 C)	62	2.3	1	0.12	--	--
School Subgroup						
50 (25 T, 25 C)	123	4.6	1	0.12	--	--

Notes: Planned MDEs were calculated assuming (1) a stratified random assignment design; (2) a two-tailed test; (3) a 5 percent significance level and an 80 percent level of power; (4) a school-level intraclass correlation of 0.15; (5) a response rate to the teacher survey of 85 percent; and (6) reductions in variance of 40 percent at the student level, 70 percent at the school level from the inclusion of covariates in the student outcome models, 10 percent at the teacher and principal levels, and 10 percent at the school level from the inclusion of covariates in the teacher outcome models. Realized MDEs account for the actual values of these parameters. Since no subgroup analysis was planned for teacher outcomes and principal outcomes we did not calculate MDEs in these cases, represented in the table by two dashes.

C= control; T= treatment.

The bottom panel of the table shows the realized statistical power based on the actual sample sizes and realized level of statistical precision of the impact estimate. The realized sample size of 102 schools, about 123 students per school, and 4.6 teachers per school were close to the assumed values, and teacher survey response rate was higher than the assumed value. Overall, the realized statistical power was better than the planned statistical power. For example, the MDE for student outcomes based on the full sample was 0.08 standard deviations, compared with the planned value of 0.12.

Matching of Schools

We implemented a matched-pair random assignment design by matching pairs of similar schools within each study district, as recommended by Imai et al. (2009). Schools in a given pair needed to be similar with respect to characteristics likely to be related to the key study outcome of student achievement. We primarily used existing administrative data to match schools but also wanted to account for harder-to-observe characteristics potentially related to student achievement that might be known to district officials but hidden from the study team.

The matching of schools within districts used a three-step process. The first step was to identify schools eligible for matching and identify subsets of schools that could be matched with one another. The goal of this step was to identify schools that were eligible and would ultimately result in the best set of matched pairs of schools. As described above, we excluded schools that were unwilling to participate or had a formal DDI program in place, such as charter and SIG schools. If the district included any magnet or Title I-eligible elementary schools, we aimed to match them with each other if possible. Second, among remaining schools, we matched those with similar observable characteristics, including enrollment, the school's racial/ethnic distribution, the percentage of students eligible for free or reduced-price meals, and the fourth- and fifth-grade proficiency rates in math and English/language arts. After matching schools based on observable characteristics, the third step involved confirming the face validity of these preliminary matched pairs with district officials. In particular, we sought feedback from district staff as to whether these pairs were well matched, particularly regarding characteristics we could not directly observe. We adjusted the pairs as appropriate on the basis of their feedback.

Random Assignment

The matching of schools resulted in 52 matched pairs; we included 51 of them in the final analysis sample. We then conducted random assignment separately (and independently) in each matched pair. Within each pair, we randomly assigned one school to be in the study's treatment group and the other in the control group. The random assignment process, and the fact that matching was designed to pair those schools similar in the characteristics used in the matching process, were designed to work together to ensure that the study's treatment and control schools were similar at baseline. Thus, at the beginning of the study, the two groups of schools differed systematically only in that schools in the treatment group had access to the DDI intervention.

We verified the baseline similarity of treatment and control schools by examining their characteristics as well as those of their students and teachers. There were few significant differences between treatment and control schools in these characteristics. At the school level, treatment and control schools were similar in their grade span, mean enrollment, and Title I and magnet status (table A.3). The two groups also were similar in various measures of student achievement. At the school level, the schools' proficiency rates in fourth and fifth grades did not significantly differ in 2014, nor did their mean assessment score in 2013. Among students in the study sample for whom baseline achievement scores were available, those in treatment and control schools had similar 2014 scores.[2]

Table A.3. Characteristics of treatment and control schools

	Treatment schools	Control schools	Difference
Characteristics of schools in sample			
Number of student enrolled (mean)	441	456	-16
School grade span (proportion)			
Pk-5 or K-5	0.73	0.73	0.00
Pk-6 or K-6	0.12	0.12	0.00
Pk-7	0.02	0.00	0.02
Pk-8	0.13	0.15	-0.02
2014 proficiency rate (mean proportion proficient)			
Math: Grade 4	0.42	0.44	-0.01
Math: Grade 5	0.44	0.48	-0.04
English/language arts: Grade 4	0.43	0.44	-0.01
English/language arts: Grade 5	0.43	0.46	-0.03
2013 assessment score (mean Z-score)			
Math	-0.31	-0.27	-0.05
English/language arts	-0.28	-0.28	0.01

[2] Students in the sample who were in fifth grade in 2016 were more likely to have valid 2014 test scores. Those who were in fourth grade in 2016 were unlikely to have 2014 scores because most were in second grade in that year (and thus were not tested). At the school level, we lacked 2014 test scores or proficiency rates for three districts—one in each of three states—because they participated in a pilot test of the Smarter Balanced assessment. Thus, to assess baseline achievement levels in treatment and control schools, we used a combination of 2013 and 2014 scores.

Table A.3. *(continued)*

	Treatment schools	Control schools	Difference
School status (proportion)			
Title I	0.77	0.83	0.06
Schoolwide Title I	0.75	0.71	0.04
Magnet	0.10	0.12	-0.02
Characteristics of students in sample			
Male (proportion)	0.52	0.52	0.00
Race/ethnicity (proportion)			
White	0.48	0.51	-0.03*
Black	0.22	0.16	0.06*
Hispanic	0.19	0.23	-0.03*
Other	0.11	0.10	0.01
Other student characteristics (proportion)			
Special education	0.15	0.15	0.00
English language learner	0.12	0.17	-0.05*
Eligible for free/reduced-price meals	0.66	0.65	0.00
2014 student achievement (mean z-score)			
Math	-0.22	-0.17	-0.04
English/language arts	-0.22	-0.23	0.01
2014 student attendance rate (mean proportion)	0.95	0.96	0.00
Characteristics of teachers in sample			
Female (proportion)	0.83	0.87	-0.04
White (proportion)	0.86	0.89	-0.03
Have master's degree (proportion)	0.27	0.31	-0.04
Years of experience teaching (mean)	13.0	12.4	0.6
Number of schools	51	51	
Number of teachers	212-219	209-215	
Number of students	5,925	6,111	

Source: 2012-2013 Common Core of Data, State Departments of Education and websites, GreatSchools.org, district administrative data, teacher survey.

Note: We estimated differences in 2014 proficiency rates among the 70 schools that fielded state tests in spring 2014. We estimated differences in 2013 school-level assessment scores using the full sample of 102 schools.

*Difference is statistically significant at the .05 level, two-tailed test.

The characteristics of students in treatment and control schools at baseline were also similar in most other respects. Each group of schools had the same proportion of male and special education students; a very similar share in each group was eligible for free or reduced-price meals. The attendance rate in each group of schools was also very similar. There were differences in variables related to students' race/ethnicity. Students in treatment schools were more likely to be Black, less likely to be White or Hispanic, and less likely to be English language learners. Finally, there were no significant differences between teachers in treatment and control schools in the characteristics we examined.

Student Attrition

Although we randomly assigned schools, our key outcome is based on data collected from students. Thus, we had to define what groups of students should be included in the treatment and control groups. One option was to define the study sample based on students' enrollment during the 2015-2016 intervention year, when we collected the outcome data. This option would have ensured that all students in the treatment group potentially would have been affected by the DDI intervention, and would maximize the proportion of students for whom outcome data would be available. On the other hand, impact estimates based on this sample would be affected by any influence the DDI intervention had on student mobility and the resulting composition of students in treatment and control schools. To address this possibility, we could define the study sample based on enrollment in spring 2014, just before random assignment. This approach would ensure that the composition of students in the treatment and control groups would not be influenced by the intervention itself. However, there were two down sides to this approach. First, some students in the treatment group would not have attended treatment schools during the 2015-2016 school year and thus would not have been affected by the intervention in that year. Second, we would be less likely to obtain outcome data for students in each group because we would not have information on outcomes for students who left the district between spring 2014 and spring 2016.

For our main analysis, we defined the treatment and control group student samples based on student enrollment in spring 2016. We based the student samples on spring 2016 enrollment because we believed it was unlikely that the intervention would have affected student mobility, given the nature of DDI.[3] The student sample included 6,168 students in 51 treatment schools and 6,367 students in 51 control schools, for a total of 12,535 students.

Using this definition of the student sample, attrition would result only from students enrolled in a treatment or control school in spring 2016 but for whom we did not have a valid test score. Based on this definition, attrition from the sample was about 4 percent (table A.4), including students who did not take any test, took a nonstandard test such as one for students with intellectual disabilities, or had a score that implied a clerical error or an extreme value.[4] The difference in the rate of attrition between treatment and control students was small and not statistically significant.[5]

---

[3] One district in the study was able to provide data only on students enrolled in study schools in spring 2014. Student records received from that district included students enrolled in a relevant grade in spring 2014 but not those who joined a study school sometime between spring 2014 and spring 2016. The sample for this district included only students enrolled in a study school in both 2014 and 2016. We estimated impacts across all districts in the sample using a similarly constructed set of students in each district (those in a study school in spring 2016 who had also been enrolled in a study school in spring 2014) to assess whether impact estimates were sensitive to this aspect of our analysis methods. We found that the results of this sensitivity analysis were very similar to the main impact results presented in the report.

[4] The last category is made up of three students whose 2016 math scores were more than five standard deviations from the mean z-score. Appendix tables C.8 and C.9 present impact estimates produced using a sample that includes these three students.

[5] These estimates of student attrition do not include the two schools dropped from the analysis during the 2015-2016 school year. These schools were in the same matched pair and were dropped after the study team learned that the control school had converted to serving only kindergarten through second grade.

Table A.4. Attrition from the spring 2016 and spring 2014 enrollment samples

	Math analysis sample			ELA analysis sample		
	Treatment schools	Control Schools	Difference	Treatment schools	Control schools	Difference
2016 enrollment sample						
Enrolled in study school spring 2016	6,168	6,367		6,168	6,367	
Enrolled in study school and had valid test score, spring 2016						
n	5,925	6,111		5,918	6,100	
%	96.1	96.0	0.1	95.9	95.8	0.1
Attrition (%)	3.9	4.0	-0.1	4.1	4.2	-0.1
2014 enrollment sample						
Enrolled in study school, spring 2014	6,500	6,569		6,500	6,569	
Enrolled in study district and had valid test score, spring 2016						
n	5,326	5,450		5,330	5,450	
%	81.9	83.0	-1.1	82.0	83.0	-1.0
Attrition (%)	18.1	17.0	1.1	18.0	17.0	1.0

Source: District administrative data.

We also assessed whether attrition could have led to differences between the treatment and control students with valid outcome data—that is, students included in the main analysis, which would not include those dropped because of attrition. The results from this comparison were similar to those from the comparison of the baseline characteristics of the full treatment and control groups presented above. In particular, there were differences in the racial/ethnic composition of students, but treatment and control students in the analysis sample were similar in sex, special education status, free or reduced-price meal eligibility, baseline achievement, and baseline school attendance (table A.5). This finding suggests that attrition did not cause any imbalance in baseline characteristics between treatment and control groups.

Table A.5. Baseline characteristics of students in the analysis sample (those enrolled in a study school in spring 2016 and with valid test scores)

	Sample with 2016 math scores			Sample with 2016 ELA scores		
	Treatment schools	Control schools	Difference	Treatment schools	Control schools	Difference
Characteristics of students in sample						
Male (proportion)	0.51	0.52	0.00	0.51	0.51	0.00
Race/ethnicity (proportion)						
White	0.48	0.51	-0.03*	0.48	0.51	-0.03*
Black	0.22	0.16	0.06*	0.22	0.16	0.06*
Hispanic	0.19	0.23	-0.03*	0.19	0.23	-0.03*
Other	0.11	0.10	0.01	0.11	0.10	0.01
Other student characteristics (proportion) (n = 7,926-10,432)						
Special education	0.14	0.14	0.00	0.14	0.14	0.00
English language learner	0.12	0.17	-0.05*	0.12	0.17	-0.05*
Eligible for free/reduced-price meals	0.65	0.65	0.00	0.65	0.65	0.00
2014 student achievement (z-score) (n = 3,357-3,365)						
Math	-0.21	-0.17	-0.04	-0.21	-0.17	-0.04
English/language arts	-0.21	-0.22	0.01	-0.21	-0.22	0.01
2014 student attendance rate (mean proportion) (n = 9,549-9,559)	0.96	0.96	0.00	0.96	0.96	0.00
Number of schools	**51**	**51**		**51**	**51**	
Number of students	**5,925**	**6,111**		**5,918**	**6,100**	

Source: District administrative data.

Notes: Unless otherwise noted, data on student characteristics were available for all students and schools in the sample. English/language arts and math test scores from 2014 were available in 70 schools, 2014 free and reduced-price lunch eligibility information was available in 80 schools, and 2014 attendance information was available in 94 schools.

ELA = English/language arts.

If the assumption that the DDI intervention would be unlikely to affect student mobility is correct, we should see little difference in the number of students entering treatment versus control schools from 2014 to 2016. About 24 percent of students enrolled in a study school in 2016 had been enrolled in a different school in 2014 (table A.6). However, the difference in this percentage among students in treatment versus control schools in 2016 was less than one percentage point. Similarly, about 26 percent of students who had been enrolled in a study school in 2014 were enrolled in a different school in spring 2016, but the difference in this percentage among students in treatment versus control schools in 2014 was less than two percentage points.[6]

Table A.6. Student mobility into analysis sample, by treatment group

	Treatment schools	Control schools	Difference
Enrolled in study school, spring 2016	6,168	6,367	
Percentage enrolled in a different school, spring 2014	24	24	0.6
Enrolled in study school, spring 2014	6,500	6,569	
Percentage enrolled in a different school, spring 2016	27	25	1.5

Source: District administrative data.

Note: The percentage enrolled in a different school ignores students who switched between schools in the same experimental group; for example, a student enrolled in treatment school A in 2014 but in treatment school B in 2016 would not be counted as being enrolled in a different school for purposes of this table.

To assess whether our sample definition would affect estimates of the impact of the DDI intervention, we conducted a sensitivity test, in which we estimated impacts using the sample of students enrolled as second- and third-graders in treatment and control schools in spring 2014 (who would be fourth- and fifth-graders in 2015-2016 in these schools if they progressed normally). The estimated impacts were similar to the main impact estimates using the sample of students enrolled in spring 2016. See appendix tables C.8 and C.9 for details. The similarity of impact estimates from these alternative samples and the similar rates of student mobility in treatment and control schools indicate that there is no evidence of impact estimates being driven by changes in the composition of study schools.

Data Collection

We collected three types of data for the study. First, we collected information from data coaches as part of the study team's monitoring of DDI implementation in treatment schools. Second, we conducted surveys of study school principals and teachers to further describe implementation and measure key intermediate outcomes. Third, we obtained student-level administrative data from participating districts to measure student characteristics and achievement.

Program Monitoring Data

Data coaches provided information on DDI implementation through weekly data coach logs and two semi-structured interviews. They were required to complete weekly logs to describe any DDI intervention activities they had completed that week, summarize implementation progress, and note any important challenges they faced. We used data from the logs to determine how often data coaches engaged in specific activities related to DDI, focusing mainly on data coach activities during the 2015-2016 school year.

[6] It is possible that some students who left treatment schools moved to control schools, or vice versa, introducing the possibility of contamination. However, the analysis of impacts for the sample of students enrolled in student schools in 2014 eliminates this possibility (see tables C.8 and C.9 for details of this analysis).

Data coaches at all treatment schools completed logs on at least some weeks of the 2015-2016 school year. Most data coaches completed logs on most weeks, but some did not always do so. We expected that data coaches would complete 36 or more logs, or every week during 2015-2016.[7] Overall, 35 percent of data coaches completed 36 or more logs during that school year (every week), 45 percent completed 27 to 35 logs, 14 percent completed 18 to 26 logs, and 6 percent completed fewer than 18 (less than half the weeks of the year).

Data coaches completed semi-structured interviews in fall 2015 (late September/early October) and spring 2016 (April/May). The interviews provided more detailed information about the implementation of the DDI intervention and data coach attitudes and perceptions about it. Response rates on these interviews were 90 percent and 92 percent, respectively.

Principal and Teacher Surveys

We conducted surveys of the principals and of fourth- and fifth-grade teachers in both treatment and control schools in spring 2016. The surveys provided information about data-related training and professional development they had received, the activities of school leaders in supporting data use in their schools, and the activities of teachers in analyzing and using data to improve their instruction.

Overall, data from the principal and teacher surveys addressed two key objectives. The first was to describe implementation of the DDI intervention, along with the resulting difference in data-related activities between the treatment and control schools. The second was to measure intermediate outcomes that the logic model indicated should have been affected by DDI, including principals' and teachers' data use, and teachers' instructional practices.

Administration of the surveys began in March 2016 and continued until early June 2016. Well over 90 percent of principals and teachers in the sample completed the survey, and there were no significant differences in response rates among those in treatment and control schools. In particular, the overall response rate on the principal survey was 95 percent, including 98 percent among treatment school principals and 92 percent among control school principals. The response rate on the teacher survey was 93 percent, including 96 percent among teachers in treatment schools and 90 percent among those in control schools.

Student Administrative Data

We collected district administrative records on students enrolled in treatment and control schools in spring 2016 (as well as for the alternative sample of those enrolled in spring 2014). Data included information on students' background characteristics and their scores on state assessments in math and English/language arts, covering the 2014-2015 through 2015-2016 school years. We used the data to measure the impact of the DDI intervention on student achievement. We also collected data on student attendance, out-of-school suspensions, and—when possible—in-school suspensions.

---

[7] Because data coaches worked half time in the study's treatment schools, in theory they could maintain a full-time schedule in some weeks but be outside of the school in other weeks. In practice, however, data coaches tended to work at least some hours at a given school during each week.

We collected data collected from each study district, including all fourth- and fifth-grade students in study schools in those districts in spring 2016. As shown in table A.4 above, we received valid test scores that could be included in the analysis for 96 percent of students enrolled in a study school in spring 2016.

Analysis Methods

Standardizing Test Score Outcomes

We measured student achievement in math and English/language arts using student scores on statewide standardized assessments. Because different state tests measured student achievement using different assessments and scales, we used standardized scores (z-scores) to measure achievement on a consistent scale across districts. For each subject and year, we constructed a z-score measure by subtracting the statewide mean score from each student's score and then dividing the result by the statewide standard deviation of student scores. Thus, the z-score measures the number of standard deviations above or below the statewide mean for each student's score in each subject and year.

Estimating Impacts on Student Outcomes

In this section, we describe the statistical models used to estimate the impacts of the DDI intervention on student achievement and behavioral outcomes, as well as on teacher and principal perceptions and practices related to data use, teacher collaboration, and instructional practices. We then discuss the statistical model used to estimate the student achievement impacts within subgroups defined by student grade and prior achievement. Finally, we describe how we analyzed the variation in impacts across study districts and matched pairs of schools.

Impact model. To estimate the impact of DDI on student achievement and student behavior, we used a regression model that reflected the study's random assignment design. The model accounts for the clustered nature of data collected from students enrolled in study schools, as well as for the pairs of similar schools matched before random assignment.

$$(1) \qquad y_{ijk} = \beta T_{jk} + X^{'}_{ijk}\delta + Z^{'}_{jk}\gamma + W^{'}_{k}\pi + \varepsilon_{ijk}$$

In the model, y_{ijk} is the outcome measure (for example, math z-score) for student *i* in school *j* in matched pair *k*, T_{jk} is the treatment indicator that takes the value one if a student was in a treatment school and zero if a student was in a control school, $X^{'}_{ijk}$ is a vector of student-level covariates included in the model, $Z^{'}_{jk}$ is a vector of school-level covariates, and $W^{'}_{k}$ is a vector of matched pair indicators. The terms δ, γ, and π are vectors of coefficients; β is the estimated impact of DDI on the outcome; and ε_{ijk} is the student-specific error term.

Covariates. The model includes student-level covariates that are indicators for grade, gender, race/ethnicity, special education status, and English language learner status, all measured during the 2013-2014 school year. The model also includes the school-grade-level average z-score on the spring 2013 state math and English/language arts assessment, interactions between these two variables and state indicators, and the percentage of students at the school eligible for free or reduced-price meals.[8]

We chose to include 2013 test scores measured at the school level rather than individual sample members' 2014 scores as covariates because the latter were available for only about one-quarter of the sample. There were three reasons for this. First, 3 of the 12 districts in our sample did not field statewide standardized assessments during the 2013-2014 school year because they participated in a pilot test for the Smarter Balanced assessment. Second, students who were fourth-graders in spring 2016 were in second grade in spring 2014 and thus did not take a state test. Finally, some students who were fifth-graders in spring 2016 did not have test scores available from spring 2014 for some other reason, presumably because they were not enrolled in the district at that time. With limited student-level data on baseline achievement, we used school- and grade-level average test scores for students in grades 4 and 5. (We refer to these scores as school-level scores for simplicity.) Because of the issue mentioned above with three districts not having 2014 scores, we used school-level scores from spring 2013.

To examine how the impact estimates might change under alternative approaches to accounting for baseline achievement levels, we conducted a set of sensitivity analyses, with results presented in appendix tables C.8 and C.9. One of these alternative models includes additional student-level covariates that capture students' 2014 math and English/language arts scores, and interactions between those variables and state indicators. This model also includes a student-level indicator for eligibility for free and reduced-price meals, and the student's attendance rate in the 2013-2014 school year. To address the fact that a substantial number of students are missing 2014 test scores and the other student-level covariates included in this sensitivity analysis, we used the dummy adjustment method described below.

Weights. We applied sample weights to the analysis so that each school and grade in the sample contributed equally to the impact estimates, and to ensure that the relative weight of the treatment and control schools within each matched pair was the same. The sample weight for each student observation was calculated as follows:

$$(2) \qquad w_{igj}^{st} = \left(\frac{1}{n_{gj}}\right) * \left(\frac{1}{2J}\right) \sum_{j=1}^{J} \sum_{g=4}^{5} n_{gj}$$

where the index j (= 1,…,J) represents a student's school, g (= 4 or 5) represents the student's grade, and i represents the student. There are sample sizes of n_{gj} students (with valid data for the outcome) within grade g of school j.

[8] Specifically, each school-grade-level variable measures the average 2013 z-score among fourth-graders in each school in 2013 and fifth-graders in each school in 2013; thus, it is a measure of school grade-level achievement among prior cohorts of students. For example, for a student in the analysis sample who was in fourth grade in spring 2016, this variable takes the value of the fourth-grade average test score in 2013 at the school in which they were enrolled in 2016. For a student enrolled in fifth grade in spring 2016, this variable takes the value of the fifth-grade average test score in 2013.

Missing data. We imputed missing values for the model's covariates based on other information we collected from the student so that he or she could be included in the impact analysis. The imputation procedure, known as multiple imputation by chained equations, used nonmissing values of baseline covariates and outcomes to estimate values for observations with missing baseline data. The method first generated multiple datasets with imputed values for missing values. We calculated a separate impact estimate with data from each of the imputed datasets and combined these impact estimates using procedures described in Rubin (1987) that accounted for the variability of estimates calculated using the different imputed datasets. We adjusted the standard error of each combined impact estimate to reflect this variability. Within each imputed dataset, we estimated the standard errors of coefficient estimates using the Huber-White sandwich estimator (Huber 1967) to account for the clustering of students' outcomes within schools by allowing correlation in errors within blocks of observations defined by school. As shown in table A.7 below, only two covariates in the main model had missing data.

Table A.7. Rates of missing data on model covariates in the main impact model and sensitivity model with additional covariates

	Percentage of observations missing data in sample with 2016 math or ELA scores		
	Treatment schools	Control schools	Difference
Covariate			
Male	0	0	0
Race/ethnicity			
White	0	0	0
Black	0	0	0
Hispanic	0	0	0
Other	0	0	0
Other student characteristics			
Special education	13	13	0
English language learner	17	17	0
School characteristics			
Percentage eligible for free/reduced-price meals	0	0	0
2013 student achievement (school mean z-score)			
Math	0	0	0
English/language arts	0	0	0
Covariates included only in sensitivity analyses			
2014 student math achievement	69	70	-1
Block missing	66	66	0
2014 student ELA achievement	69	70	-1
Block missing	66	66	0
2014 student FRL participation	34	35	0
Block missing	25	25	0
2014 attendance rate	19	20	-1
Block missing	8	8	0
Number of schools	**51**	**51**	
Number of students	**5,936**	**6,131**	

Table A.7. ***(continued)***

Source: District administrative data and Common Core of Data

Notes: Proportions of missing data were the same within the math analysis sample as they were within the reading sample. Rows labeled "block missing" present the percentage of the student sample that had missing data for the entire grade, school, or district, for example, because students who were 4th graders in spring 2016 were not in a tested grade in spring 2014, or because a whole district did not provide attendance information. Numbers in the difference column may not exactly equal the difference between the rounded proportions presented in the treatment and control columns.

ELA = English/language arts; FRL = free or reduced-price lunch.

Imputation was conducted separately by treatment group, and all baseline characteristics included as covariates in the main impact model were included in the imputation model. The predictor set in the imputation model also included the student achievement and behavior outcome measures and indicators for randomization blocks, the school matched pairs. Finally, we imputed values only for model covariates with missing data, not for outcome measures. Although we used these imputed baseline covariates in our analysis of impacts of DDI, we did not include any of the imputed values in the tests of baseline equivalence. For the analysis of baseline equivalence, we simply treated students missing data on a given variable as being missing from the sample.

To estimate the sensitivity of the impact estimates to models that included student characteristics missing for a large proportion of sample members (student-level test score data, student free/reduced-price meals eligibility, and student attendance rates from 2014), we used the dummy adjustment approach (Puma et al. 2009). In this approach, missing values are assigned a single value (for example, the mean among nonmissing values of that variable), and the impact model includes a binary indicator (or dummy variable) for each of these variables that takes on a value of one for these missing values and zero for nonmissing values.[9] We used this approach for students whose data was missing due to factors that affected all students within a given grade, school or district, referred to as block-level factors. For example, in some districts all students' free or reduced-price lunch participation information was missing because the district did not provide it for any students. There would have been little information to form the basis for multiple imputation of missing values in these situations.

Subgroup model. Teachers receiving the DDI intervention may change instructional practices differently for different groups of students, particularly on the basis of student achievement levels. To explore this hypothesis, we estimated the impacts of the DDI intervention on student achievement within two sets of student subgroups—one set defined by students' spring 2016 enrolled grade and the other based on their spring 2015 test scores.[10] Specifically,

[9] Even for these covariates with a large amount of missing data, we used regression imputation to impute the values for students who were not part of a "block missing" pattern, for example students who were missing attendance information but attended a school and grade from which we did receive attendance information for a large majority of students.

[10] Technically, a student's 2015 test scores were not measured at baseline and so could have been influenced by the intervention, which began in December 2014. Thus, spring 2015 test scores were not ideal for defining subgroups. However, no true baseline achievement scores were available for about three-fourths of students in the sample. We also believed that the DDI intervention would not be as likely to substantially affect student achievement as measures on their spring 2015 assessments because of the nature of the intervention. In particular, it was complex,

we estimated impacts separately among fourth- and fifth-graders, and separately among students who had scored below the state proficiency threshold in 2015 in the same subject as the outcome measure in 2015, those at or above this threshold but below the “advanced” threshold, and those at or above the advanced threshold.

To estimate impacts within each of these subgroups, as well as the difference in impacts between subgroups, we used the following model:

$$(3) \quad \begin{aligned} y_{ijk} = {} & \beta_1 T_j + \theta_2 Group2_{ijk} + \theta_3 Group3_{ijk} + \beta_2 \left(T_j \times Group2_{ijk}\right) \\ & + \beta_3 \left(T_j \times Group3_{ijk}\right) + X'_{ijk}\delta + Z'_{jk}\gamma + W'_k \pi + \varepsilon_{ijk} \end{aligned}$$

In equation (3), the impact of DDI within groups 1, 2, and 3 (for example, below proficient, proficient, and advanced) is represented by β_1, $(\beta_1 + \beta_2)$, and $(\beta_1 + \beta_3)$, respectively. All other variables in equation (3) are the same as those defined in equation (1). We tested the statistical significance of the coefficients β_2 and β_3 to assess whether impacts within the second and third subgroups differed significantly from the impact in the first group. When estimating impacts by student grade level, our approach was identical to equation (3) except that the model did not include the indicator and interaction associated with $Group3_{ijk}$.

Estimating Impacts on Teacher and Principal Outcomes

Main estimation model. To estimate the impact of the DDI intervention on teacher and principal outcomes, we used a regression model similar to the one used with student outcomes. The model accounts for the clustered nature of data collected from teachers within study schools, as well as for the pairs of similar schools matched before random assignment. Because there was only one principal per school, there was no clustering in the model used to estimate impacts on principal outcomes. This model differs from the student model in that the unit of observation is the teacher (or principal) rather than the student. In addition, the model included no additional covariates measuring the characteristics of schools (or teachers or principals).[11]

$$(4) \quad y_{ijk} = \beta T_{jk} + W'_k \pi + \varepsilon_{ijk}$$

and many of the early activities required in the DDI intervention involved setting up and initiating school- and grade-level structures rather than immediately making changes to teachers’ instructional practices.

[11] In the case of the principal model, we could not include covariates because with the matched pair dummy variables and the treatment status variable, we could not identify the effects of additional school-level characteristics on the outcome. With just one observation per school in this model, there would be no independent variation in the covariate to allow us to determine its effect. For the teacher model, there were typically two or more teachers (observations) per school so we could identify the effects of covariates that varied within schools, such as teacher characteristics. However, we did not believe that such within-school variation in teacher characteristics would explain much of the variation in the outcomes we examined, so we excluded these covariates.

In equation (4), where y_{ijk} is the outcome measure, such as whether a teacher frequently used formative assessment data to guide instructional decisions, for teacher (or principal) *i* in school *j* in matched pair *k*, T_{jk} is the treatment indicator that takes the value one if a teacher is in a treatment school and zero if in a control school, and $W_k^{'}$ is the vector of matched pair indicators. The term π is a vector of coefficients, β is the estimated impact of DDI on the outcome, and ε_{ijk} is the teacher- or principal-specific error term. Standard error estimates account for the clustering of teacher observations within schools using the Huber-White sandwich estimator described above. This method allows correlation of teacher error terms within each school.

Missing data. When all teacher observations within a school were missing for a given outcome (that is, we had no teacher outcome data from that school), we excluded both that school and its matched treatment- or control-group school from the analysis sample. Similarly, when a principal observation was missing for a given outcome, we excluded both that observation and the matched treatment- or control-group principal observation. Because the statistical models used to estimate principal and teacher impacts did not include baseline covariates, we did not need to adopt a strategy to address missing baseline data.

Weights. We constructed a single set of teacher weights to apply across all outcomes. The weights were designed so that each school in the sample contributed equally to the impact estimates, regardless of the number of teachers in the school, and to ensure that the relative weight going to the treatment and control school was the same within each matched pair. We calculated the weight as follows:

$$w_{ij}^{tch} = \left(\frac{1}{I_j}\right) * \left(\frac{1}{J}\right) \sum_{j=1}^{J} I_j \tag{5}$$

where *i* indexes teachers and *j* indexes schools. Because there was only one principal per school, we did not need to construct weights to achieve the goal of each school in the sample contributing equally to estimated impacts.

Analysis of Variation in Impacts

This section describes how we measured the level of variation in impacts of the DDI intervention on student achievement across several groups. First, we describe estimating the variation in impacts across districts and testing the statistical significance of that variation. We then describe estimating and testing the significance of variation in impacts across matched pairs of schools. Finally, we describe how we measured each treatment school's readiness to implement the DDI intervention and how that readiness correlated with the school's impacts on student achievement.

Estimating variation in impacts across study districts. To estimate the impact of the DDI intervention in each study district, we used a regression model similar to the one presented in equation (1):

$$(6) \qquad y_{ijk} = \sum_{d=1}^{12} \beta_d \left(T_j \times I_j^{(d)} \right) + X'_{ijk}\delta + Z'_{jk}\gamma + W'_k \pi + \varepsilon_{ijk}$$

where $I_j^{(d)}$ is a set of 12 binary variables indicating the district (d) in which a student's school was located and β_d is a set of 12 district-specific impact estimates. The remaining variables are the same as in the student impact model, with the exception that the pre-baseline math and reading tests were not interacted with a series of state indicator variables as they were in the main student impact model. We used an *F*-test of the joint equality of the district-specific impact estimates to determine whether the variation in impacts across districts was statistically significant.

Estimating impacts within matched pairs. To estimate the impact of the DDI intervention in each matched pair of schools in the study, we estimated the following model:

$$(7) \qquad y_{ijk} = \sum_{k=1}^{51} \beta_k \left(T_j \times I_j^{(k)} \right) + X'_{ijk}\delta + W'_k \pi + \varepsilon_{ijk}$$

Equation (7) differs from equation (6) in two ways. First, $I_j^{(k)}$ indicates a school's matched pair rather than district, and β_k is a set of 51 matched pair-specific impact estimates. Second, the model does not include any school-level covariates because there was no variation in values of those variables independent of the matched pair indicators (W'_k) and matched pair-treatment interactions $(T_j \times I_j^{(k)})$.[12] To assess whether the variation in impacts across matched pairs was likely due to chance alone, we also estimated the following hierarchical linear model:

$$(8) \qquad y_{ijk} = \alpha_0 + \beta_1 T_j + \eta_{0k} + \eta_{1k} T_j + X'_{ijk}\delta + Z'_{jk}\gamma + I'_d \omega + \varepsilon_{ijk}$$

where η_{0k} is a random intercept at the matched-pair level, η_{1k} is a random slope on the treatment indicator that varies at the matched-pair level, $I'_d\omega$ is a set of district indicators and vector of associated coefficients, and the remaining variables are the same as in equation (1). We estimated the variation in impacts across matched pairs as the standard deviation of the random treatment coefficient. We then used the standard error of the estimated standard deviation to construct a 95 percent confidence interval to assess whether variation in impacts appears to be due to chance alone.

Examining the correlation between estimated impacts and treatment school readiness to implement DDI. After estimating DDI impacts for each treatment school (that is, within each matched pair), we examined the correlation between a school's estimated impact and its readiness to implement the DDI intervention at the beginning of the study. We measured

---

[12] Because the construction of matched pairs was based in part on school-level prior math and reading proficiency levels and levels of free/reduced-price participation, removing the school-level variables does not have a substantive effect on impact estimates from this model.

readiness to implement DDI using data from the data coach interviews conducted near the beginning of the fall 2015 semester, in which they described the presence or absence of specific structures and practices in treatment schools before December 2014, when DDI implementation began.

To measure schools' DDI readiness, we first created indicators for four aspects of readiness before December 2014: (1) whether the school had an ILT in place, (2) whether the school had teacher collaboration teams in place, (3) whether teacher collaboration teams worked with student data "almost every meeting" or "occasionally" (versus "rarely" or "never"), and (4) whether the school had data walls in place. We then used these indicators to create a binary measure of DDI readiness for the 41 treatment schools in which the coach had responded to at least three of the four items mentioned above. We coded the DDI readiness measure as one for any matched pair in which more than half of these readiness elements were in place before December 2014 and as zero otherwise. We assigned the measure for a given treatment school to its matched pair. Based on this definition, 27 percent of matched pairs (11 out of 41) were in the high-readiness group and the remainder in the low-readiness group.

We also tested the sensitivity of our results using an alternative measure of schools' DDI readiness that aimed to address a potential weakness of the above measure, which did not distinguish between the four aspects of readiness in defining the school's overall readiness. The alternative readiness measure is based on the idea that because most of the on-the-ground work of the DDI intervention involved the teacher collaboration teams working with data, the alternative measure gave priority to those aspects of readiness. To be defined as having high readiness to implement DDI based on this measure, schools were required to have teacher collaboration teams that worked with student data at least occasionally, plus at least one of the other two aspects of readiness (an ILT and/or data walls) before December 2014. This alternative measure defined schools with a teacher collaboration team in place that rarely or never worked with student data as having low readiness for DDI. The rationale for this definition was that these teams would have procedures and routines in place before the intervention that did not involve working with data, so the intervention staff would need to spend time and resources getting the team to change these procedures and routines before they could proceed to other aspects of DDI. Based on this definition, 20 percent of matched pairs (8 out of 41) were in the high-readiness group and the remainder in the low-readiness group.

We used a model similar to the subgroup impact model in equation (3) to estimate the impacts of DDI on student achievement within each subgroup of matched pairs defined by DDI readiness. The only difference from equation (3), apart from the existence of only two subgroups rather than three, was that the subgroup membership indicators were not explicitly included in the model because the matched-pair indicators $W_k^{'}$ fully accounted for their DDI readiness subgroup.

One important limitation of this analysis is that we did not measure the structures and practices underlying the DDI readiness indicator in control group schools. The result was the possibility that treatment schools with high DDI readiness may also have had other attributes that led to higher student achievement as compared with the control school to which they were matched. If so, then differences in impacts between high- and low-readiness schools may have been caused by these other positive school attributes in addition to any effect of DDI readiness.

Our analysis addresses this limitation by controlling for 2013 school-grade-level average math and reading test scores as well as the rest of the baseline covariates described in equation (1), but this analytic approach does not completely remove the possibility that differences in impacts are driven by school characteristics other than DDI readiness.

APPENDIX B

SUPPLEMENTAL FINDINGS ON IMPLEMENTATION OF THE DATA-DRIVEN INSTRUCTION INTERVENTION (CHAPTER III)

This appendix supplements information presented in Chapter III on implementation of the intervention. That chapter described evidence on implementation we obtained from staff in the study's treatment schools, as well as treatment-control differences in data-related activities.

Implementation of Support for DDI in Treatment Schools

Key Intervention Supports and Provider Services

This section provides additional details on the data coaches placed in treatment schools by the intervention, as well as their participation in the intervention's professional development sessions. The data coaches' background characteristics are presented in table B.1 and the extent to which they and principals remained at treatment schools over the course of the evaluation period are presented in figure B.1. Tables B.2 and B.3 show data coach and principal attendance at professional development sessions and the coaches' assessment of the usefulness of the training they received under the intervention.

Table B.1. Data coaches' background

Prior years of experience	Fall 2015
Years as educator (mean)	20.6
Positions held as educator (percentages)	
Teacher	100
Data coach	<7
Instructional coach or department head	36
Principal	16
Assistant principal	7
District administrator	11
Other	16
Years of experience working with student data (mean)	12.5
Years worked in assigned district (mean)	13.6
Number of coaches	**45**

Source: Coach interviews
< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics statistical standards, but that the percentage is less than or greater than the number following the < or > symbol (U.S. Department of Education 2012).

Figure B.1. Turnover among key staff in treatment schools

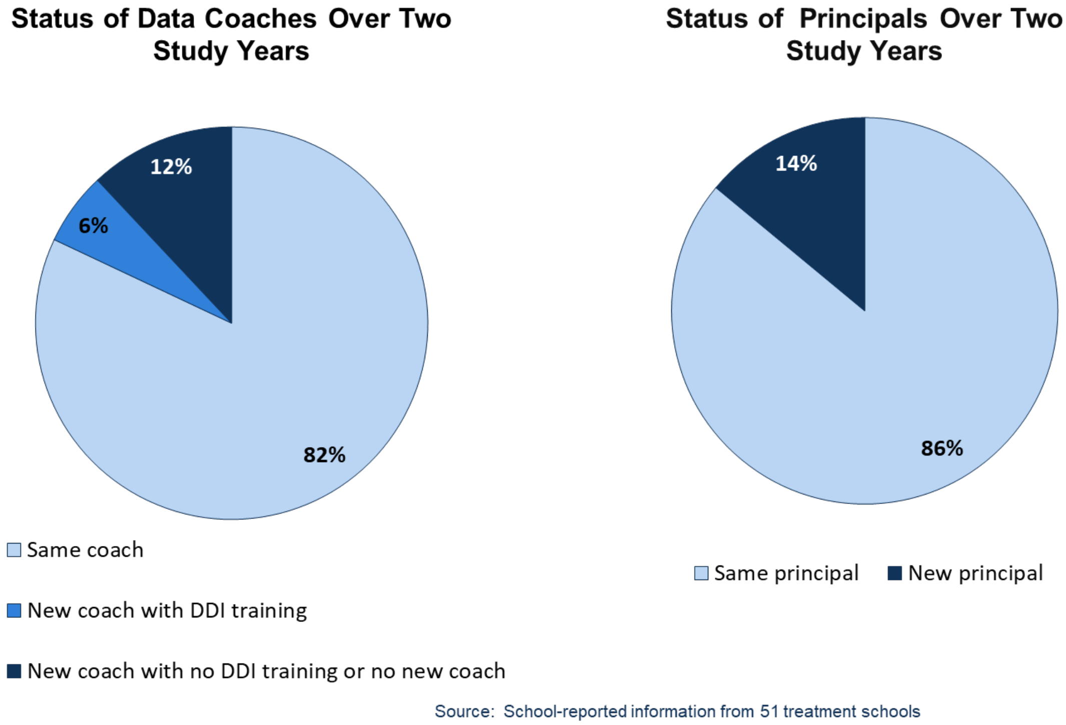


Table B.2. Attendance at and ratings of professional development (PD) sessions

PD session number	Percentage of coaches attending	Percentage of principals attending	Average participant rating of session quality (1-5)
1	>94	89	4.78
2	>94	87	4.71
3	88	79	4.71
4	87	73	4.77
5	>94	77	4.74
6	89	75	4.64
Average	**92**	**80**	**4.72**

Source: Coach interviews for 51 treatment schools

Note: Attendance data available from all 12 districts for sessions 2 and 3; from 11 districts for sessions 1, 4, and 6; and from 10 districts for session 5. Data on participants' ratings of session quality are available from 12 districts for session 3; from 11 districts for sessions 1, 2, 4, and 6; and from 10 districts for session 5.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics statistical standards, but that the percentage is less than or greater than the number following the < or > symbol (U.S. Department of Education 2012).

Table B.3. Data coach assessment of usefulness of data-driven instruction training

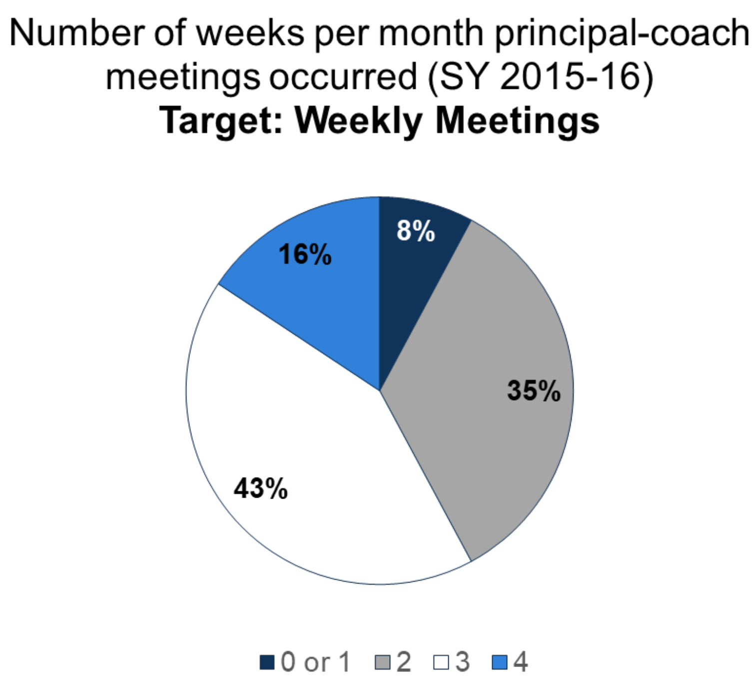
	Spring 2016
Degree to which training helped prepare coach for role (Percentages)	
Not at all prepared	0
Prepared to carry out some but not all tasks	21.7
Prepared to carry out most or all tasks	78.3
Number of coaches	**46**

Source: Coach interviews

Intervention Activities in Treatment Schools

This section provides additional details on school level activities implemented in the treatment schools, including the frequency of specific activities and perceptions of principal support. Figure B.2 shows the frequency of principal-data coach meetings and table B.4 shows data coaches' assessment of principal support for DDI. Table B.5 lists the instructional focus and best practices chosen by instructional leadership teams. Figure B.3 displays the frequency of teacher collaboration team meetings in grades 4 and 5 during the 2015-16 school year. Table B.6 shows challenges reported by data coaches related to working with instructional leadership teams and teacher collaboration teams.

Figure B.2. Frequency of meetings between principal and data coach, 2015-2016



Source: Coach logs for 51 treatment schools

Table B.4. Data coach assessment of principal's level of support of data-driven instruction

	Spring 2016
Data coach rating of principal support	
1 (low support)	0
2	<6.5
3	13.0
4	10.9
5 (high support)	71.7
Mean rating	4.5
Number of coaches	**46**

Source: Coach interviews

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics statistical standards, but that the percentage is less than or greater than the number following the < or > symbol (U.S. Department of Education 2012).

Table B.5. Instructional focus and best practices chosen by instructional leadership teams

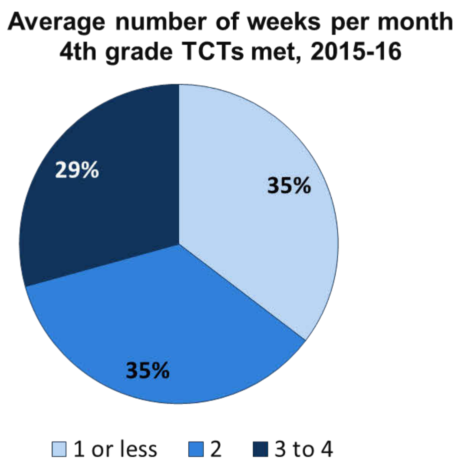
Instructional Focus	# of Schools	Instructional Best Practices Used by One or More Schools
Reading	16	Close reading (across subjects)
		Engagement and total participation
		Focus on comprehension and fluency in content areas
		Guided reading
		Smart E goals and differentiated instruction
		Text-based evidence and differentiation
		Think, Pair, Share, Write
		Use formative assessment to guide reading instruction
		Workshop model
		None identified
Math or Math and Reading	10	Close reading, numbers and operations, student engagement
		Collaborative conversations and math talks
		Communicate / demonstrate understanding in writing and orally
		Differentiated instruction / use of formative assessment
		Intentional talks
		Math fact fluency and differentiated instruction
		Math mindsets pedagogy
		Workshop model
Reading and Writing	4	Close reading (across subjects)
		Workshop Model and differentiation
Writing	3	Workshop Model
		None identified
Vocabulary	4	Direct instruction
		High Yield Effect strategies for vocabulary
		John Hattie classroom discussion strategies
		Word walls, concept maps, Accelerated Reader, word sorts
Academic Vocabulary	3	John Hattie classroom discussion strategies
		Teach academic vocabulary
		Workshop Model

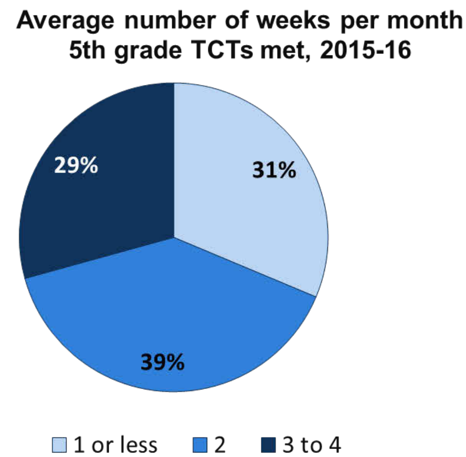
Table B.5. *(continued)*

Instructional Focus	# of Schools	Instructional Best Practices Used by One or More Schools
Differentiated instruction or Improve scores		Gradual release of responsibility
		Small group differentiation to increase student engagement
		Use differentiated instruction to integrate art into al subjects
		None identified
None	6	Active engagement
		Three-tiered layered intervention model with differentiation
		Workshop model
		None identified
Total	**50**	

Source: Coach interviews

Figure B.3. Frequency of teacher collaboration team meetings, 2015-2016





Source: Coach logs for 51 treatment schools

Table B.6. Percentage of coaches that reported challenges with teams during 2015-2016

	Fall 2015	Spring 2016
Challenges related to instructional leadership teams		
Consistent participation and attendance	17.1%	37.8%
Ensuring all opinions are heard	17.1%	17.8%
Maintaining focus on data and instruction	34.1%	31.1%
Balancing DDI work with other school priorities and responsibilities	48.8%	35.6%
Follow through on decisions	39.0%	37.8%
Challenges related to teacher collaboration teams		
Consistent participation	48.8%	40.9%
Establishing meeting norms	19.5%	18.2%
Maintaining focus on data and instructional improvement	48.8%	31.8%
Balancing DDI work with other work and priorities	65.9%	52.3%
Follow through on decisions	41.5%	56.8%

Source: Data coach interviews (n = 44-46)

Differences between Treatment and Control Schools in Data-Related Activities

This section provides additional details on the differences between treatment and control schools in the extent to which they engaged in data-related activities such as professional development for teachers, activities of instructional leadership teams and school level supports for teachers on data use. Tables B.7 displays differences between treatment and control schools on professional development and the topics of the professional development for teachers. Table B.8 compares treatment and control school principal reports of the types of activities of their instructional leadership team. Figure B.4 compares the guidance principals give teachers on data-drive instruction between treatment and control schools. Figure B.5 displays treatment-control differences in the percentage of teachers working one on one with a coach or school leader. Table B.9 shows treatment-control differences in teacher reports of data-related guidance from school leaders. Differences in school leadership planning for training are displayed in figure B.6 and teacher reports of receiving training are in figure B.7. Figure B.8 compares teacher collaboration topics between treatment and control schools.

Table B.7. Teacher professional development during 2015-2016, by teacher type

All Teachers	Treatment	Control	Difference	*p*-value
Receipt of Professional Development				
Received any professional development (percentages)	97.5	>97.7	<-0.2	0.19
Total hours of professional development over school year	38.4	34.5	4.0	0.07
Received professional development focused on topics related to how to use and analyze data to inform instructional practices (percentages)	85.7	71.1	14.6*	0.00
Hours of professional development focused on topics related to how to use and analyze data to inform instructional practices	12.6	8.3	4.3*	0.00
Professional Development Topics (percentages who reported topic was a major focus)				
How to analyze or interpret various types of student data to understand student needs	29.3	14.9	14.4*	0.00
How to use data to set individual learning goals for students	29.0	17.0	12.0*	0.00
How to change instruction based on data	22.2	15.5	6.7*	0.04
How to use student data to monitor student progress toward meeting learning goals	32.5	13.7	18.8*	0.00
How to use evidence-based instructional strategies to help students meet learning goals	23.1	16.1	7.0	0.06
Number of Teachers—Range[a]	214-219	212-214	426-433	0.00

Grade-Level Chairs	Treatment	Control	Difference	*p*-value
Receipt of Professional Development				
Received any professional development (percentages)	100.0	100.0	0.0	1.00
Total hours of professional development over school year	37.7	44.0	-6.3	0.70
Received professional development focused on topics related to how to use and analyze data to inform instructional practices (percentages)	92.6	77.4	15.2*	0.02

Table B.7. *(continued)*

Grade-Level Chairs	Treatment	Control	Difference	*p*-value
Hours of professional development focused on topics related to how to use and analyze data to inform instructional practices	17.2	12.8	4.4	0.21
Professional Development Topics (percentages who reported topic was a major focus)				
How to analyze or interpret various types of student data to understand student needs	39.2	18.5	20.7	0.05
How to use data to set individual learning goals for students	42.7	11.6	31.1*	0.00
How to change instruction based on data	36.6	22.6	14.1	0.11
How to use student data to monitor student progress toward meeting learning goals	47.7	11.3	36.3*	0.00
How to use evidence-based instructional strategies to help students meet learning goals	38.3	18.5	19.8*	0.01
Number of Grade-Level Chairs—Range[a]	40-40	43-44	83-84	

Other Teachers	Treatment	Control	Difference	*p*-value
Receipt of Professional Development				
Received any professional development (percentages)	96.4	>97.7	<-1.3	0.07
Total hours of professional development over school year	39.2	32.0	7.2*	0.00
Received professional development focused on topics related to how to use and analyze data to inform instructional practices (percentages)	82.6	72.7	9.9*	0.01
Hours of professional development focused on topics related to how to use and analyze data to inform instructional practices	11.4	7.4	4.0*	0.00
Professional Development Topics (percentages who reported topic was a major focus)				
How to analyze or interpret various types of student data to understand student needs	26.9	14.6	12.3*	0.00
How to use data to set individual learning goals for students	26.7	18.1	8.6	0.09
How to change instruction based on data	19.2	14.6	4.6	0.25
How to use student data to monitor student progress toward meeting learning goals	28.7	13.6	15.1*	0.00
How to use evidence-based instructional strategies to help students meet learning goals	20.4	15.6	4.8	0.25
Number of Other Teachers—Range[a]	153-157	133-135	286-292	

Source: Teacher survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

[a] Sample sizes are presented as a range based on the data available for each row in the table.

* Difference is statistically significant at the .05 level, two-tailed test.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics statistical standards, but that the percentage is less than or greater than the number following the < or > symbol (U.S. Department of Education 2012).

Table B.8. Instructional leadership team activities as reported by principals during 2015-2016

Activities	Treatment	Control	Difference	p-value
	Percentage reporting at least monthly or several times per term			
Set and monitor instructional focus and instructional best practices				
Identify evidence-based instructional strategies (best practices) that teachers should use	85.1	51.1	34.0*	0.00
Provide guidance on evidence-based instructional strategies (best practices) teachers should use	78.7	51.1	27.7*	0.00
Monitor and provide feedback to teachers on their implementation of evidence-based instructional strategies (best practices)	70.2	40.4	29.8*	0.00
Set schoolwide achievement growth goals and monitor progress				
Analyze student data to set schoolwide and grade-level achievement or proficiency improvement goals	72.3	16.8	25.5*	0.00
Provide guidance to teachers on achievement or proficiency improvement goals for their students	87.2	57.4	29.8*	0.00
Analyze student data to monitor progress toward achievement or proficiency improvement goals	85.1	59.6	25.5*	0.01
Provide feedback to teachers on students' progress toward meeting achievement improvement goals	70.2	51.1	19.1*	0.05
Communicate importance of data-driven culture				
Provide guidance on how often teachers should examine student data	80.9	39.3	42.6*	0.00
Provide guidance on the types of student data that teachers should examine	52.6	34.8	47.8*	0.00
Provide guidance on the protocols or strategies that teachers should use to analyze student data	74.5	34.0	40.4*	0.00
Monitor and provide feedback to teachers on their use of data to guide instruction	68.1	42.6	25.5*	0.01
Lead or plan professional development for teachers				
Analyze student data to determine the professional development needs of teachers	66.0	51.1	14.9	0.18
Plan other structured supports for teachers (for example, coaching) on data use	42.3	36.2	36.2*	0.00
Develop or plan professional development for teachers on data use	83.0	44.7	38.3*	0.00
Designate time for teachers to work collaboratively to plan instruction based on data	80.9	57.4	23.4*	0.00
Number of Principals—Range[a]	46-47	46-47	92-94	

Source: Principal survey.

Notes: Treatment-control differences may not equal the impact shown in the table because of rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Difference is statistically significant at the .05 level, two-tailed test.

Figure B.4. Percentage of principals providing guidance to teachers on aspects of data-driven instruction at least monthly or several times per term during 2015-2016

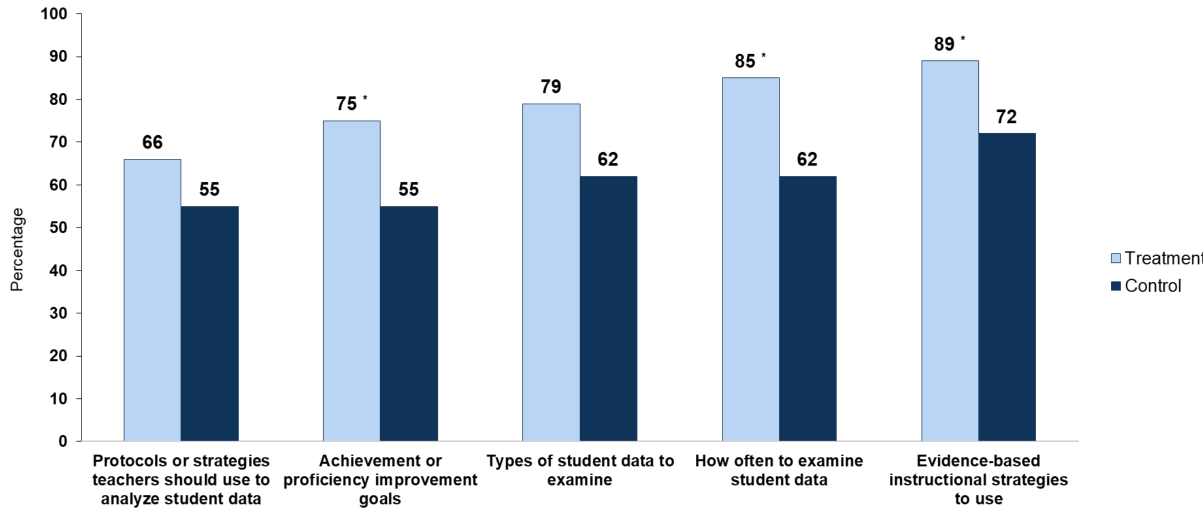
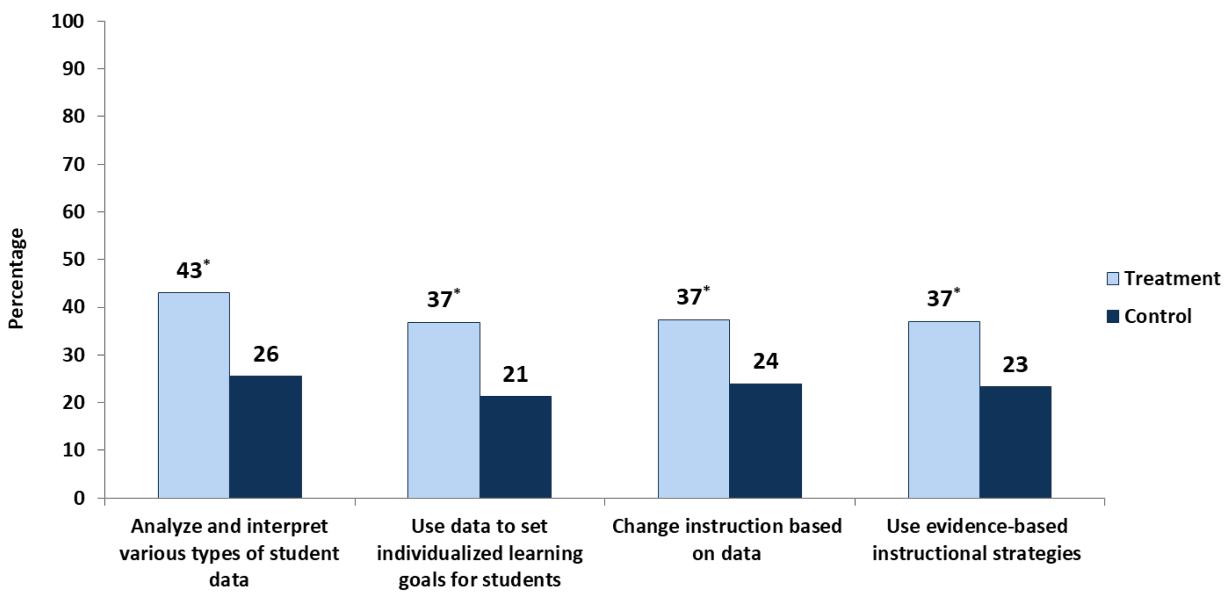


Figure B.5. Percentage of teachers who reported working one on one with a coach or school leader on different aspects of data use at least weekly or several times per month during 2015-2016



Source: Teacher survey (n = 432–433).

*Difference is statistically significant at the .05 level, two-tailed test.

Table B.9. Percentage of teachers who reported guidance or feedback from school leaders on data-related topics at least weekly or several times per month during 2015-2016

	Treatment	Control	Difference	*p*-value
Guidance on:				
How often to examine student data	56.7	33.9	22.8*	0.00
Types of student data to examine	46.8	30.0	16.8*	0.00
Protocols or strategies to use to analyze student data	37.8	22.9	14.9*	0.00
Achievement or proficiency improvement goals for students	35.6	23.3	12.3*	0.00
Evidence-based instructional strategies (best practices) to use	34.0	20.0	13.8*	0.00
Feedback on:				
Data analysis and use of data to guide instruction	32.3	15.8	16.5*	0.00
Implementation of evidence-based instructional strategies (best practices)	28.9	17.2	11.7*	0.01
Student progress toward meeting achievement of proficiency improvement goals	29.0	19.0	10.0*	0.01
Number of teachers (range)[a]	218-219	214-215	432-434	

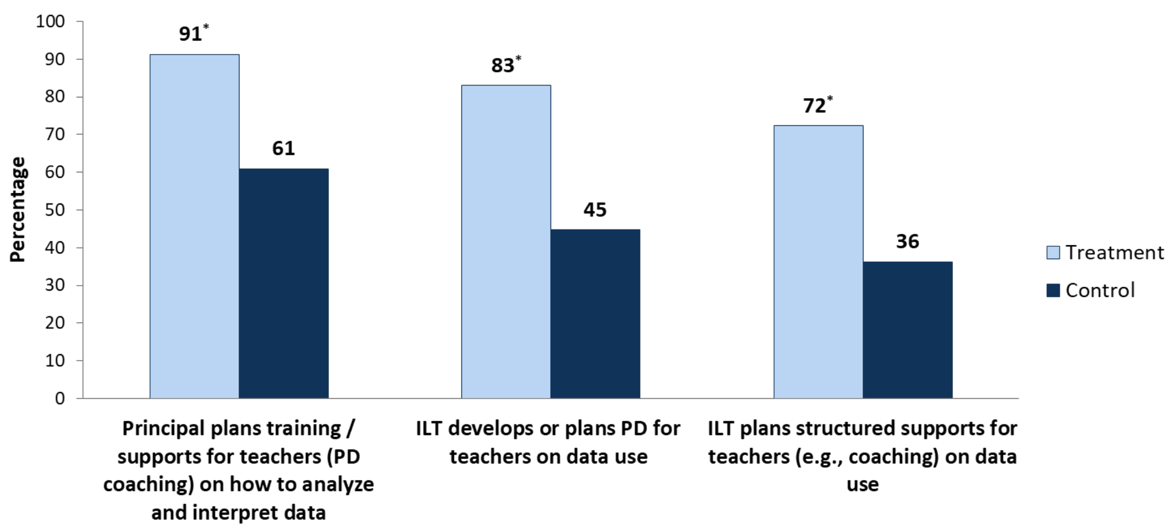
Source: Teacher survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

[a]Sample sizes are presented as a range based on the data available for each row in the table.

*Treatment-control difference is statistically significant at the .05 level, two-tailed test.

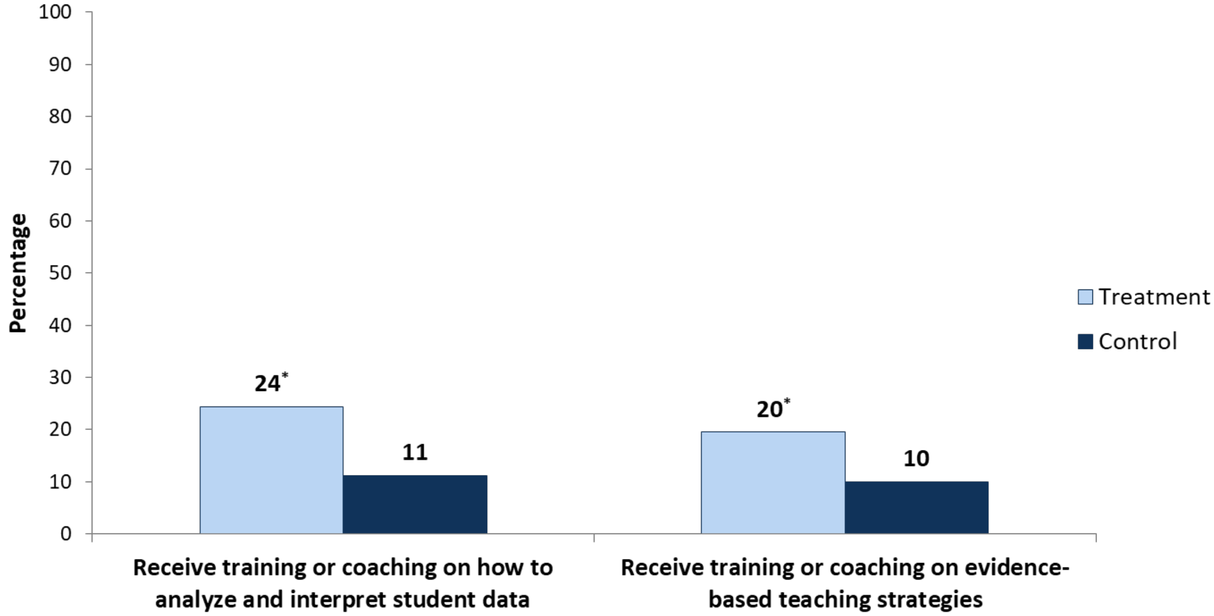
Figure B.6. Planning by School Leaders of Data-Related Training or Structured Supports for Teachers during 2015-2016



Source: Principal survey (N=92-94)

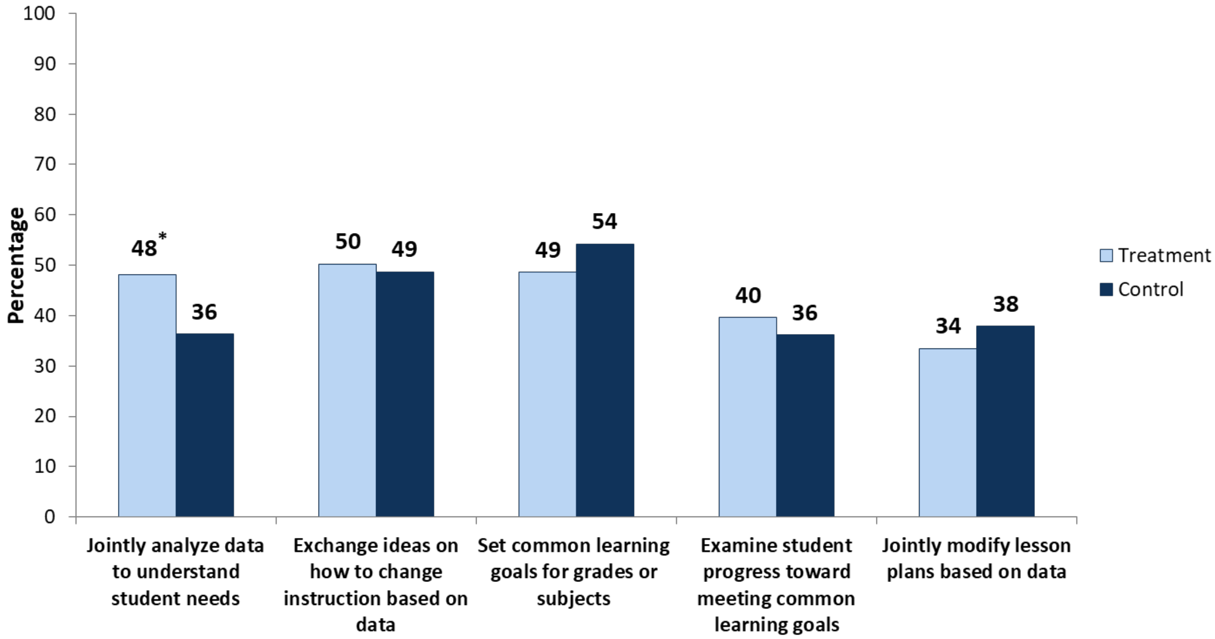
*Difference is statistically significant at the .05 level, two-tailed test.

Figure B.7. Percentage of teachers who reported receiving training or coaching on data-related activities in collaboration with other teachers during common planning periods at least weekly or several times per month during 2015-2016



Source: Teacher survey (n = 421-422).
*Difference is statistically significant at the .05 level, two-tailed test.

Figure B.8. Percentage of teachers who reported collaboration with other teachers during common planning periods on data-related activities at least weekly or several times per month during 2015-2016



Source: Teacher survey (n = 421-422).
*Difference is statistically significant at the .05 level, two-tailed test.

APPENDIX C

SUPPLEMENTAL FINDINGS ON THE IMPACTS OF THE DATA-DRIVEN INSTRUCTION INTERVENTION (CHAPTER IV)

This appendix supplements information presented in Chapter IV on the impacts of the data-driven instruction (DDI) intervention. That chapter described evidence of the impacts on intermediate outcomes reported by principals and teachers, as well as on student outcomes obtained through administrative data (see appendix A for details about the analytical approach).

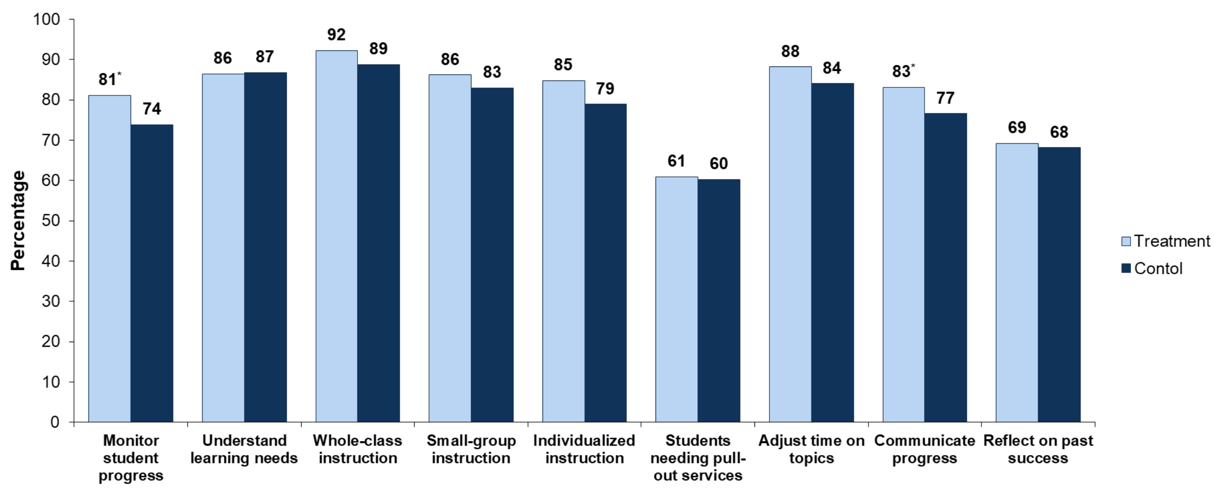
Impacts of the Support for DDI intervention on Intermediate Outcomes

Teachers' Access to and Use of Data

This section provides additional details on the estimated impacts of the DDI intervention on teachers' access to and use of data to improve their instruction. Figures C.1 and C.2 report on the impacts of DDI on teachers' data use using a different cutoff for what constitutes regular data use. In Chapter IV, we required teachers to report using data at least several times per week for a given practice in order to consider the teacher as having engaged in that data use practice. Figures C.1 and C.2 consider a teacher to have engaged in that practice if they at least several times per month.

Tables C.1 through C.4 cover impacts on teachers' perceptions of their access to data as well as the barriers to their access to and use of data, including access to data (table C.1.), the types of data they use (table C.2), the degree to which they find the data useful (table C.3), and their confidence in using data (table C.4).

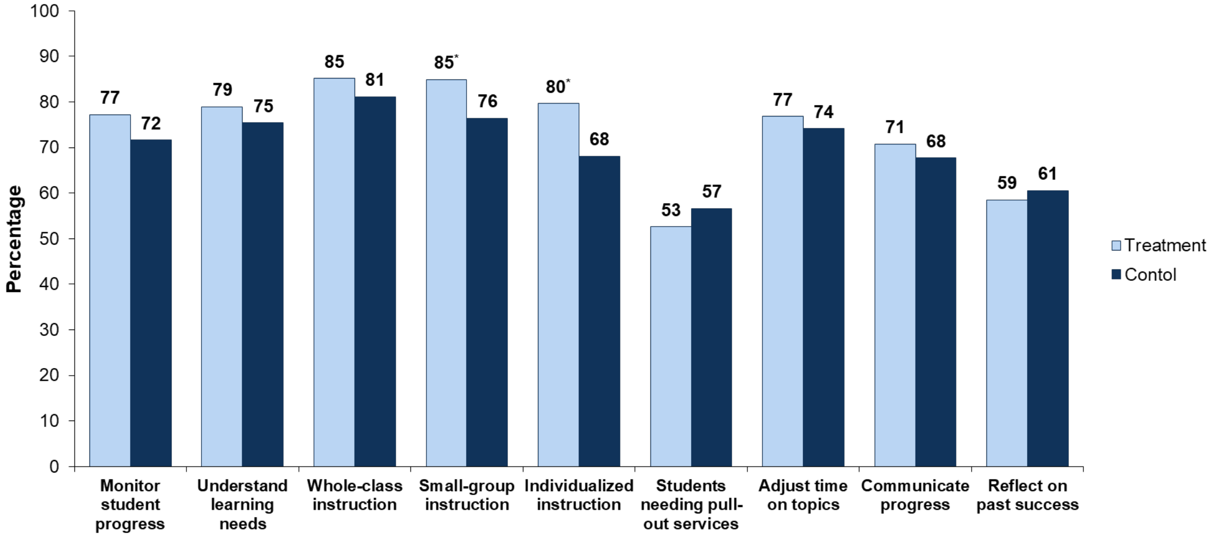
Figure C.1. Percentage of grade 4 and 5 teachers who reported using data at least several times per month for different purposes in math during 2015-2016



Source: Teacher survey (n = 396-398).

*Difference is statistically significant at the .05 level, two-tailed test.

Figure C.2. Percentage of grade 4 and 5 teachers who reported using data at least several times per month for different purposes in English/language arts during 2015-2016



Source: Teacher survey (n = 408-411).

*Difference is statistically significant at the .05 level, two-tailed test.

Table C.1. Impacts of support for DDI on teachers' reports of access to specific types of student data (percentages)

Type of student data	Treatment	Control	Impact	*p*-value
Assessment data				
Summative assessment results	98.0	97.7	0.3	0.75
Interim assessment results	95.8	94.0	1.9	0.28
Formative assessment results	>98.6	>98.6	---	0.81
Achievement data, broken down by student characteristics	59.3	52.2	7.1	0.13
Other types of data				
Samples of student work *for other students*	82.6	76.6	6.0	0.09
Grades from prior year	77.6	80.9	-3.3	0.34
Attendance	95.3	97.0	-1.7	0.44
School behavior	89.8	90.1	-0.3	0.93
Readiness for grade promotion or graduation	71.3	71.8	-0.6	0.88
Number of teachers—range[a]	**219-219**	**214-215**	**433-434**	

Source: Teacher survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics statistical standards, but that the percentage is less than or greater than the number following the < or > symbol (U.S. Department of Education 2012).

Table C.2. Impacts of support for DDI on the types of data teachers used to guide math and reading instruction (percentages who reported using each type of data at least monthly or several times per term)

Type of student data	Math				Reading			
	Treatment	Control	Impact	*p*-value	Treatment	Control	Impact	*p*-value
Assessment data								
Summative assessment results	81.2	78.6	2.6	0.39	76.8	78.5	-1.7	0.65
Interim assessment results	78.5	71.6	6.9*	0.05	76.8	75.4	1.3	0.70
Formative assessment results	93.1	94.2	-1.1	0.64	92.7	90.7	2.0	0.40
Student achievement data, broken down by student background characteristics	28.8	22.9	5.9	0.15	26.5	21.8	4.7	0.17
Other types of data								
Samples of student work *for other students*	66.2	62.6	3.6	0.46	75.1	68.3	6.8	0.10
Grades from prior year	12.0	11.5	0.4	0.84	13.7	14.2	-0.5	0.83
Attendance	43.7	41.7	2.0	0.59	41.1	38.9	2.3	0.59
School behavior	57.3	50.7	6.6	0.11	49.9	45.8	4.0	0.28
Readiness for grade promotion or graduation	35.3	30.4	4.9	0.29	32.7	30.4	2.3	0.54
Number of teachers—range[a]	**195-196**	**199-201**	**395-397**		**205-206**	**203-205**	**409-411**	

Source: Teacher survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

Table C.3. Impacts of support for DDI on teachers' reports that specific types of student data are "very useful" when making instructional decisions (percentages)

Type of student data	Treatment	Control	Impact	*p*-value
Assessment data				
Summative assessment results	45.9	46.5	-0.6	0.89
Interim assessment results	43.3	43.9	-0.6	0.88
Formative assessment results	80.3	79.6	0.7	0.85
Student achievement data, broken down by student background characteristics	17.0	14.2	2.8	0.44
Other types of data				
Samples of student work *for other students*	76.1	71.5	4.6	0.13
Grades from prior year	7.1	4.0	3.1	0.11
Attendance	22.8	21.9	0.9	0.80
School behavior	35.4	33.5	1.8	0.66
Readiness for grade promotion or graduation	17.8	16.1	1.8	0.54
Number of teachers—range[a]	**219-219**	**214-215**	**433-434**	

Source: Teacher survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

Table C.4. Impacts of support for DDI on teachers' level of confidence in using data to guide their instruction (percentages)

	Treatment	Control	Impact	*p*-value
Confident or very confident in ability to use data to guide instruction	79.1	77.2	1.9	0.57
Number of teachers—range[a]	**218-218**	**215-215**	**433-433**	

Source: Teacher survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

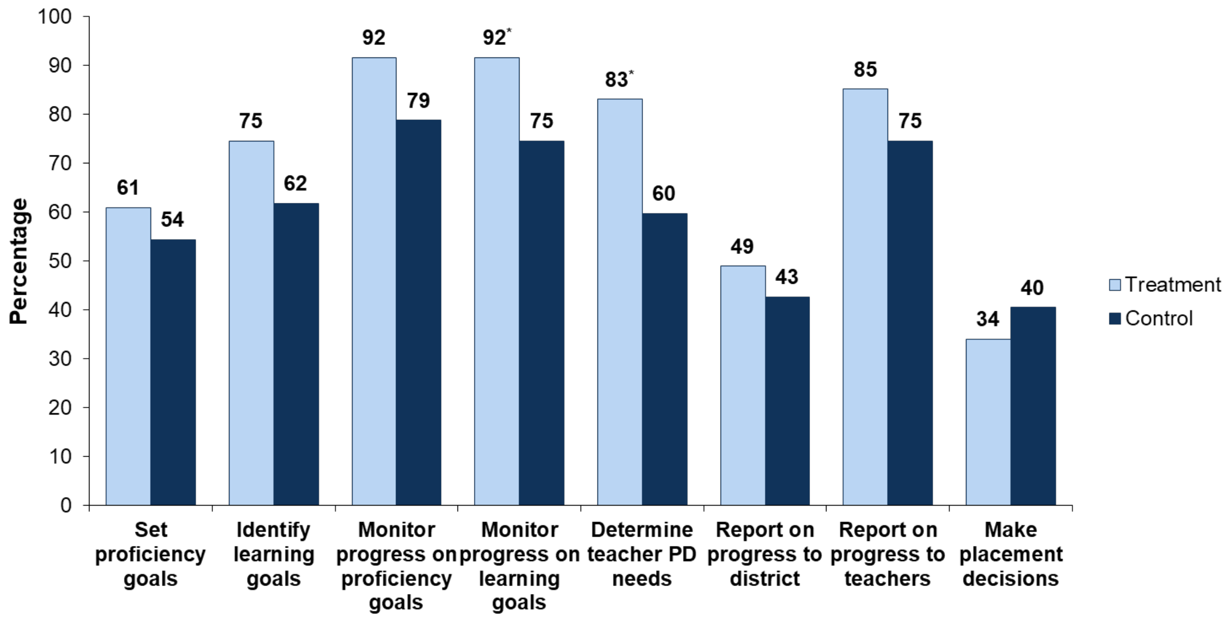
[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

Principals' Access to and Use of Data

Next, we present estimates of the impact of the DDI intervention on a set of outcomes related to principals' data use. Figure C.3 shows estimated impacts of DDI on principals' data use, while the tables in this section show estimates of impacts on the kinds of data principals' used (table C.5) and their reported barriers to data use (table C.6).

Figure C.3. Percentage of principals who used data at least monthly or several times per term for different purposes during 2015-2016



Source: Principal survey (n = 92-94).

*Difference is statistically significant at the .05 level, two-tailed test.

Table C.5. Impacts of support for DDI on the types of student data principals used to inform school decisions (percentages who reported at least monthly or several times per term)

Type of student data	Treatment	Control	Impact	*p*-value
Assessment data				
Summative assessment results	74.5	57.4	17.0	0.07
Interim assessment results	87.2	68.1	19.1*	0.03
Formative assessment results	87.2	78.7	8.5	0.32
Student achievement data, broken down by student background characteristics	53.2	31.9	21.3*	0.02
Other types of data				
Samples of student work	68.1	59.6	8.5	0.32
Past course grades	22.2	15.6	6.7	0.26
Attendance	83.0	72.3	10.6	0.20
School behavior	80.9	74.5	6.4	0.41
Readiness for grade promotion or graduation	23.4	29.8	-6.4	0.44
Number of principals—range[a]	**45-47**	**45-47**	**90-94**	

Source: Principal survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

Table C.6. Impacts of support for DDI on principals' reported barriers to accessing student data (percentages who reported a moderate or major barrier)

	Treatment	Control	Impact	*p*-value
Lack of access to:				
Student-level data in a usable form	21.7	21.7	0.0	1.00
Technology and tools to help track and analyze student data	10.6	12.8	-2.1	0.71
Analysis and reports of student data in a usable form	17.8	26.7	-8.9	0.21
Formal training on how to analyze and use student data to inform instructional practice	14.9	48.9	-34.0*	0.00
Coaching, mentoring, or other one-on-one support on how to analyze and interpret student data	10.6	53.2	-42.6*	0.00
Coaching, mentoring, or other one-on-one support on how to change instruction based on data	17.0	57.4	-40.4*	0.00
Information or other resources on evidence-based instructional strategies (best practices)	6.4	38.3	-31.9*	0.00
Number of principals—range[a]	**45-47**	**45-47**	**90-94**	

Source: Principal survey.

Notes: The difference between the treatment and control estimates may not equal the impact shown in the table because of rounding.

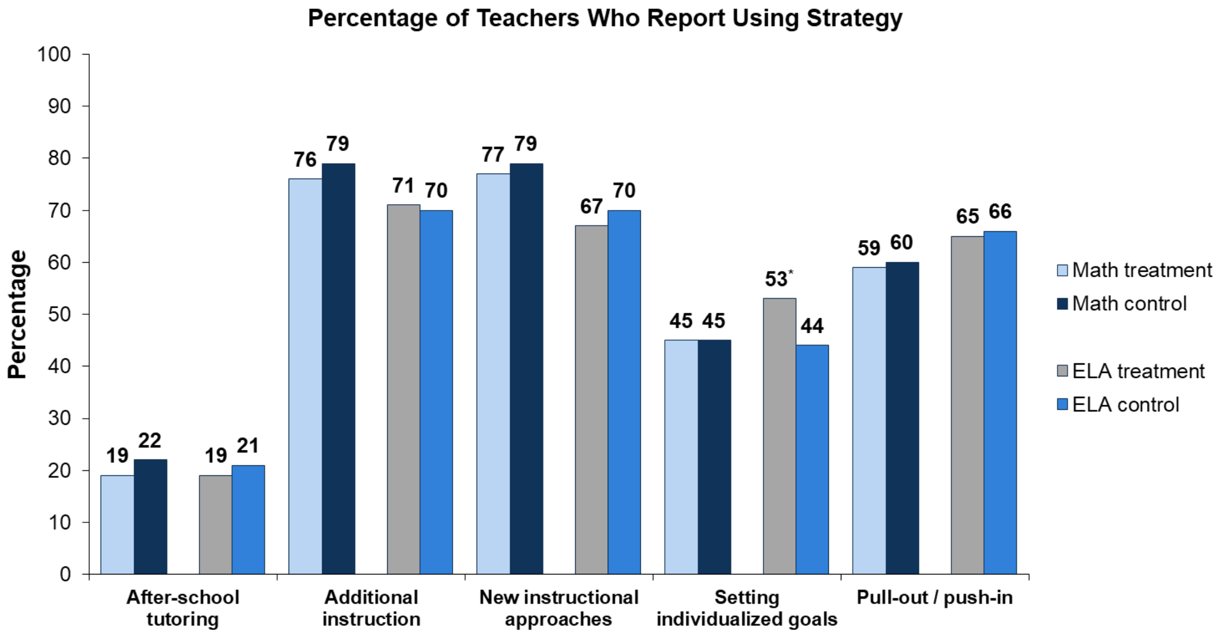
[a]Sample sizes are presented as a range, based on the data available for each row in the table.

*Impact is statistically significant at the .05 level, two-tailed test.

Teachers' Use of Instructional Strategies Associated with DDI

Next, we present estimates of the impacts of the DDI intervention on teachers' use of instructional practices, based on a different cutoff than that used in Chapter IV. Figure C.4 shows impacts on whether teachers used a particular practice at least several times per month.

Figure C.4. Percentage of grade 4 and 5 teachers who reported using instructional practice at least several times per month for different purposes in English/language arts and math during 2015-2016



Source: Teacher survey (n = 397-411).

ELA = English/language arts.

*Difference is statistically significant at the .05 level, two-tailed test.

Impacts of the Support for DDI Intervention on Student Outcomes

This section provides additional details on the estimation of impacts of the DDI intervention on student outcomes.

Estimated Impacts on Student Achievement in Spring 2017

The DDI model requires teachers to acquire and use new skills related to data interpretation, instructional planning, and delivery of instruction, and these skills might take more than a single year to learn and put in practice. If teachers continued integrating the DDI model into their work beyond the 2015-2016 school year, then the intervention might have led to positive impacts on student achievement in later years even though the study supported data coaches only through the 2015-2016 school year. In order to assess this hypothesis, we gathered an additional year of test scores and estimated the impact of DDI on the math and English/language arts achievement of 4th and 5th graders in spring 2017, one year after formal implementation of the DDI intervention ended. This section briefly describes the data and methods used to estimate these impacts, and presents the findings of this analysis.

We tested the impact of DDI on student math and English/language arts achievement in spring 2017 using average test scores within each study school and grade. We collected these data (hereafter referred to as school-grade data) from states, which provided estimated average test scores in spring 2016 and spring 2017 for 4th and 5th graders in each study school. Although we collected and analyzed the data as average scores, we would expect the school-grade data to yield impact estimates similar to those from the student-level data. Conceptually, the impacts from student-level data reflected differences between the treatment and control group in the average achievement level in each grade and school, because random assignment took place at the school level and our main student-level impact estimates weighted student records to give each grade and school equal weight. However, in practice, there are multiple reasons why average test scores based on the school-grade data might not perfectly match those based on student-level data.

First, average test scores based on the school-grade level and average scores based on the student-level data might differ if the underlying sample of students in any given school and grade is not identical. For example, the student-level data included any student who was present in a study school on testing day and took the statewide assessment in spring 2016, even if that student spent most of the year at a different school. In contrast, the school-grade data reflect state policy governing which school a student should be associated with for accountability purposes. In addition, in the student-level data, one district was only able to provide test scores for students who had attended a study school in spring 2014, so students who had joined a study school during the 2015 or 2016 school year were not included in the student-level sample, but are presumably included in the school-grade data.

Second, test scores from these two sources might differ because different procedures were used to determine which scores to include. Differences in average test scores arising from this reason were generally small, with one exception. In one district, the average scores in the school-grade data included scores from an alternate exam. The study team excluded these alternate scores from the student sample since they were measured on a different scale from the primary score. To address this issue, we replaced the original 2016 school-grade average scores in this district with those created from the student-level data (that excluded alternate scores). We did not have this option for 2017, so we adjusted the 2017 school-grade average scores in this district using a school-grade specific adjustment factor reflecting the 2016 difference in average scores between the two data sources.[13]

To assess whether we could use the school-grade data to estimate the impact of DDI on student achievement in 2017, we examined whether the two data sources gave a similar impact estimate in 2016. To make this comparison, we estimated a model that included the following covariates: school-grade average test scores from spring 2013 in the same subject as the outcome measure and school level demographic variables including the percent male, percent black, percent Hispanic, percent other (nonwhite) race, percent with an individualized education plan, percent eligible for free or reduced-price meals, and percent who were English language learners.

[13] Specifically, we used a three-step process. First, we calculated the difference: [school-grade mean of 2016 score based on student-level data minus school-grade mean of 2016 score based on school-grade data]. Second, we standardized this difference by dividing it by the statewide standard deviation of 2016 scores in that grade and subject. Third, we added this factor (which varies at the school, grade and subject level) to the 2017 standardized score in each school, grade, and subject. The result is an adjusted score that approximates what we would have measured as the school-grade mean of student scale scores in 2017, if we had collected student-level data in 2017 and the mean scores from that data source had the same relationship with the average score based on school-grade data in 2017 as existed in 2016.

The first three rows of Table C.7 summarize the results of this analysis. Based on the school-grade data (row 1), the estimated impact of DDI on student achievement in math and English/language arts was not statistically significant. The estimate is within 0.03 of the main impact estimate based on student-level data shown in row 2. The estimate based on school-grade data is even closer to the impact estimate based on student-level data when the analysis was conducted at the school level—that is, the student test scores and other characteristics were converted into school-level averages (row 3). Overall, while the estimates based on school-grade data are not identical to those based on student-level data, they are close.

We used the same model with school-grade data to estimate the impact of DDI on student achievement in 2017. The results of this analysis indicate that DDI did not affect student math or English/language arts achievement in 2017. The second panel of Table C.7 presents two sets of impact estimates, with and without the adjustment described above. In each case, the estimated impact of DDI is close to zero and not statistically significant.

Table C.7. Impacts of support for DDI on 2016 and 2017 student achievement

2016 Achievement	Math				English/Language Arts			
Data and model	**Treatment**	**Control**	**Impact**	**SE**	**Treatment**	**Control**	**Impact**	**SE**
School-grade data	-0.31	-0.24	-0.07	0.038	-0.28	-0.24	-0.04	0.035
Student-level data, main study estimate	-0.28	-0.23	-0.04	0.026	-0.25	-0.25	-0.01	0.027
Student-level data, school averages	-0.31	-0.23	-0.08	0.045	-0.27	-0.25	-0.02	0.047

2017 Achievement	Math				English/Language Arts			
Data and model	**Treatment**	**Control**	**Impact**	**SE**	**Treatment**	**Control**	**Impact**	**SE**
School-grade records	-0.26	-0.23	-0.03	0.047	-0.23	-0.19	-0.04	0.037
School-grade records, no adjustment	-0.31	-0.27	-0.04	0.047	-0.26	-0.23	-0.03	0.038

Source: District and state administrative data; National Center for Education Statistics (NCES) Common Core of Data.

Notes: Models using school-grade data included the following covariates: 2013 school-grade average test score in the same subject as the outcome, interacted with indicators of the school's state, and school-level percent FRL, SPED/IEP, ELL, Black, Hispanic, Other race/ethnicity, and male. None of the impacts are statistically significant at the .05 level, two-tailed test.

Sensitivity of Estimated Impacts on Student Achievement to Alternative Specifications

This section provides additional details on the analysis of the sensitivity of the estimated impacts on student achievement using models based on different assumptions. Tables C.8 and C.9 present estimated impacts on math and English/language arts achievement, respectively. Impact estimates are presented in standard deviation units. Model 1 is the main impact estimation model described in appendix A. The models presented in the other rows of this table differ from the primary model in the following ways:

- Assumptions about the structure of the error term in the model: random effects (model 2) rather than fixed effects (model 1)

- The covariates included in the model: no covariates (model 3), an expanded or rich set of covariates (model 4), or alternative baseline achievement covariates included (model 11) rather than the standard set of covariates (model 1)
- The sample definition: the spring 2014 sample (model 5); the 2016 sample, including three students with extreme values of test scores (model 6); or the sample present in study schools from spring 2014 to spring 2016 (model 7) rather than the main 2016 sample (model 1)
- The unit of analysis: school-level analysis (models 8 and 9) rather than student-level analysis (model 1)
- Inclusion of district fixed effects: district fixed effects and matched-pair random effects (model 10) rather than matched-pair fixed effects and no district effects (model 1)
- Alternative approaches to missing values for covariates: no imputation for covariates and observations with missing values dropped (model 12); single imputation (model 13) versus multiple imputations (model 1)

Overall, the results were not greatly sensitive to these alternative specifications. In the case of math, each impact estimate was less than 0 and ranged from -0.012 to -0.080. Most of the estimates were not statistically significant, although three estimates were significant. In the case of English/language arts, the estimates ranged from -0.029 to -0.002 and none was statistically significant.

Table C.8. Impacts of support for DDI on students' math scores

#	Model description	Impact	se	*p*-value	N_t	N_c
1	Covariates, FE, cluster SE	-0.043	0.026	0.11	5925	6111
2	Covariates, RE†	-0.044	0.037	0.23	5925	6111
3	No covariates, FE, cluster SE	-0.080*	0.032	0.01	5925	6111
4	Rich covariates, FE, cluster SE†	-0.044	0.026	0.10	5925	6111
5	2014 sample, covariates, FE, cluster SE	-0.062*	0.023	0.01	5326	5450
6	Model 1, but including three students with extreme z-scores†	-0.042	0.026	0.12	5926	6113
7	Model 1, including only "stayers," for whom School_{2016} = School_{2014}†	-0.046	0.027	0.10	4401	4589
8	School-level, no covariates, FE	-0.080	0.045	0.08	51	51
9	School-level, no covariates, RE	-0.080	0.044	0.07	51	51
10	Covariates, district FE, MP RE, and MP-varying random treatment coefficient†	-0.051	0.038	0.17	5925	6111
11	Model 1, but including 2015 z-scores as covariates instead of 2013 school z means†	-0.012	0.024	0.63	5925	6111
12	Model 1 with nonimputed dataset	-0.051	0.026	0.06	4932	5046
13	Model 1 with single-imputed dataset†	-0.043	0.026	0.11	5925	6111

Source: District and state administrative data; National Center for Education Statistics (NCES) Common Core of Data.

Notes: For brevity, the table uses several abbreviations: matched pair (MP), random effects (RE), fixed effects (FE), and standard errors calculated adjusting for clustering at the school level (Cluster SE). Model descriptions with a dagger (†) indicate those estimated using a single-imputed dataset. The imputation method for that dataset was identical to the multiple imputation procedure, but only one imputed dataset was retained rather than 20, and the model estimation used standard, single-dataset estimation procedures rather than estimation techniques that combine parameter estimates from multiple imputed datasets.

*Impact is statistically significant at the .05 level, two-tailed test.

Table C.9. Impacts of Support for DDI on students' English/language arts scores

#	Model description	Impact	se	*p*-value	N_t	N_c
1	Covariates, FE, cluster SE	-0.006	0.027	0.82	5918	6100
2	Covariates, RE†	-0.009	0.036	0.80	5918	6100
3	No covariates, FE, cluster SE	-0.024	0.033	0.47	5918	6100
4	Rich covariates, FE, cluster SE†	-0.023	0.025	0.37	5918	6100
5	2014 sample, covariates, FE, cluster SE	-0.012	0.026	0.65	5330	5450
6	Model 1, but including students with extreme z-scores[a]	--	--	--	--	--
7	Model 1 including only "stayers," for whom $\text{School}_{2016} = \text{School}_{2014}$†	-0.002	0.031	0.95	4406	4593
8	School-level, no covariates, FE	-0.024	0.047	0.61	51	51
9	School-level, no covariates, RE	-0.024	0.046	0.61	51	51
10	Covariates, district FE, MP RE, and MP-varying random treatment coefficient†	-0.014	0.038	0.70	5918	6100
11	Model 1, but including 2015 z-scores as covariates instead of 2013 school z means†	-0.020	0.017	0.25	5918	6100
12	Model 1 with nonimputed dataset	-0.014	0.030	0.64	4926	5039
13	Model 1 with single-imputed dataset†	-0.005	0.027	0.84	5918	6100

Source: District and state administrative data, NCES Common Core of Data.

Notes: For brevity, the table uses several abbreviations: matched pair (MP), random effects (RE), fixed effects (FE), standard errors calculated adjusting for clustering at the school level (Cluster SE). Model descriptions with a dagger (†) indicate those estimated using a single-imputed dataset. The imputation method for that dataset was identical to the multiple imputation procedure, but only one imputed dataset was retained rather than 20, and the model estimation used standard, single-dataset estimation procedures rather than estimation techniques that combine parameter estimates from multiple imputed datasets.

[a] There were no students with extreme z-scores in the ELA analysis sample, so this sensitivity test does not apply.

*Impact is statistically significant at the .05 level, two-tailed test.

Impacts on Student Behavior

This section provides additional details on estimated impacts of the DDI intervention on student behavior outcomes. Table C.10 presents the impacts of DDI on five measures of student behavior during the 2015-2016 school year. All of the impacts were small, and none was statistically significant. We requested data from each study district on the number of school days a student attended school in the district during the 2015-2016 school year and the two previous years, as well as the number of days each student was enrolled in the district during each year. We calculated the attendance rate as the proportion of enrolled days in which a student attended school. We also requested data on the total number of days a student spent in out-of-school suspensions during the 2015-2016 school year, the total number of episodes or incidents of out-of-school suspensions, and the same measures for in-school suspensions. We used these data to calculate the total number of school days spent in each form of suspension and the total number of each type of suspension incident each student experienced during the 2015-2016 school year.

Table C.10. Impacts of support for DDI on student behavioral outcomes

Student behavioral outcome	Treatment	Control	Impact	*p*-value
Attendance rate	0.95	0.95	0.00	0.41
Days of out-of-school suspension (OSS)	0.17	0.13	0.03	0.26
Number of OSS incidents	0.09	0.08	0.01	0.59
Days of in-school suspension (ISS)	0.08	0.08	0.00	0.89
Number of ISS incidents	0.07	0.06	0.01	0.26

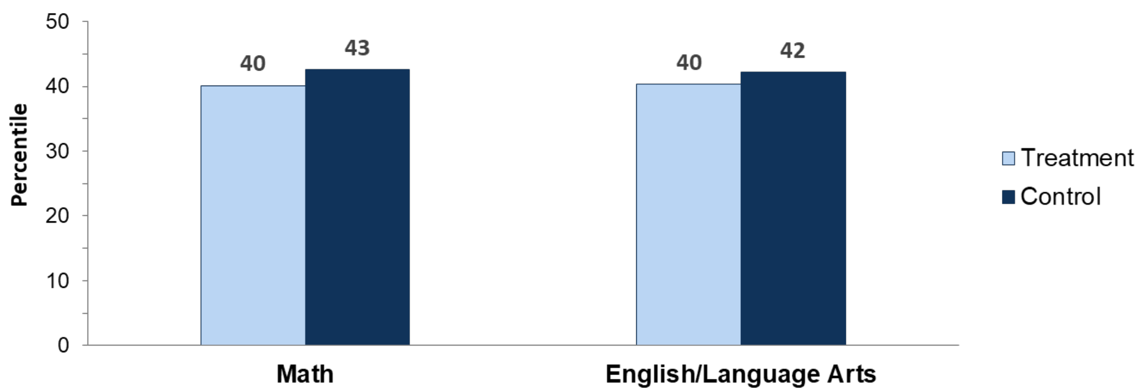
Source: District student records, n = 10,931-12,535 students

Note: None of the estimated impacts is statistically significant at the .05 level, two-tailed test.

Impacts Among Schools with an Instructional Focus that Included English/Language Arts

This section shows estimates of the impacts of the DDI intervention on student achievement in math and English/language arts among schools with an instructional focus in the area of English/language arts or a focus in both English/language arts and math (figure C.5). Among these schools, neither the estimated impact of the intervention on student achievement in English/language arts or math was statistically significant.

Figure C.5. Mean student achievement on 2016 state assessments in math and English/language arts, among schools with an instructional focus in an area of English/language arts (including those that also focused on math)



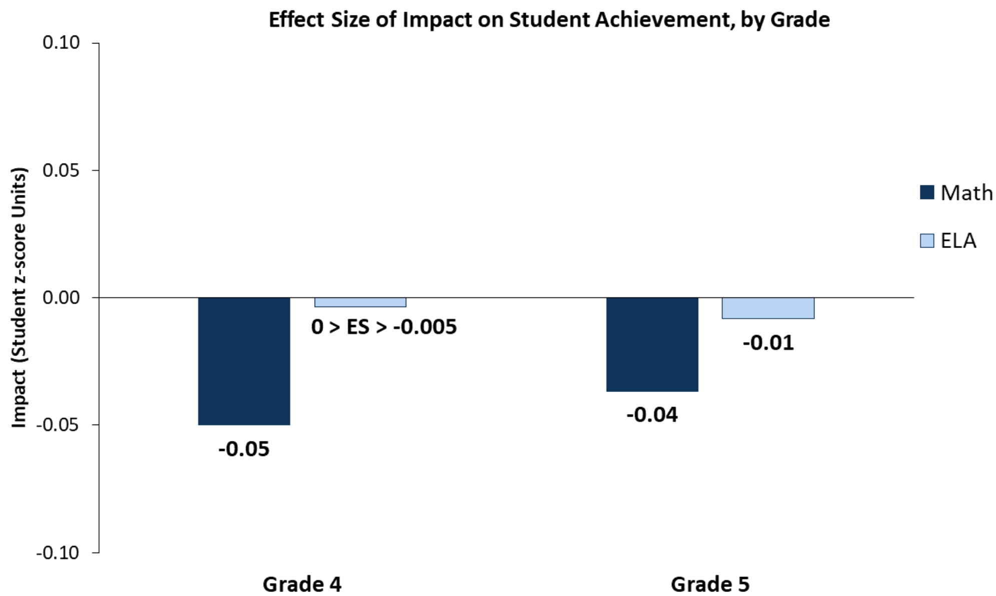
Source: District student records (n = 8,232-8,254).
Neither difference is statistically significant at the .05 level, two-tailed test.

Impacts on Student Achievement, by Student Subgroup

This section provides additional details on the estimated impacts of the DDI intervention within subgroups, as defined by students' grade level and 2015 achievement level. Figure C.6 presents the impacts on student math and reading achievement among students enrolled in fourth and fifth grade in spring 2016. Figure C.7 presents the impacts of DDI on student math and reading achievement within three groups defined by performance categories on the spring 2015 state assessment in the same subject as the outcome being analyzed. Subgroups are defined based on student test scores falling below proficient, at or above the proficiency threshold but below the advanced threshold, and at or above the advanced threshold. The estimated impacts of the

DDI intervention on achievement in both subjects were similar for each of these subgroups, with no significant differences in impacts for different subgroups.

Figure C.6. Impacts of support for DDI on students' math and English/language arts scores, by grade level

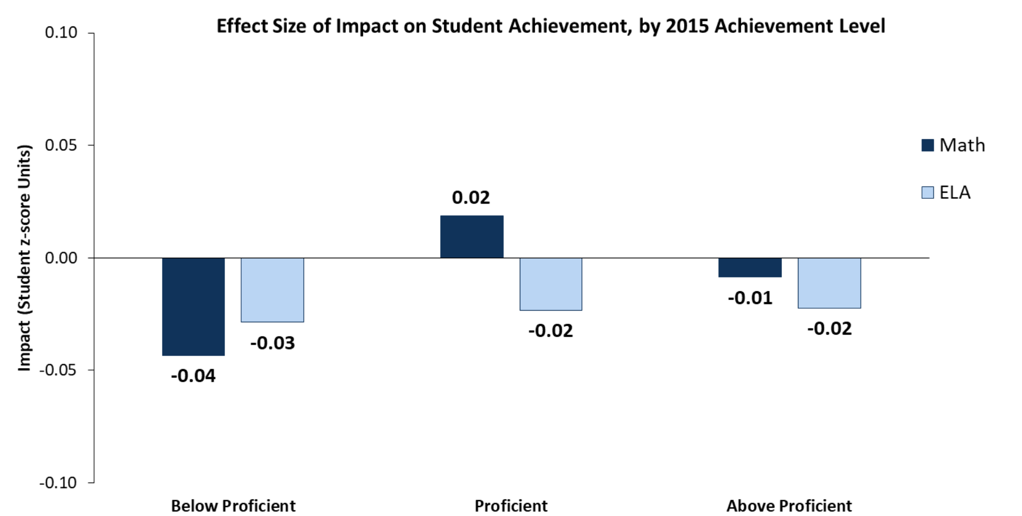


Source: District student records, n = 12,036 students.

Note: Neither the estimated impacts nor the difference between grades in impacts are statistically significant at the .05 level, two-tailed test.

ELA = English/language arts.

Figure C.7. Impacts of support for DDI on students' math and English/language arts scores, by students' baseline proficiency level



Source: District student records, n = 11,213 students.

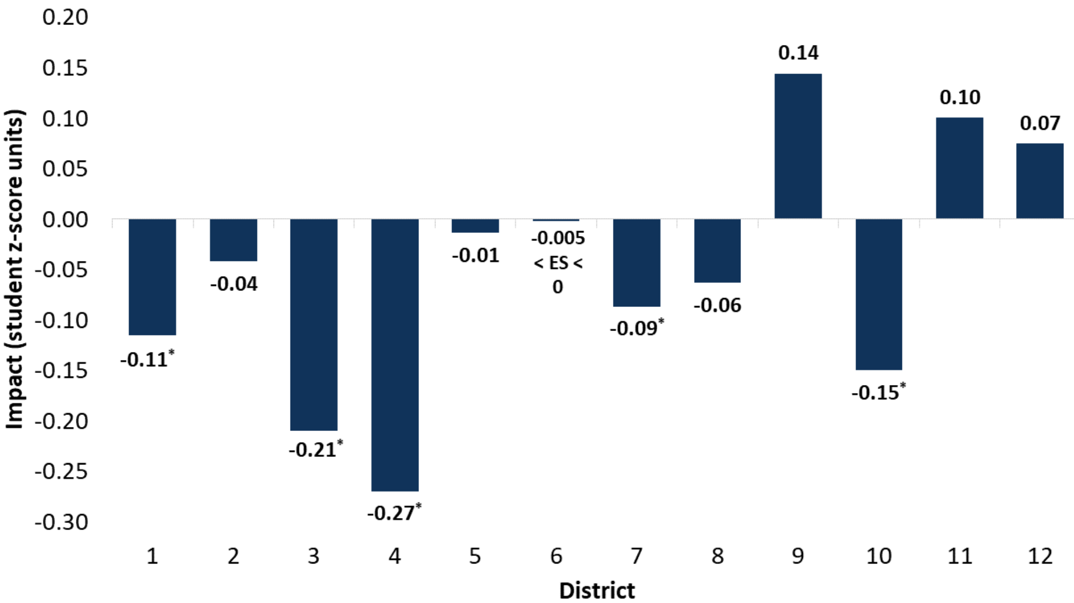
Note: Neither the estimated impacts nor the differences between any of the proficiency levels in impacts is statistically significant at the .05 level, two-tailed test.

ELA = English/language arts.

Variation Across Districts and School Readiness in Impacts on Student Achievement

This section presents evidence on how much impacts differed across study districts. Figure C.8 shows impacts of DDI on students' math achievement and figure C.9 shows impacts on English/language arts achievement. Impacts in each subject differed between individual districts by a statistically significant and substantial amount. Figure C.10 shows the impacts based on the alternative definition of school readiness described in Chapter IV (box 4).

Figure C.8. Estimated impact of support for DDI on student achievement on 2016 state assessments in math, by district

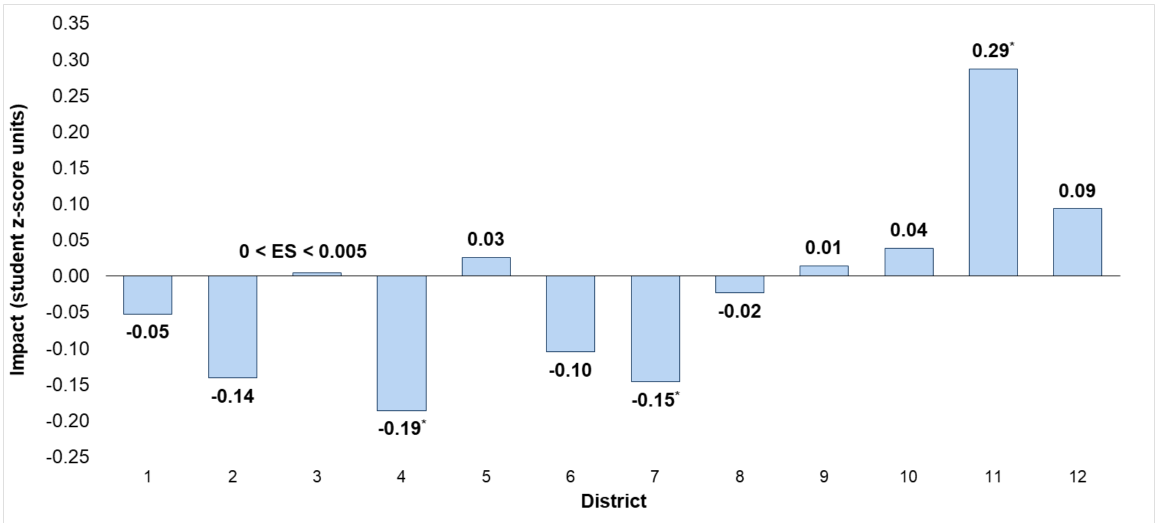


Source: District student records (n = 12,036).

*Estimated impact in district is statistically significant at the .05 level, two-tailed test.

Variation in estimated impacts across districts is statistically significant at the .05 level, two-tailed test.

Figure C.9. Estimated impact of support for DDI on student achievement on 2016 state assessments in English/language arts, by district

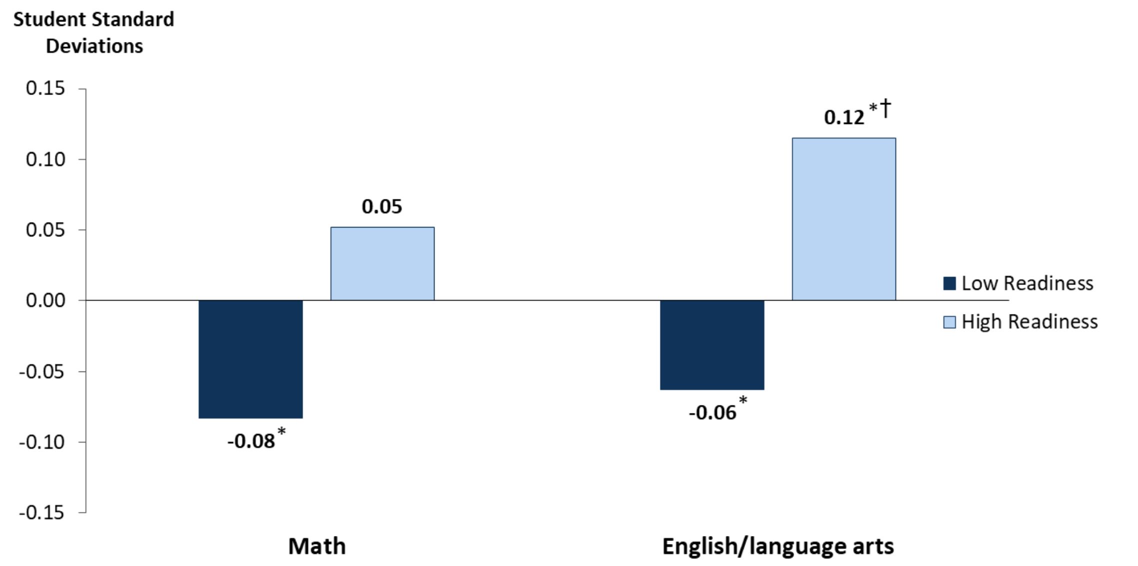


Source: District student records (n = 12,018).

*Estimated impact in district is statistically significant at the .05 level, two-tailed test.

Variation in estimated impacts across districts is statistically significant at the .05 level, two-tailed test.

Figure C.10. Estimated impact of support for DDI on student achievement on 2016 state assessments in math and English/ language arts, by school's implementation readiness (alternative measure)



Source: District student records (n = 9,693 – 9,708); data coach interviews.

*Estimated impact within subgroup is statistically significant at the .05 level, two-tailed test.

†Difference in impact estimates between subgroups is statistically significant at the 0.05 level, two-tailed test.

Variation in Impacts on Student Achievement and Teacher Practices Across Levels of Coach Preparation

Data coaches' level of preparation might affect their performance as coaches, influencing the impact of DDI on teacher data use and instructional practices as well as student achievement. To explore this hypothesis, we estimated impacts of DDI on student achievement and three measures of teacher practice, within subgroups of study schools with coaches who had either higher or lower levels of preparation to be a data coach (based on two measures of coach preparation). The outcome measures, student achievement and number of teacher practices in several areas, are defined in the same way as elsewhere in the report: student achievement is measured using z-scores on the spring 2016 math and English/language arts assessments, while teacher practices are measured as the number of practices a teacher engaged in at least several times per week. The first measure of coach preparation, presented in table C.11, simply identifies coaches who reported having previously served as an instructional coach before joining the DDI study. The second measure, presented in table C.12, distinguishes coaches who reported that the DDI training provided by Focus Schools left them "prepared to carry out most or all tasks" (38 coaches) from those who reported feeling either "prepared to carry out some but not all tasks" or "not at all prepared" (12 coaches).

Results of these analyses suggest that the impacts of DDI on student achievement and teacher practices were less positive and/or more negative in schools where the coach had prior experience as an instructional coach (table C.11). In contrast, the relationship between impacts and coaches' reported adequacy of training was less consistent (table C.12). There is some evidence that coaches who reported less adequate training had more negative impacts on student achievement and teacher instructional practices in math, but this pattern was not consistent between math and reading. These results should be interpreted with caution since coaches' prior experience or preparation were not randomly assigned. In particular, it could be that factors other than coaches' prior experience are affecting the estimates in table C.11. For example, if schools with more anticipated barriers to successful implementation of DDI were more likely to hire coaches with prior experience, then these subgroup differences might reflect the effect of those implementation barriers rather than the effect of prior experience. In addition, coaches were asked about their level of preparation after they started working in the schools, so coaches who were facing greater challenges may have reported they felt less prepared.

Table C.11. Impact of support for DDI on student achievement and teacher practices in math and English/language arts, by coach's prior experience as instructional coach

	Math			English/Language Arts		
	Impact	se	*p-value* of difference in impacts	Impact	se	*p-value* of difference in impacts
Student achievement						
No experience	-0.02	0.045		0.04	0.044	
Prior experience	-0.10*	0.042	0.284	-0.09*	0.038	0.041
Teacher data use practices						
No experience	0.90*	0.282		0.87*	0.279	
Prior experience	-0.89	0.560	0.005	-0.63	0.588	0.024
Teacher instructional practices						
No experience	0.16	0.149		0.15	0.128	
Prior experience	-0.70*	0.194	< 0.005	-0.27	0.238	0.125
Teacher collaborative activities (not subject-specific)						
No experience	0.06	0.370				
Prior experience	0.79	0.436	0.208			

Source: District student records, n = 10,657 – 10,666 students, teacher survey, n = 356 – 379, and data coach interview, n= 43 – 45.

*Estimated impact within subgroup is statistically significant at the .05 level, two-tailed test.

Table C.12. Impact of support for DDI on student achievement and teacher practices in math and English/language arts, by coach-reported adequacy of DDI training

	Math			English/Language Arts		
	Impact	se	*p-value* of difference in impacts	Impact	se	*p-value* of difference in impacts
Student achievement						
Lower adequacy	-0.13*	0.052		-0.05	0.055	
Higher adequacy	-0.01	0.031	0.060	0.01	0.031	0.285
Teacher data use practices						
Lower adequacy	0.20	0.555		0.20	0.379	
Higher adequacy	0.12	0.304	0.896	0.13	0.339	0.886
Teacher instructional practices						
Lower adequacy	-0.41*	0.198		-0.07	0.226	
Higher adequacy	-0.10	0.144	0.215	-0.08	0.132	0.980
Teacher collaborative activities (not subject-specific)						
Lower adequacy	-1.23*	0.485				
Higher adequacy	0.55	0.302	0.002			

Source: District student records, n = 11,819 – 11,838 students, teacher survey, n = 398 – 422, and data coach interview, n= 48 – 50.

*Estimated impact within subgroup is statistically significant at the .05 level, two-tailed test.

ACKNOWLEDGEMENTS

This study would not have been possible without the contributions of many individuals. We are grateful for the cooperation of the many data coaches, teachers, principals, and district leaders and staff who helped implement the data-driven instruction intervention, assisted with the study's data collection, and provided important information that shaped the study. A team of leaders and expert consultants from Focused Schools were primarily responsible for shaping the intervention and providing the professional development and technical assistance for the study. They included Focused Schools' senior executives Jan Leight, Joseph Palumbo, and Kerry Purcell, and consultants Mary Jo Bangert, Brett Bishop, Maureen Ciccone, Kathy Crum, Janice Hanada, Anne Midden, and Barbara Rudiak.

Several individuals made enormous efforts to collect and analyze data successfully for this study. Sheila Heaviside provided excellent leadership over our administration of teacher and principal surveys, and additional support was provided by Alice Gao, Scott Reid, Kathy Shepperson, and Boris Shkolnik. A large team of staff from Mathematica, Abt, Evidence Based Education Research and Evaluation (EBERE), and Synergy contributed to a successful effort of recruiting schools into the study, including Porsha Cropper, Brittany English, Sherri Lauver, Barbara Harris, Chris Jones, Alyssa Ruff Fountain, and many others. Acquiring administrative data and analyzing both administrative and survey data required a large effort with contributions from Juan DelaCruz, Naihobe Gonzales, Sophie MacIntyre, Steve Malick, Sam McCormick, Kelley Monzella, Sarah Osborn, Juha Sohlberg, Rayna Thornton, Charles Tilley, and Rian Watt.

Many people contributed to the study design and the analysis and interpretation of the study's data and the production of this report. The study received useful advice from our technical working group, consisting of Geoffrey Borman, Russell Gersten, Laura Hamilton, Ken Koedinger, Spyros Konstantopoulos, Leslie Nabors-Olah, Ellen Mandinach, Jon Supovitz, Jeff Wayman, and Martin West. At Mathematica, Josh Furgeson, Irma Perez-Johnson, and Alison Wellington helped shape and guide the evaluation and provided expert contributions and advice and various stages of the study. At Abt, Amy Checkoway, Kim Dadisman, and Beth Gamse provided important input into the design and conduct of the study. Molly Cameron and Jim Cameron edited the report, and Shantal James prepared the report for publication.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The research team for this evaluation included staff from Mathematica and subcontractors Abt Associates, Evidence Based Education Research and Evaluation (EBERE), and Synergy, Inc.. None of the research team members has financial interests that could be affected by findings from this evaluation. No one on the ten-member technical working group, convened by the research team to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Evaluation and Regional Assistance
Matthew Soldner
Commissioner

September 2019

The report was prepared for the Institute of Education Sciences under Contract No. ED-IES-12-C-0086. The project officer is Erica Johnson in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

Gleason, Philip, Sarah Crissey, Greg Chojnacki, Marykate Zukiewicz, Tim Silva, Sarah Costelloe, and Fran O'Reilly. (2019). Evaluation of Support for Using Student Data to Inform Teachers' Instruction—APPENDICES (NCEE 2019-4008). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

This report is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.