Widening the Pathway

Implementation and Impacts of Alternative Teacher Preparation Programs Across Three Contexts—Appendixes

ALICE HUGUET, CHRISTOPHER JOSEPH DOSS, BENJAMIN K. MASTER, FATIH UNLU, JESSICA L. SOUSA, KAREN CHRISTIANSON, GARRETT BAKER

Sponsored by TNTP

RAND CORPORATION

For more information on this publication, visit **www.rand.org/t/RRA281-1**.

Published by the RAND Corporation, Santa Monica, Calif.

About This Report

TNTP (formerly The New Teacher Project) is committed to ending educational inequities by promoting the recruitment, training, and retention of high-quality teachers and school leaders. The U.S. Department of Education's Supporting Effective Educator Development (SEED) grant has funded TNTP to implement the Teacher Effectiveness and Certification (TEACh) program in three regions. Through TEACh, TNTP works with organizational partners, such as school districts, to develop a local process to recruit, prepare, and certify teacher candidates, as well as support them with coaching and other training in their first year. The RAND Corporation conducted an evaluation of TEACh, investigating each program's implementation, the effects of TEACh on the recruitment and retention of teachers, and the relative performance of those teachers.

This report presents the results of a study of the implementation and impact of the TEACh program established in partnerships between TNTP and three local partner organizations. The report offers guidance to policymakers and leaders interested in recruiting teachers for positions in K–12 schools through alternative pathways and preparation programs. The findings from this study may also inform teacher preparation practices more broadly. The main report can be found at Huguet et al. (2021).

RAND Education and Labor

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs; workforce development; and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. This study was sponsored by TNTP with funding through a U.S. Department of Education's SEED grant program, grant number U367D170012.

More information about RAND can be found at www.rand.org. Questions about this report should be directed to ahuguet@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

Appendix A. Methods

We fielded a mixed-methods study to holistically analyze the TEACh program. In this appendix, we detail the implementation and impact study methods. We first present the implementation study methods, detailing the design of the implementation research, associated data collection, and data analysis methods. We then present the details of the impact study including the aims of the study, our estimation approach to minimize omitted variables bias, and the models used to estimate effects of TEACh on student achievement and teacher retention. We conclude with limitations.

Implementation Study Methods

Design

Data for the implementation portion of this report were collected through five research activities: (1) interviews with TNTP and district staff, (2) interviews with coaches working with the TEACh candidates, (3) coaching logs related to the frequency of coaching sessions and their primary foci, (4) phone interviews with teaching candidates, (5) phone interviews with principals who hired program teaching candidates, and (6) documents collected from program staff. These data collection activities were designed with a set of implementation measures in mind, which were developed through a collaboration between TNTP and researchers at the RAND Corporation (see Table A.1).

Table A.1. Implementation Measures Guiding Data Collection Planning and Analyses

	Implementation Measures
Hiring Metrics	Recruitment activities attract a diverse range of applicants.
	Program selection and district hiring processes are well defined and transparent to applicants and program staff.
	Partner organizations allocate sufficient personnel and resources to recruitment activities and processes.
	School leaders have knowledge of, and buy into, the program.
Teacher Effectiveness	Coaches and trainers responsible for pre- and in-service support are adequately prepared for their roles.
	Pre-service training is focused on skills considered essential to the district and program.
	Pre-service training provides considerable opportunities for practice.
	In-service coaching provides specific, actionable feedback on instructional practice.
	In-service coaches support teachers using a baseline of agreed-upon strategies and tools.

Equitable Distribution	Hiring and placement processes are in place to match effective teachers to high-need vacancies.
Retention	Program teachers intend to remain in high-needs positions beyond their first year of teaching.
Program Sustainability	Program demonstrates responsiveness to district priorities and adapts when able.
	A plan for district ownership of the program is clearly articulated to district staff, and the plan is provided in an adequate time frame to allow for transition preparations.

Data Sources

Two RAND team members made three in-person visits to each program (in summer 2018, fall 2018, and fall 2019). During both the summer and fall visits, we conducted interviews with TNTP and school district staff, including coaches. We collected documents from program staff, such as stepbacks (reflections on program progress developed by TNTP for each internal team), training content, and enrollment data. Each spring, we conducted phone interviews with program candidates who were in their first year of teaching, as well as phone interviews with principals who had hired TEACh program candidates that school year.

We wrote interview protocols in alignment with the project's implementation measures (see Table A.1). Interviews with program staff lasted one hour to 90 minutes, while interviews with candidates and principals were each 30 minutes. We recorded our interviews and had them professionally transcribed, then uploaded them into the qualitative analysis software Dedoose.

TNTP staff at each program helped select interviewees and schedule interviews. We interviewed all essential staff members in the programs, both those who worked for TNTP and those who worked at the partner organization. We selected candidates for spring phone interviews based on the grade and subject area that they taught, as well as whether they taught at charter or traditional district schools, seeking variety across these dimensions. We invited them to the phone interview via email. We invited the principals of candidates to participate, also via email. In the case of candidate and principal phone interviews, we compensated participants for their time with a $25 Amazon gift code following the interview. Information about interviewees can be found in Table A.2.

Table A.2. Total Number Interviews over Two Years of Data Collection

	Program A	Program B	Program C
Coaches	10	2[a]	10
TNTP staff	17	18	11
Partner organization staff	9	16	13
Residents	12	14	12
Principals	8	7	6

[a] During the first year in Program B, coaches were often also TNTP staff, where they are classified in this table. In the second year, we did not interview coaches in Program B because they had not been assigned during the time of our visit.

In addition to these interviews, we administered a coaching log across programs. We designed the log based on research and knowledge about coaching practices. The logs included questions about the amount of time that coaches spent with their assigned candidates, the kinds of practices they used with them, and whether or not they felt they were improving. Prior to administration, we piloted the logs with coaches in summer 2018. The logs could be completed on a computer or on a smartphone or other device; we intended for coaches to complete them soon after they finished their sessions, and they had to complete them within one week of their session. The intention of the logs was to collect real-time data about coaching and to monitor how it changed over the course of the year(s).

In year 1, we asked every program coach to complete these logs each time they worked with their candidates. In the second year of data collection, changes in the coaching structure presented challenges to this request. In Program B, coaches were not working with candidates until at least January, and the COVID-19 pandemic disrupted coaching a few months later. During the time when coaches were working with candidates in Program B, year 2, each coach completed a log for only a portion of their candidates, as requested by staff at the school district. This was intended to limit the amount of extraneous time coaches were spending on the log. In Program A, coaching responsibilities transferred from TNTP to the partner organization in the second year, and staff from the district requested that we limit each coach to completing a log for *one* of their teaching candidates during the year. To summarize, the percentage of candidates for whom we were receiving coaching logs varied

Table A.3. Total Number of Coaching Logs Completed

	Program A	Program B	Program C
Year 1	626	452	259
Year 2	193	326	303
Total	819	778	562

NOTE: One log equals one *entry* in the online platform.

from year to year and program to program. Information about the number of logs completed in each program, each year, is included in Table A.3.

We offered coaches honorarium for completing logs, which also varied by district, as we in some cases received feedback from the partner organization about how to improve the incentives. In some cases, we offered $10 per week that at least one log was completed. In other cases, we offered $5–$10 per log; and in still others, we offered graduated incentives (in which coaches could earn more *per* log as they filled out more logs overall). These also changed by year as the coaching shifted from TNTP to the partner organizations' purview.

We collected documents from program staff. These included enrollment and hiring data, training information, stepbacks, and other documents. Stepbacks are internal presentations that TNTP and program staff generate to reflect on successes, challenges, and progress in their pre-service training (PST) and the school year. Stepbacks often included TNTP and the programs' internal analyses of candidate and principal surveys. The stepbacks provided opportunities to confirm or disconfirm themes that emerged from our other data sources.

Analysis

Interview Analysis

We developed the coding scheme to align with the study's implementation measures, outlined in Table A.1. We applied deductive codes that were descriptive (e.g., we identified when an interviewee was discussing recruitment) as well as analytic (e.g., we identified when the discussion was about strength of the program). We also remained open to emergent codes that arose as we reviewed our transcripts and notes. For example, if coders saw that a topic was discussed by multiple interviewees but we did not have an existing code for it, we would create a code for that topic. In each round of coding, a group of three or four coders practiced coding on the same transcripts, followed by team debriefs to identify, discuss, and correct any areas of misalignment. After establishing consensus on codes for several transcripts coded by multiple individuals, coders coded transcripts individually.

During analysis, we focused on areas of successes or concerns that were shared across programs, as well as identifying when a unique issue in a single program was mentioned by multiple interviewees. We triangulated emergent themes by reviewing multiple interviewees' input and examining documents and coaching logs.

Coaching Log Analysis

When calculating summary averages presented in the main report, we first averaged coach ratings by teacher in the given time frame and then across teachers. In this way, each observed teacher's rating received equal weight in the overall average. We chose this method of averaging because coaches had the flexibility to choose which teachers they reported on and sometimes did not report on teachers every week. Thus, instead of weighting results by coaches that were more diligent in filling out logs or by teachers more consistently rated, we provide unweighted averages that give each teacher-coach unit equal weight.

Impact Study Methods

The impact analysis explores TEACh's progress toward fulfilling three of their goals. One goal, as an alternative teacher pipeline program, was to fill hard-to-staff teaching positions such as special education (SPED), English language learner (ELL), and secondary math teaching positions. Another goal is to recruit a more diverse teaching force in terms of the race and ethnicity of the teacher candidates. To understand TEACh's progress in these dimensions, we performed descriptive analyses where we look at the proportion of new teachers recruited by TEACh that further the staffing goal and compare it to the proportion of new teachers recruited by the district through their normal course of business in the same program and year that also contribute to the staffing goal. These results are presented in the main body of the report.

A second goal is to train teachers to improve student learning. To this end, we conducted statistical analyses that compared the effectiveness of TEACh teachers to the effectiveness of teachers recruited by the program through their normal channels in the same year. We measure effectiveness as the ability to improve student performance on standardized tests of English Language Arts (ELA) and math. In the 2018–2019 school year, we were able to gather performance on state standardized tests in all districts. The COVID-19 pandemic caused all states to cancel these tests in the 2019–2020 school year. In only one of our programs were we able to obtain an alternative measure of student achievement, the Measures of Academic Progress (MAP), for that school year. In this program, the MAP is traditionally given as a formative assessment three times a year: in the fall, winter, and spring. During the 2019–2020 school year, only the fall and winter administrations were completed before schools closed due to the pandemic. Thus, we were able to look at half-year effects in the program that year. Table A.4 summarizes the contributions of each program to each academic outcome. This is a re-creation of Table 1.1 in the main body. Ultimately, we are interested in a single estimate of the differential effect of TEACh teachers on student achievement after one and two years in the classroom. Where possible, we take program and cohort estimates and perform a fixed-effect meta-analysis to create an overall estimate of effectiveness. This method weights each constituent estimate by its precision (Cooper, Hedges, and Valentine, 2009).[1]

Our last set of analyses seeks to understand if the short-term retention of TEACh teachers differs from their peers hired in the same year through traditional district recruitment channels. We define short-term retention as staying in the district after one and two years of service. We did not define recruitment as staying in a school, as the TEACh program recruited candidates for the district, not particular schools. Once again, where appropriate,

[1] We believe a fixed-effect meta-analytic framework is a reasonable approach here to estimate the average effect of the intervention in this sample of districts. Because the sample of districts in this study is small, just four districts across three programs, it is not well suited to using a random-effect model to generalize the results to other populations.

fixed-effect meta-analytic averages were calculated to provide summative estimates of differential retention. We are able to look at teacher retention in only two programs due to the delays in the release of data in one program. Table A.4 summarizes which programs contribute to the retention analyses.

Table A.4. Summary of TEACh Teachers Included in Analyses of Impacts on Retention and Student Achievement

	Retention Outcomes	State Assessments	MAP Assessments
Districts/Regions/Programs	Programs A and C	Programs A, B, and C	Program A
TEACh Cohorts and Years	Cohort 1 Years 1 and 2; Cohort 2 Year 1	Cohort 1 Year 1	Cohort 1 Years 1 and 2; Cohort 2 Year 1
N of TEACh teachers	144	73	101

NOTE: Program B retention outcomes for cohort 1, year 1, are not yet available but will be included in a future, updated version of this report.

TEACh Impacts on Student Achievement

To estimate the differential effect of TEACh teachers on increasing student achievement, we compare achievement outcomes of students taught by TEACh teachers to students taught by teachers hired in the same district and year but through the district's traditional recruitment channels. This study analyzes the program in the 2018–2019 and 2019–2020 school years. We label the TEACh teachers recruited in 2018–2019 to be *cohort 1* teachers. We were able to estimate *one-year* effects across all districts and *two-year* effects for the single district that had alternative student achievement data in the 2019–2020 school year. We label the TEACh teachers recruited in the 2019–2020 school year as *cohort 2* teachers and were able to look at *one-year* effects in the district for which we have alternative assessments of student achievement. We consider all teachers hired into a district in a school year as "first-year teachers" because we are unable to determine prior teaching experience in other districts.

When estimating the effect on ELA outcomes, we look at students taught by TEACh teachers in ELA classrooms and compare their performance on assessments to students taught by other teachers hired in the same year and program in ELA classrooms. Separate samples were created with math classrooms when looking at math outcomes. It is possible that students taught by TEACh teachers can also be taught by non-TEACh teachers, either across subjects, across years, or both. To avoid any spillover of the effect of TEACh teachers to the comparison group, we remove any student that was ever taught by a TEACh teacher from the pool of comparison students.[2]

Naive comparisons of student achievement between TEACh and non-TEACh teachers will likely produce biased estimates. Teachers and students are not randomly assigned to each

[2] These are a relatively small number of students. Models that do not make this restriction produce qualitatively similar results with no changes in inferences.

other, and TEACh programs made an effort to recruit candidates in hard-to-staff positions, exacerbating this nonrandom assignment. In Programs A and B, where the number of students and teachers were large, to guard against this type of bias, we instituted propensity score weights as part of our models (Rosenbaum and Rubin, 1983). To create the propensity scores, we predict the probability that a student was taught by a TEACh teacher in their ELA or math class based on their observable characteristics such as grade, prior achievement in ELA and math, race/ethnicity, special program participation, etc. Please see Table A.5 for a full list of variables used in the model. This probability is called the propensity score. We then calculated an average treatment on the treated (ATT) effect by using the propensity score to weight comparison-group students to look like treatment students. With this method, students in the comparison group that look more like the students in the treatment group are given more weight. Those that look more different are given less weight. To estimate the propensity score, we employed a machine learning technique called generalized boosted models (GBMs; Ridgeway, Madigan, and Richardson, 1999). GBMs are a nonparametric approach where the machine learning algorithm chooses a functional form to minimize the chosen test statistic. The flexibility of the approach allows for interaction and nonlinear terms and removes the responsibility of choosing the correct functional form from the analyst.[3]

If the propensity score is specified correctly, adding weights should eliminate bias due to the observable variables in the propensity score model. In addition to including the weights in our regressions, we also include the variables in the propensity score model as covariates in our outcome regressions to produce "doubly robust" estimates. Including the controls will increase precision if the propensity score weights completely account for bias based on the observables. If there is lingering bias that can be accounted for using observable variables, the controls will account for the remainder of the bias while minimizing the risk of extrapolating between treatment and control groups because the propensity weighting will increase the region of common support. However, there may still be bias in our estimates due to differences in comparison groups that we cannot observe with our available data. Table A.5 also lists the variables included as controls in our model.

In Program C, the number of teachers and students were too small to achieve balance on key variables including baseline student achievement levels via the doubly robust propensity score models. Thus, in that district we use ordinary least-squares models, with a parsimonious set of controls, to analyze ELA outcomes. Table A.5 indicates which control variables are included in those models. Samples were too small to give reliable results in math. We therefore caution the reader that the ELA results in Program C may suffer from omitted variables bias not seen in the estimates from the two other programs.

---

[3] To employ the GBMs, we use the TWANG package in Stata (Griffin et al., 2020). We allow for 5,000 iterations and three-way interactions. The program minimizes the mean effect size of the covariates in the model.

Table A.5. Variables in Student Achievement Propensity Score Models and Outcome Regressions, by Program

	Programs A and B		Program C
	Propensity Score Model	Outcome Regression	Outcome Regression (ELA Only)
Student Covariates			
Grade Level	✓	✓	✓
Black	✓	✓	
Asian	✓	✓	
Hispanic	✓	✓	
Other Race/Ethnicity	✓	✓	
Female	✓	✓	
ELL	✓	✓	
SPED	✓	✓	
Econ. Disadvantaged	✓	✓	
Gifted[a]	✓	✓	
Baseline Attendance	✓	✓	
Baseline ELA Achievement	✓	✓	✓
Baseline Math Achievement	✓	✓	✓
Teacher-Level Covariates			
Proportion of Students SPED	✓	✓	✓
Proportion of Students ELL	✓	✓	✓
School-Level Covariates			
Proportion Black		✓	
Proportion Hispanic		✓	
Proportion Asian		✓	
Proportion Other Race/Ethnicity		✓	
Proportion Female		✓	
Proportion ELL		✓	
Proportion SPED		✓	
Proportion Econ. Disadvantaged		✓	
School Level		✓	
Baseline ELA Achievement	✓	✓	
Baseline Math Achievement	✓	✓	
Baseline Attendance		✓	

NOTES: SPED = special education, ELL = English language learner.
[a] indicates only available in Program A. All districts provided a district-specific measure of being economically disadvantaged. In one district in Program C, that was an indicator for free or reduced-price lunch eligibility; for all other districts, it was a district-specific measure of economic disadvantage.

The final model takes the following form:

$$Y_{ipst} = \beta_0 + \beta_1 TEACh_{pst} + \boldsymbol{X_{ipst}\beta_2} + \boldsymbol{C_{pst}\beta_3} + \boldsymbol{S_{st}\beta_4} + \varepsilon_{ipst}, \quad (1)$$

where Y_{ipst} represents the ELA or math outcome of student, *i,* of teacher, *p*, in school, *s*, in year, *t*. $TEACH_{pst}$ is an indicator for whether the teacher was a part of the TEACh program. $\boldsymbol{X_{ipst}}$, $\boldsymbol{C_{pst}}$, and $\boldsymbol{S_{st}}$ are vectors of student characteristics, teacher characteristics, and school characteristics, respectively. The contents of each vector are detailed in Table A.5. Finally, ε_{ipst} is a student-level stochastic error term. Districts provided state standardized test scores in scale score units and MAP scores as Rasch Unit (RIT) scores. For each outcome, we standardized by grade and district to aid an effect-size interpretation. All regressions included the aforementioned propensity score weights.[4] It is worth noting that we did not control for any characteristics of teachers themselves, such as demographic characteristics, because the TEACh program was designed to recruit teachers who are more diverse and can staff certain positions. Thus, we wanted to program estimates to include the effects of these characteristics to capture the total effect of a program that has these stated goals. However, we did control for classroom-level student characteristics, such as the proportion of SPED and ELL students, as well as formal SPED and ELL roles to account for any systematic differences in students that TEACh teachers taught due to the efforts to recruit for these hard-to-staff roles. Standard errors were clustered at the teacher level to account for the correlation of student outcomes by teacher.

Table A.5 shows that more school-level characteristics were added as controls but were not found included in the propensity score models. Ideally, all school-level covariates would have been added to the models to ensure that the schools of students of TEACh teachers looked similar to the schools of students of non-TEACh teachers. However, we found that the addition of too many school-level variables resulted in worse overall balance of covariates. We therefore included a more parsimonious set of variables and controlled for the full set. We present balance statistics on all variables in Appendix B and find that this approach provides good balance on all covariates, regardless of whether they were in the propensity score model. These results provide some measure of comfort that the propensity score model eliminated major sources of bias.

Last, as with most administrative data, some students were missing values for some covariates. To avoid excluding these students, we impute covariate data using the multiple imputation, -mi-, command in Stata. The command executes a multiple imputation routine that is based on the concepts of Rubin (1987) and Schafer (1997). Multiple imputation is a regression-based imputation method that uses the known information about students in the data set to impute values for the missing data. To capture uncertainty in the imputed values, the routine creates five imputations and aggregates the results.

[4] Propensity score weights were added to OLS models in Stata using the weighting option. Standard errors are adjusted to account for the weights.

The estimation approach delineated above is very similar to that used by Kaufman et al. (2020) in their analysis of the TEACh program in three other urban districts. We chose a similar analysis, in part, to aid comparisons of effects across studies of the iterations of the program in different urban districts.

TEACh Impacts on Teacher Retention

A similar analytic approach was taken when analyzing the effect of the TEACh program on teacher retention in Programs A and C where data were available.[5] Teachers were defined as retained if they were present in the district after one or two years in the classroom. Our second-year retention outcome is a cumulative measure so as to include all teachers who entered in a program in a cohort. We chose district retention, and not school retention, as the outcome of interest because the TEACh program is meant to recruit candidates for the district, not for particular schools.

Naive comparisons of retention rates of TEACh teachers to peers hired in the same program and year by traditional channels would be biased for the same reasons as the student achievement analyses above—namely, students and teachers are not randomly assigned to each other. We mitigate this bias in the same way. First, we created teacher-level propensity scores for being a TEACh teacher via GBM models. We then leveraged doubly robust regression models that include the variables in the propensity score models as controls in addition to weighting the regression by the propensity score. Table A.6 delineates the variables used in our propensity score models and those used as controls in the outcome regression equation. Once again, additional school-level variables were added as controls to obtain the best balance on covariates overall.

Teacher retention models take the following form:

$$Y_{ist} = \beta_0 + \beta_1 TEACh_{ist} + \boldsymbol{C_{ist}\beta_3} + \boldsymbol{S_{st}\beta_4} + \varepsilon_{ipst} \qquad (2),$$

where Y_{ist} represents the first- or second-year retention of teacher, *i,* in schools, *s*, in year, *t*. $TEACh_{ist}$ is an indicator for whether the teacher is in the TEACh program. $\boldsymbol{X_{ist}}$**,** $\boldsymbol{C_{ist}}$**,** and $\boldsymbol{S_{st}}$ are vectors of classroom and school-level characteristics, respectively, listed in Table A.6. Finally, ε_{ist} is a teacher-level stochastic error term. All models included robust standard errors. As in the student academic achievement models, we do not account for teacher-level characteristics such as demographic variables and certifications so as to capture the recruitment priorities of the program as part of the impact estimate. However, similar to the student achievement analyses, we control for special program participation at the classroom level and indicators for being an ELL or SPED teacher in order to ensure that the characteristics of students taught are balanced between TEACh and non-TEACh students. Last, the same multiple imputation procedure was used to account for missing covariate data.

[5] We are able to leverage the doubly robust propensity score method in Program C for teacher retention because the teacher sample was not limited to ELA and math teachers in tested grades.

Table A.6. Variables in Teacher Retention Propensity Score Models and Outcome Regressions, by Program

	Programs A and C	
	Propensity Score Model	Outcome Regression
Classroom Covariates		
Grade Level	✓	✓
Proportion Black	✓	✓
Proportion Asian	✓	✓
Proportion Hispanic	✓	✓
Proportion Other Race/Ethnicity	✓	✓
Proportion Female	✓	✓
Proportion Econ. Disadvantaged	✓	✓
Baseline Attendance	✓	✓
Baseline ELA Achievement	✓	✓
Baseline Math Achievement	✓	✓
Proportion of Students SPED	✓	✓
Proportion of Students ELL	✓	✓
School-Level Covariates		
Proportion Black	✓	✓
Proportion Hispanic	✓	✓
Proportion Asian		✓
Proportion Other Race/Ethnicity		✓
Proportion Female		✓
Proportion ELL		✓
Proportion SPED		✓
Proportion Econ. Disadvantaged	✓	✓
School Level		✓
Baseline ELA Achievement	✓	✓
Baseline Math Achievement	✓	✓

NOTES: SPED = special education, ELL = English language learner. Program B was unable to provide data for teacher retention analyses.

Accounting for Multiple Comparisons

As part of the impact study, we estimated the effect of student and teacher outcomes on many samples and across two years. For the student achievement outcomes, this was exacerbated given the incomplete student achievement data in the 2019–2020 school year due to the COVID-19 pandemic. Conducting multiple statistical tests on the same outcome can increase the likelihood of detecting an effect by chance, even if no effect exists. To guard against these false positives, we delineate our primary and secondary estimates as follows:

Student Achievement:

- When choosing among samples contributing to year 1 or year 2 effects, the primary estimate is the one to which the greatest number of students across all programs contribute. For first-year effects, this is the meta-analytic average of cohort 1 on state standardized tests in Programs A and B in the 2018–2019 school year. Results from Program C were not reliable and thus were not included in the meta-analytic average. For second-year effects, this estimate is the estimate in Program C when looking at the MAP assessment.
- Our primary outcome of interest is second-year effects because that represents the effects of TEACh teachers after receiving the full training curriculum. We acknowledge that the year 2 evidence in our study is based on one program due to unforeseen complications from the pandemic.

Teacher Retention:

- We take the same approach in the teacher retention outcomes. When choosing among the samples contributing to year 1 or year 2 effects, the primary estimate is the one to which the greatest number of students across all cohorts and programs applies. In the case of year 1 estimates, this estimate is the meta-analytic average of the first-year effects across all programs and both cohorts. For year 2 effects, this estimate is the meta-analytic average across all programs for cohort 1.
- Our primary outcome of interest is second-year effects for the same reason articulated in the student achievement section. Further, since retention is a cumulative measure, it does not suffer from the same selection issues as the student achievement analysis.

We consider the remainder of the outcomes presented in the main body to be exploratory analyses.

Correlations Between Coaching and Student Achievement Gains

In additional exploratory analyses, we investigated whether the specific coaching practices described in coach logs were associated with variation in the adjusted student achievement of coached teachers. When calculating correlations between coaching ratings and student outcomes, we aggregate student achievement by teacher. The outcome measure is overall measure of achievement calculated by averaging ELA and math performance on state standardized tests. We then regress the spring average performance measure on the baseline ELA and math scores, district fixed effects, school level, and the coaching log measure of interest. Each measure is included in separate regressions. All continuous variables were standardized within district to provide an effect-size interpretation. Models included robust standard errors. These exploratory results are shown in Table A.7. These estimates are correlational in nature and should not be interpreted as the effects of coaching on teacher performance. They may in some cases indicate responses of coaches to variation in teacher performance.

Table A.7. Associations Between Coaching Behaviors and Teachers' Adjusted Student Achievement, Across TEACh Teachers

Log Variable	ELA and Math Compoprogram Outcome
Hours Coach Planned/Worked with Teacher	0.083*
	(0.032)
Hours Coached Spent on Administration	0.012
	(0.04)
Hours Coached Spent on Own Personal Development	–0.067
	(0.037)
Hours Coach Assisted Others in School	–0.061
	(0.037)
Hours Spent Planning or Provide Personal Development	–0.025
	(0.052)
Hours Spent Seeking Resources	–0.033
	(0.045)
Hours Spent Traveling	0.065
	(0.039)
Hours on Other	0.026
	(0.043)
Minutes Worked with Teacher per Week	–0.009
	(0.042)
Advise on Positive Learning Environment	–0.059
	(0.036)
Hours Noninstructional Issues	–0.069*
	(0.032)
Teacher Knowledge of Content or Standards	0.015
	(0.044)
Co-teach with Teacher	–0.041
	(0.039)
Co-plan Lessons	–0.025
	(0.031)
Formal Evaluation	0.053
	(0.032)
Talk About Less Pre- or Post-Observation	–0.021
	(0.041)
Connect Teacher and Resources	–0.034
	(0.04)
Discuss Curriculum	0.009
	(0.041)
Discuss Classroom Management	–0.034
	(0.038)

Log Variable	ELA and Math Compoprogram Outcome
Discuss Differentiation	–0.084*
	(0.038)
Discuss Assessment of Learning	–0.024
	(0.032)
Discuss Time Management	–0.06*
	(0.03)
Discuss Technology	–0.062
	(0.033)
Discuss Goal Setting	–0.042
	(0.046)
Guide Reflection on Practices	–0.069
	(0.041)
Discuss Grading Course Modules or Online Content	0.017
	(0.026)
Model Practice for Teacher	–0.022
	(0.045)
Observe Instructional Practice	–0.096**
	(0.034)
Observe Modeling a Practice	–0.061
	(0.033)
Provide Emotional Support	–0.061
	(0.041)
Review Data or Student Work	0.074*
	(0.036)
Provided Small-Group/Individual Support to Students	–0.065*
	(0.025)
Other	–0.041
	(0.053)
Did Teacher Improve on Practice from Last Time?	–0.001
	(0.029)
Interact via Phone	–0.05
	(0.034)
Interact via Text	–0.026
	(0.035)
Interact in Person	–0.005
	(0.038)
Interact via Email	–0.039
	(0.036)
Interact in Other Way	0.024
	(0.054)

Log Variable	ELA and Math Compoprogram Outcome
Interact via Email	–0.03
	(0.04)
Maximum Number of Observations	68

NOTES: Robust standard errors in parentheses. Each cell reports the results of a separate regression of the compoprogram student achievement outcome on the coaching practice of interest and controls. Controls include one-year lagged measure of outcomes in ELA and math, school level (elementary or middle school), and district fixed effects. Test score outcomes are standardized. * indicates $p < 0.05$, ** $p < 0.01$.

Limitations

The above analyses have several limitations. First, none of our analyses in isolation or in combination can evaluate the full effect of the TEACh program on the overall recruitment pipeline of each district in each program because the recruitment efforts of the TEACh program affected the traditional recruitment efforts of the district. For example, the efforts of TEACh to lower barriers into the teaching profession, recruit a more diverse teaching force, and to recruit for hard-to-staff positions may have attracted candidates to the applicant pool that otherwise may not have been available. The selection of these candidates by TEACh from the larger pool may leave more traditionally competitive teachers available for the remainder of the positions the district recruited for through traditional channels. A true understanding of how TEACh affected the teacher pipeline would require application data in years prior to and after the institution of TEACh to understand how the presence of the program shifted applications and selections. Unfortunately, that type of detailed data was not available from programs or districts.

Second, the use of propensity score weighting and doubly robust estimators can account for a substantial amount of bias compared to naive OLS estimates. However, the main limitation of this method is that it can only account for bias from observable student, teacher, and school characteristics. The key assumption is that accounting for these observable characteristics also accounts for *all* unobservable differences between students, teachers, and schools. This is a somewhat strong assumption that can never be fully tested. It is possible that students, teachers, and schools differ in unobserved ways that confound our estimates. The mission of TEACh to recruit for hard-to-staff positions exacerbates this problem. Because TEACh was successful in meeting this goal, fewer non-TEACh teachers were hired in the same year and program in similar roles that taught similar students. This is particularly salient for ELL and SPED teachers who teach students that generally have weaker academic outcomes and face a special set of needs compared to the general education population. With fewer teachers in the comparison group in these positions, the propensity score weighting algorithm can have more difficulty in weighting the comparison group to look similar to the treatment group along these dimensions and on all other observable (and therefore unobservable) variables. Further, due to imperfect data, we were not able to match all teachers to students with the student-link files, and we must also assume that this rate of missing is unrelated to TEACh status after applying the propensity score weighting model and controlling for covariates.

Third, despite pooling results across three programs, the sample sizes are relatively modest, especially of TEACh teachers. This limitation is of greatest concern in Program C, which is composed of two relatively smaller districts. The small sample sizes in Program C had a direct effect on our ability to estimate student achievement outcomes because limiting teachers to those in tested grades and subjects, the number of TEACh teachers was extremely small. For this reason, we used a parsimonious set of covariates in OLS models with no propensity score weighting when looking at ELA outcomes. Sample sizes were so small in math that we could not reliably estimate effects, even in these parsimonious OLS models. We caution the reader when looking at ELA estimates in Program C, as they can contain more bias than estimates in Programs A and B. We were able to use our full propensity score weighting methodology in Program C for teacher retention outcomes, as we were able to incorporate TEACh teachers who did not teach in tested grades and subjects. Even pooling estimates across programs, the relatively small overall sample sizes means that we cannot rule out economically meaningful effects of the TEACh program in cases where we found statistically insignificant effects. In addition, when pooling effects across programs, a limitation of our approach is that we put more weight on some districts than others, with our results more heavily weighted toward districts (typically larger districts) where our estimates are more precise.

Fourth, even in cases where we do pool estimates across programs, we are not attempting to generate an estimate that would necessarily generalize to a broader population of districts around the country. With a sample of just three programs spanning four districts, we did not have sufficient sample to estimate the degree of variability in typical program effects that may occur across district contexts, but we acknowledge that such variability in impact across programs likely exists. Instead, our focus is on estimation of the average effect in this instance of TEACh spanning these particular programs. Moreover, it is worth keeping in mind that this average effect does not weight all three programs equally but instead weights some programs (generally Programs A and B) more than others, specifically those where our estimates are more precise.

Finally, this study was able to look at only the short-term effects of the TEACh program. Year 1 effects are estimated from two cohorts, and year 2 effects are estimated from only one cohort. Year 1 effects could be smaller than the true programmatic effects because teachers are still being trained while in the classroom. Year 2 effects represent the first year after training is complete, but some TEACh (and comparison) teachers have left the classroom. In our study in particular, our ability to analyze year 2 student achievement effects was limited because of the lack of state standardized testing in year 2019–2020 due to the COVID-19 pandemic. Overall, our findings indicate whether TEACh teachers were more effective than other novice teachers, but we cannot speak to whether there is a differential effectiveness between TEACh teachers and those with similar experience in the longer term.

Appendix B. Additional Impact Estimates

In this appendix, we provide additional descriptive statistics, model estimates, and covariate balance checks. For the student outcome estimates, the main MAP estimates use fall administrations of the assessment in the same year as the baseline measure in lieu of the spring administration of the prior year. In one sense, the fall baseline measure may control for some of the TEACh effect if there are any that accumulate in the beginning of the school year. On the other hand, the fall baseline measure allows us to include all students in tested grades in the analysis without imputing baseline values of the covariate for students in the youngest grade who do not have scores in the previous spring. Point estimates are similar when using the two baseline measures. In this appendix, we provide estimates from both models.

Table B.1. Teacher Characteristics, Cohorts Pooled, by Program

	All Programs		Program A		Program C	
	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh
N	169	1,615	137	937	32	678
Female	75	65	77	71	69	58
Black	23	15	23	21	22	8
Hispanic	4	9	4	8	6	10
White	62	71	63	65	56	79
Other Race/Ethnicity	10	5	9	6	16	3
SPED Teacher	10	13	1	9	50	19
ELL Teacher	9	4	4	3	34	5

NOTES: SPED = special education, ELL = English language learner. Program B was unable to provide demographic data on teachers. Only TEACh teachers and other new hires are included in the table. All statistics combine cohorts 1 and 2 hired in school years 2018–2019 and 2019–2020, respectively.

Table B.2. Sample Characteristics, Cohort 1, Year 1, Outcomes with State Tests

	Program A				Program B			
	ELA		Math		ELA		Math	
Variable	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh
Student Characteristics								
Grade	4.46	4.56	4.38	4.6	3.86	5.46	5.77	5.39
Black	23	23	23	23	51	45	51	44
Asian	2	1	2	1	3	5	4	5
Hispanic	42	43	43	40	33	31	32	30
Other Race/Ethnicity	12	13	13	14	3	2	2	2
Female	49	50	49	50	48	49	48	48
ELL	34	37	35	33	17	16	15	18
SPED	19	21	19	23	6	9	9	10
Econ. Disadvantaged	85	81	87	81	54	46	49	46
Prior Year ELA Achievement	–0.13	0.02	–0.18	0.02	–0.29	–0.15	–0.32	–0.2
Prior Year Math Achievement	–0.07	0.04	–0.19	0.03	–0.35	–0.06	–0.18	–0.14
Prior Year Attendance	94	94	94	94	95	95	95	95
Classroom Characteristics								
Percentage SPED	22	22	21	23	7	9	10	11
Percentage ELL	28	32	31	28	16	15	15	17

Variable	Program A				Program B			
	ELA		Math		ELA		Math	
	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh
School-Level Characteristics								
Percentage Black	24	22	24	23	51	43	46	42
Percentage Hispanic	41	42	43	39	30	30	29	28
Percentage Asian	2	1	2	1	5	6	6	7
Percentage Other Race/Ethnicity	14	14	14	15	3	3	3	3
Percentage Female	50	49	49	49	49	49	49	49
Proportion ELL	33	32	35	30	16	15	13	15
Proportion SPED	18	17	19	17	9	9	10	9
Percentage Econ. Disadvantage	82	77	83	77	51	45	46	43
Elementary School	93	93	100	89	62	40	40	45
Prior Year ELA Achievement	0.28	0.32	–0.19	–0.02	–0.22	–0.11	–0.11	–0.07
Prior Year Math Achievement	–0.13	–0.03	–0.16	–0.03	–0.16	–0.02	–0.07	0.02

NOTES: SPED = special education, ELL = English language learner. All results are unweighted.

Table B.3. MAP Sample Characteristics, Year 1 and 2 Effects, by Cohort

	Year 1 Effect Samples												Year 2 Effect Sample			
	Cohort 1 State Test Sample				Cohort 1 Full Sample				Cohort 2 Full Sample				Cohort 1 Full Sample			
	ELA		Math		ELA		Math		ELA		Math		ELA		Math	
Variable	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh
Student Characteristics																
Grade	4.46	4.56	4.38	4.6	3.54	3.35	3.34	3.29	3.15	3.28	3.44	3.89	3.28	3.06	3.07	3.04
Black	23	23	23	23	25	23	25	22	28	22	28	23	22	22	21	22
Asian	2	1	2	1	2	1	2	1	1	2	1	2	2	1	2	1
Hispanic	42	43	43	40	44	42	44	42	42	42	42	40	41	44	40	44
Other Race/Ethnicity	12	13	13	14	11	14	13	14	13	14	12	15	17	15	17	15
Female	49	50	49	50	49	50	49	49	52	49	52	48	48	48	47	48
ELL	34	37	35	33	37	35	37	35	34	35	33	31	36	36	33	36
SPED	19	21	19	23	18	18	18	19	15	16	15	16	19	15	20	15
Econ. Disadvantaged	85	81	87	81	86	82	88	80	92	83	90	84	88	86	88	85
Fall ELA Achievement	–0.13	0.02	–0.18	0.02	–0.12	0.01	–0.17	0.02	–0.12	–0.13	–0.16	–0.06	–0.10	–0.01	–0.19	–0.06
Fall Math Achievement	–0.07	0.04	–0.19	0.03	–0.07	0.04	–0.19	0.04	–0.15	–0.08	–0.15	–0.03	0.01	–0.03	–0.11	–0.06
Prior Year Attendance	94	94	94	94	94	94	93	94	93	94	93	94	92	94	93	94
Classroom Characteristics																
Percentage SPED	22	22	21	23	19	19	18	19	15	15	15	15	18	15	20	15

	Year 1 Effect Samples												Year 2 Effect Sample			
	Cohort 1 State Test Sample				Cohort 1 Full Sample				Cohort 2 Full Sample				Cohort 1 Full Sample			
	ELA		Math		ELA		Math		ELA		Math		ELA		Math	
Variable	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh
Percentage ELL	28	32	31	28	32	32	35	32	33	33	33	30	35	37	36	37
School-Level Characteristics																
Percentage Black	24	22	24	23	24	23	25	22	28	24	27	24	23	24	23	24
Percentage Hispanic	41	42	43	39	42	42	43	40	37	37	37	37	38	39	38	39
Percentage Asian	2	1	2	1	2	1	2	1	2	2	2	2	2	1	2	1
Percentage Other Race/Ethnicity	14	14	14	15	14	15	14	15	16	15	16	16	16	16	16	16
Percentage Female	50	49	49	49	49	49	49	49	49	49	49	49	49	49	49	49
Proportion ELL	33	32	35	30	34	32	35	31	30	28	29	24	31	31	32	31
Proportion SPED	18	17	19	17	18	17	18	17	18	18	19	18	18	17	18	17
Percentage Econ. Disadvantage	82	77	83	77	83	78	84	77	92	82	92	83	89	87	90	86
Elementary School	93	93	100	89	94	95	100	93	98	85	93	68	92	95	100	96
Fall ELA Achievement	–0.13	–0.03	–0.19	–0.02	–0.17	–0.04	–0.23	–0.02	–0.19	0.00	–0.19	–0.01	–0.15	–0.06	–0.16	–0.06
Fall Math Achievement	–0.10	–0.03	–0.16	–0.03	–0.15	–0.04	–0.21	–0.02	–0.2	0.01	–0.2	0.00	–0.12	–0.07	–0.14	–0.07

NOTES: SPED = special education, ELL = English language leaner. All results are unweighted.

Table B.4. Teacher Retention Sample Characteristics, by Cohort

	Cohort 1				Cohort 2			
	Program A		Program C		Program A		Program C	
	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh	TEACh	Non-TEACh
Classroom Characteristics								
Grade	3.42	3.89	6.91	5.63	3.00	4.21	6.56	6.05
Percentage Black	28	27	14	12	25	26	14	9
Percentage Asian	2	1	3	1	1	1	1	1
Percentage Hispanic	36	36	77	76	40	37	76	79
Percentage Other Race/Ethnicity	15	16	1	2	16	16	2	1
Percentage Female	49	47	34	46	51	49	34	46
Percentage Econ. Disadvantaged	92	88	72	75	91	84	90	85
Baseline ELA Achievement	–0.32	–0.29	–0.76	–0.22	–28	–33	–0.79	–0.21
Baseline Math Achievement	–0.36	–0.25	–0.66	–0.18	–27	–29	–0.77	–0.21
Baseline Attendance	93	92	93	93	92	92	93	92
Percentage SPED	18	25	56	37	16	25	51	27
Percentage ELL	29	26	52	27	33	28	33	24
School-Level Characteristics								
Percentage Black	27	26	12	12	27	28	13	10
Percentage Hispanic	36	35	77	78	37	35	78	77
Percentage White	27	26	7	8	19	20	7	11
Percentage Female	36	35	47	47	49	48	50	49
Proportion ELL	30	25	34	27	29	23	31	25
Proportion SPED	18	19	29	29	19	19	27	26
Percentage Econ. Disadvantage	91	88	71	72	91	86	73	65
Elementary School	83	70	14	48	88	59	17	44
Baseline ELA Achievement	–0.22	–0.12	–0.17	–0.12	–0.17	–0.06	–0.06	–0.02
Baseline Math Achievement	–0.22	–0.12	–0.13	–0.11	–0.18	–0.06	–0.03	–0.03

NOTES: SPED = special education, ELL = English language learner. Program B was unable to provide retention data. All statistics are unweighted.

Table B.5. Estimates of the Year 1 Effect of TEACh on Student Achievement, by Cohort, Program, and Test Outcome

	Baseline Measure	Meta-analytic Average		Program A		Program B		Program C	
		ELA	Math	ELA	Math	ELA	Math	ELA	Math
Panel A: Cohort 1 Year 1 (2018–2019)									
State Tests	Prior Year Spring	–0.012	–0.010	0.002	0.058	–0.021	–0.020	—	—
		(0.018)	(–0.015)	(0.029)	(0.042)	(0.024)	(0.016)	—	—
N(Students)				3,002	3,013	8,131	8,021	1,235	491
N(TEACh Teachers)				30	32	17	25	5	3
N(Non-TEACh Teachers)				97	99	215	205	24	14
Spring MAP (State Test Sample)	Same Year Fall	N/A		–0.023	0.005	N/A		N/A	
				(0.034)	(0.049)				
Spring MAP (State Test Sample)	Prior Year Spring	N/A		–0.016	0.033	N/A		N/A	
				(0.036)	(0.051)				
N(Students)				3,002	3,013				
N(TEACh Teachers)				30	32				
N(Non-TEACh Teachers)				97	99				

	Baseline Measure	Meta-analytic Average		Program A		Program B		Program C	
		ELA	Math	ELA	Math	ELA	Math	ELA	Math
Spring MAP (Full Sample)	Same Year Fall	N/A		–0.043	–0.018	N/A		N/A	
				(0.034)	(0.044)				
Spring MAP (Full Sample)	Prior Year Spring	N/A		–0.024	0.004	N/A		N/A	
				(0.033)	(0.046)				
N(Students)				4,371	4,587				
N(TEACh Teachers)				46	49				
N(Non-TEACh Teachers)				160	169				
Panel B: Cohort 2 Year 1 (2019–2020)									
Winter MAP (Full Sample)	Same Year Fall	N/A		0.022	–0.024	N/A		N/A	
				(0.041)	(0.033)				
Winter MAP (Full Sample)	Prior Year Spring	N/A		0.054	–0.017	N/A		N/A	
				(0.039)	(0.038)				
N(Students)				3,764	4,490				
N(TEACh Teachers)				43	44				
N(Non-TEACh Teachers)				119	129				

NOTES: Each estimate is the result of a separate regression of the relevant student achievement outcome on an indicator for being a TEACh teacher. Panel headers indicate cohort, and row and column headers indicate student achievement outcome. All models in Programs A and B include propensity score weights and student and school covariates. Due to sample-size limitations, Program C results were not calculated because sample sizes were too small. All test score outcomes are standardized by grade and district.

Table B.6. Estimates of the Year 2 Effect of TEACh on Student Achievement, Program A, MAP Assessment

	Baseline Measure	Program A	
		ELA	Math
Spring MAP (Full Sample)	Same Year Fall	0.014	0.076*
		(0.038)	(0.033)
Spring MAP (Full Sample)	Prior Year Spring	0.044	0.082+
		(0.044)	(0.046)
N(Students)		3,653	3,715
N(TEACh Teachers)		37	35
N(Non-TEACh Teachers)		129	127

NOTES: Each estimate is the result of a separate regression of the relevant student achievement outcome on an indicator for being a TEACh teacher. Panel headers indicate cohort, and row and column headers indicate student achievement outcome. All models in Programs A and B include propensity score weights and student and school covariates. Due to sample-size limitations, Program C uses OLS and a select number of covariates. Please see Appendix A for details on the model. Sample sizes are as follows: Program A: N(ELA State Sample) = 3,002, N(Math State Sample) = 3,013, N(Cohort 1 MAP Full Sample) = 4,371, N(Cohort 2 MAP Full Sample) = 4,587. Program B: N(ELA State Sample) = 4,253 , N(Math State Sample) = 9,196. Program C: N(ELA State Sample) = 1,300, N(Math State Sample) = 575. All test score outcomes are standardized by grade and district. Program C was not included in the meta-analytic average because propensity score weighted models were not used. + indicates $p < 0.01$, * $p < 0.05$.

Table B.7. The Effect of TEACh on Year 1 Teacher Retention, by Program

	Meta-analytic Average	Program A	Program C
Panel A: Cohort-Specific Results			
Cohort 1	–0.001	0.019	–0.096
	(0.056)	(0.062)	(0.136)
N(TEACh Teachers)		61	7
N(Non-TEACh Teachers)		324	147
Cohort 2	0.084	0.079	0.301
	(0.064)	(0.065)	(0.421)
N(TEACh Teachers)		58	18
N(Non-TEACh Teachers)		262	156
Panel B: Cross-Cohort Results			
Meta-analytic Average	0.036	0.048	–0.058
	(0.042)	(0.045)	(0.129)

NOTES: Each estimate is the result of a separate regression of an indicator for remaining in the district after the first year on being a TEACh teacher. Column headers indicate programs included in the analyses. All models include propensity score weights, and student and school covariates were used in program-specific regressions. Fixed-effect meta-analytic average models were used in column 1.

Table B.8. Year 2 Teacher Retention, by Program

	Meta-analytic Average	Program A	Program C
	0.054	0.120+	–0.148
Cohort 1	(0.061)	(0.070)	(0.122)
N(TEACh Teachers)		61	7
N(Non-TEACh Teachers)		324	147

NOTES: Each estimate is the result of a separate regression of an indicator for remaining in the district after the first year on being a TEACh teacher. Column headers indicate programs included in the analyses or meta-analytic model. All models include propensity score weights, and student and school covariates were used in program-specific regressions. Fixed-effect meta-analytic average models were used in column 1. + indicates $p < 0.10$.

Table B.9. Covariate Balance on Year 1 Effect with State Standardized Tests, by Program

	Program A, Cohort 1				Program B, Cohort 1			
	ELA (Unweight)	ELA (Weight)	Math (Unweight)	Math (Weight)	ELA (Unweight)	ELA (Weight)	Math (Unweight)	Math (Weight)
Student Covariates								
Grade 3	0.204	0.005	0.309	0.005	0.859+	0.222	–0.082	–0.256
	(0.239)	(0.270)	(0.267)	(0.314)	(0.514)	(0.565)	(0.245)	(0.282)
Grade 4	–0.150	–0.133	–0.115	–0.045	1.100*	0.689	0.279	0.005
	(0.227)	(0.241)	(0.207)	(0.216)	(0.487)	(0.545)	(0.354)	(0.400)
Grade 5	–0.012	0.119	–0.091	0.026	–0.054	–0.220	–0.224	–0.248
	(0.252)	(0.249)	(0.217)	(0.225)	(0.412)	(0.464)	(0.298)	(0.325)
Grade 6	–0.025	0.001	–0.070	0.012	—	—	–0.660**	–0.222
	(0.241)	(0.240)	(0.231)	(0.229)	—	—	(0.213)	(0.184)
Grade 8	—	—	—	—	—	—	0.671	0.641
	—	—	—	—	—	—	(0.428)	(0.444)
Black	0.004	0.002	–0.015	–0.069	0.132	–0.045	0.135	0.090
	(0.100)	(0.105)	(0.080)	(0.098)	(0.108)	(0.113)	(0.105)	(0.107)
Asian	0.068	0.064	0.068	0.081	–0.091*	–0.052	–0.058	–0.002
	(0.071)	(0.074)	(0.084)	(0.085)	(0.045)	(0.051)	(0.070)	(0.070)
Hispanic	–0.030	0.010	0.056	0.082	0.038	0.054	0.033	–0.038
	(0.123)	(0.126)	(0.104)	(0.112)	(0.123)	(0.126)	(0.091)	(0.099)
Other	–0.041	–0.042	–0.034	–0.011	0.012	0.005	0.002	–0.027
	(0.046)	(0.055)	(0.047)	(0.062)	(0.045)	(0.055)	(0.036)	(0.042)
Female	–0.017	–0.004	–0.019	–0.030	–0.011	0.012	0.009	–0.008
	(0.034)	(0.036)	(0.033)	(0.033)	(0.065)	(0.071)	(0.034)	(0.040)
ELL	–0.058	–0.049	0.040	0.017	0.027	–0.065	–0.080	–0.125
	(0.125)	(0.123)	(0.108)	(0.099)	(0.099)	(0.101)	(0.081)	(0.082)
SPED	–0.035	–0.081	–0.087	–0.104+	–0.089+	–0.115+	–0.037	–0.052
	(0.076)	(0.078)	(0.058)	(0.054)	(0.045)	(0.060)	(0.043)	(0.055)
Economically Disadvantaged	0.085	0.049	0.160	0.003	0.165*	0.026	0.063	–0.011
	(0.119)	(0.115)	(0.108)	(0.086)	(0.066)	(0.076)	(0.071)	(0.071)

	Program A, Cohort 1				Program B, Cohort 1			
	ELA (Unweight)	ELA (Weight)	Math (Unweight)	Math (Weight)	ELA (Unweight)	ELA (Weight)	Math (Unweight)	Math (Weight)
Gifted	–0.024	0.014	–0.138	–0.014	–0.262**	–0.060	–0.209**	–0.068
	(0.112)	(0.108)	(0.095)	(0.078)	(0.073)	(0.057)	(0.076)	(0.066)
Attendance	–0.148	–0.067	–0.203	–0.003	–0.136	–0.014	–0.126	–0.078
	(0.143)	(0.146)	(0.123)	(0.101)	(0.086)	(0.104)	(0.078)	(0.075)
Prior Year ELA	–0.110	–0.091	–0.215+	–0.099	–0.301**	–0.117	–0.035	0.042
State Test	(0.143)	(0.139)	(0.115)	(0.083)	(0.101)	(0.124)	(0.115)	(0.115)
Prior Year Math	–0.067	0.000	–0.047	0.056	0.008	0.046	–0.024	0.006
State Test	(0.062)	(0.065)	(0.050)	(0.061)	(0.045)	(0.065)	(0.043)	(0.049)
Teacher Characteristics								
Proportion SPED	–0.033	–0.168	–0.114	–0.150	–0.157	–0.150	–0.090	–0.075
Taught	(0.151)	(0.141)	(0.154)	(0.138)	(0.131)	(0.154)	(0.119)	(0.144)
Proportion ELL	–0.168	–0.107	0.130	0.126	0.075	–0.139	–0.076	–0.166
Taught	(0.215)	(0.200)	(0.216)	(0.202)	(0.232)	(0.237)	(0.186)	(0.188)
School Covariates								
Prior Year ELA	–0.207	–0.018	–0.367	–0.011	–0.287+	–0.063	–0.109	0.024
State Test	(0.222)	(0.210)	(0.230)	(0.184)	(0.173)	(0.155)	(0.220)	(0.212)
Prior Year Math	–0.135	0.005	–0.285	0.040	–0.339+	–0.034	–0.222	–0.035
State Test	(0.217)	(0.204)	(0.229)	(0.186)	(0.187)	(0.181)	(0.228)	(0.226)
Observations	2,203–3,002		2,152–3,013		7,487–8,131		7,397–8,021	

NOTES: Standard errors are clustered by teacher. Only weighted models include propensity score weights. All covariates are standardized within district. Economically disadvantaged in one district in Program C is an indicator for being eligible for free or reduced-price lunch. All variables are standardized. + indicates $p < 0.10$; * $p < 0.05$, ** $p < 0.01$.

Table B.10. Covariate Balance on Year 1 Effects with MAP Assessment, Program A Only, Fall Baseline

	Cohort 1 State Sample MAP Fall Baseline				Cohort 1 Full Sample, Fall Baseline				Cohort 2 Full Sample, Fall Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
Student Covariates												
Kindergarten	—	—	—	—	–0.154	–0.115	–0.188	–0.155	0.061	–0.058	0.067	–0.090
	—	—	—	—	(0.137)	(0.143)	(0.142)	(0.149)	(0.204)	(0.234)	(0.195)	(0.225)
Grade 1	—	—	—	—	–0.080	–0.075	–0.095	–0.085	–0.073	–0.128	–0.006	–0.136
	—	—	—	—	(0.169)	(0.187)	(0.171)	(0.186)	(0.190)	(0.217)	(0.184)	(0.216)
Grade 2	—	—	—	—	0.037	0.134	0.172	0.244	–0.066	–0.132	–0.022	–0.128
	—	—	—	—	(0.171)	(0.169)	(0.194)	(0.194)	(0.191)	(0.210)	(0.154)	(0.174)
Grade 3	0.204	0.145	0.309	–0.024	0.207	0.105	0.266	0.107	–0.046	–0.044	0.034	0.018
	(0.239)	(0.270)	(0.267)	(0.348)	(0.202)	(0.224)	(0.226)	(0.247)	(0.200)	(0.202)	(0.181)	(0.185)
Grade 4	–0.150	–0.111	–0.115	0.050	–0.088	–0.127	–0.077	–0.147	0.154	0.126	0.146	0.016
	(0.227)	(0.262)	(0.207)	(0.212)	(0.187)	(0.215)	(0.160)	(0.194)	(0.231)	(0.243)	(0.184)	(0.214)
Grade 5	–0.012	–0.027	–0.091	–0.164	0.030	–0.002	–0.050	–0.151	0.140	0.171	0.172	0.160
	(0.252)	(0.273)	(0.217)	(0.251)	(0.213)	(0.224)	(0.171)	(0.194)	(0.179)	(0.188)	(0.174)	(0.183)
Grade 6	–0.025	0.007	–0.070	0.151	0.020	0.059	–0.047	0.176	0.109	0.100	0.124	0.229
	(0.241)	(0.244)	(0.231)	(0.207)	(0.206)	(0.205)	(0.179)	(0.149)	(0.240)	(0.245)	(0.258)	(0.244)
Grade 7	—	—	—	—	—	—	—	—	–0.269+	–0.060	–0.477*	–0.093
	—	—	—	—	—	—	—	—	(0.146)	(0.085)	(0.241)	(0.208)

	Cohort 1 State Sample MAP Fall Baseline				Cohort 1 Full Sample, Fall Baseline				Cohort 2 Full Sample, Fall Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
Black	0.004	–0.015	–0.015	–0.140	0.041	–0.000	0.074	–0.000	0.150	0.084	0.134	0.093
	(0.100)	(0.116)	(0.080)	(0.112)	(0.081)	(0.094)	(0.070)	(0.080)	(0.098)	(0.106)	(0.096)	(0.100)
Asian	0.068	0.074	0.068	0.072	0.078	0.054	0.089	0.066	–0.048	–0.043	–0.048	–0.074
	(0.071)	(0.074)	(0.084)	(0.087)	(0.053)	(0.059)	(0.059)	(0.063)	(0.056)	(0.058)	(0.049)	(0.052)
Hispanic	–0.030	–0.060	0.056	0.145	0.029	0.005	0.038	0.048	–0.002	–0.028	0.047	0.026
	(0.123)	(0.140)	(0.104)	(0.133)	(0.094)	(0.097)	(0.084)	(0.087)	(0.099)	(0.101)	(0.098)	(0.097)
Other	–0.041	0.006	–0.034	–0.000	–0.092*	–0.099+	–0.052	–0.084+	–0.028	–0.040	–0.066	–0.106+
	(0.046)	(0.056)	(0.047)	(0.076)	(0.041)	(0.053)	(0.040)	(0.049)	(0.053)	(0.054)	(0.052)	(0.056)
Female	–0.017	0.001	–0.019	–0.032	–0.014	–0.019	–0.003	–0.018	0.044	0.039	0.079*	0.065
	(0.034)	(0.038)	(0.033)	(0.032)	(0.030)	(0.035)	(0.029)	(0.032)	(0.036)	(0.038)	(0.039)	(0.040)
ELL	–0.058	–0.099	0.040	0.045	0.032	0.014	0.055	0.041	–0.030	–0.076	0.045	–0.037
	(0.125)	(0.136)	(0.108)	(0.109)	(0.097)	(0.101)	(0.086)	(0.089)	(0.100)	(0.101)	(0.088)	(0.089)
SPED	–0.035	–0.082	–0.087	–0.072	0.008	–0.018	–0.027	–0.067	–0.026	–0.033	–0.018	–0.004
	(0.076)	(0.083)	(0.058)	(0.056)	(0.065)	(0.071)	(0.053)	(0.058)	(0.057)	(0.054)	(0.058)	(0.052)
Economically Disadvantaged	0.085	0.006	0.160	–0.013	0.112	0.006	0.192*	0.051	0.258**	–0.011	0.175*	0.008
	(0.119)	(0.112)	(0.108)	(0.092)	(0.089)	(0.081)	(0.078)	(0.059)	(0.089)	(0.041)	(0.086)	(0.064)
Gifted	–0.024	0.036	–0.138	–0.015	–0.016	0.043	–0.110	–0.024	–0.124+	0.003	–0.109	0.003
	(0.112)	(0.111)	(0.095)	(0.082)	(0.097)	(0.094)	(0.076)	(0.060)	(0.066)	(0.046)	(0.075)	(0.058)

	Cohort 1 State Sample MAP Fall Baseline				Cohort 1 Full Sample, Fall Baseline				Cohort 2 Full Sample, Fall Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
Fall ELA MAP	–0.126	–0.050	–0.123	–0.002	–0.094	–0.026	–0.114	0.016	0.008	0.073	–0.042	0.050
	(0.116)	(0.121)	(0.110)	(0.098)	(0.085)	(0.088)	(0.080)	(0.072)	(0.073)	(0.059)	(0.074)	(0.064)
Fall Math MAP	–0.092	–0.038	–0.151	–0.069	–0.051	0.026	–0.095	–0.009	–0.087	0.039	–0.097	0.006
	(0.115)	(0.112)	(0.103)	(0.081)	(0.085)	(0.085)	(0.076)	(0.065)	(0.068)	(0.052)	(0.074)	(0.063)
Attendance	–0.067	0.008	–0.047	0.026	0.010	0.086	–0.021	0.055	–0.171*	–0.015	–0.118	–0.018
	(0.062)	(0.063)	(0.050)	(0.052)	(0.054)	(0.061)	(0.051)	(0.057)	(0.075)	(0.091)	(0.077)	(0.083)
Teacher Characteristics												
Proportion	–0.033	–0.154	–0.114	–0.161	–0.002	–0.033	–0.068	–0.098	–0.029	0.003	–0.038	–0.008
SPED Taught	(0.151)	(0.141)	(0.154)	(0.133)	(0.125)	(0.122)	(0.122)	(0.116)	(0.141)	(0.125)	(0.123)	(0.115)
Proportion ELL	–0.168	–0.170	0.130	0.215	0.015	0.011	0.113	0.123	0.001	–0.026	0.130	0.039
Taught	(0.215)	(0.212)	(0.216)	(0.214)	(0.165)	(0.162)	(0.166)	(0.155)	(0.176)	(0.175)	(0.163)	(0.161)
School Covariates												
Fall ELA	–0.114	0.014	–0.165*	–0.025	–0.130*	0.003	–0.184**	–0.037	–0.041	0.034	–0.096	–0.008
MAP	(0.078)	(0.069)	(0.082)	(0.061)	(0.058)	(0.050)	(0.057)	(0.041)	(0.052)	(0.045)	(0.063)	(0.055)
Fall Math MAP	–0.073	0.041	–0.139+	0.004	–0.089	0.035	–0.149*	–0.010	–0.106*	0.008	–0.127*	–0.023
	(0.077)	(0.064)	(0.084)	(0.058)	(0.057)	(0.048)	(0.058)	(0.040)	(0.053)	(0.042)	(0.061)	(0.051)
Observations	2,739–3,002		2,742–3,013		3,958–4,371		4,045–4,587		3,026–3,764		3,367–4,490	

NOTES: Standard errors are clustered by teacher. Only weighted models include propensity score weights. All covariates are standardized within district. Economically disadvantaged in one district in Program C is an indicator for being eligible for free or reduced-price lunch. UW = unweighted , W = weighted. All variables are standardized.
+ indicates $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table B.11. Covariate Balance on Year 1 Effects with MAP Assessment, Program A Only, Spring Baseline

	Cohort 1, State Sample MAP, Prior Year Spring Baseline				Cohort 1, Full Sample, Prior Year Spring Baseline				Cohort 2, Full Sample, Prior Year Spring Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
Student Covariates												
Kindergarten	—	—	—	—	–0.154	–0.231	–0.188	–0.196	0.061	–0.022	0.067	–0.022
	—	—	—	—	(0.137)	(0.168)	(0.142)	(0.174)	(0.204)	(0.234)	(0.195)	(0.214)
Grade 1	—	—	—	—	–0.080	–0.120	–0.095	–0.066	–0.073	–0.142	–0.006	–0.092
	—	—	—	—	(0.169)	(0.199)	(0.171)	(0.215)	(0.190)	(0.223)	(0.184)	(0.205)
Grade 2	—	—	—	—	0.037	0.160	0.172	0.332+	–0.066	–0.152	–0.022	–0.165
	—	—	—	—	(0.171)	(0.163)	(0.194)	(0.191)	(0.191)	(0.215)	(0.154)	(0.182)
Grade 3	0.204	0.041	0.309	0.084	0.207	0.079	0.266	0.005	–0.046	–0.008	0.034	0.004
	(0.239)	(0.260)	(0.267)	(0.293)	(0.202)	(0.229)	(0.226)	(0.306)	(0.200)	(0.194)	(0.181)	(0.186)
Grade 4	–0.150	–0.066	–0.115	–0.102	–0.088	–0.081	–0.077	–0.099	0.154	0.158	0.146	–0.053
	(0.227)	(0.222)	(0.207)	(0.238)	(0.187)	(0.203)	(0.160)	(0.194)	(0.231)	(0.240)	(0.184)	(0.226)
Grade 5	–0.012	0.012	–0.091	–0.031	0.030	0.088	–0.050	–0.063	0.140	0.154	0.172	0.157
	(0.252)	(0.252)	(0.217)	(0.222)	(0.213)	(0.215)	(0.171)	(0.191)	(0.179)	(0.187)	(0.174)	(0.178)
Grade 6	–0.025	0.016	–0.070	0.059	0.020	0.040	–0.047	0.055	0.109	0.125	0.124	0.170
	(0.241)	(0.242)	(0.231)	(0.223)	(0.206)	(0.215)	(0.179)	(0.178)	(0.240)	(0.245)	(0.258)	(0.252)
Grade 7	—	—	—	—	—	—	—	—	–0.269+	–0.129	–0.477*	–0.023
	—	—	—	—	—	—	—	—	(0.146)	(0.104)	(0.241)	(0.199)

	Cohort 1, State Sample MAP, Prior Year Spring Baseline				Cohort 1, Full Sample, Prior Year Spring Baseline				Cohort 2, Full Sample, Prior Year Spring Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
Black	0.004	–0.072	–0.015	–0.077	0.041	–0.002	0.074	0.002	0.150	0.055	0.134	0.083
	(0.100)	(0.104)	(0.080)	(0.089)	(0.081)	(0.084)	(0.070)	(0.093)	(0.098)	(0.102)	(0.096)	(0.097)
Asian	0.068	0.038	0.068	0.073	0.078	0.018	0.089	0.043	–0.048	–0.033	–0.048	–0.074
	(0.071)	(0.076)	(0.084)	(0.085)	(0.053)	(0.067)	(0.059)	(0.069)	(0.056)	(0.056)	(0.049)	(0.055)
Hispanic	–0.030	0.003	0.056	0.029	0.029	–0.012	0.038	0.068	–0.002	0.000	0.047	0.031
	(0.123)	(0.121)	(0.104)	(0.112)	(0.094)	(0.094)	(0.084)	(0.100)	(0.099)	(0.099)	(0.098)	(0.095)
Other	–0.041	–0.041	–0.034	–0.027	–0.092*	–0.074+	–0.052	–0.036	–0.028	–0.026	–0.066	–0.087
	(0.046)	(0.049)	(0.047)	(0.061)	(0.041)	(0.044)	(0.040)	(0.057)	(0.053)	(0.054)	(0.052)	(0.054)
Female	–0.017	–0.018	–0.019	–0.019	–0.014	–0.019	–0.003	–0.022	0.044	0.036	0.079*	0.056
	(0.034)	(0.035)	(0.033)	(0.037)	(0.030)	(0.034)	(0.029)	(0.031)	(0.036)	(0.038)	(0.039)	(0.041)
ELL	–0.058	–0.039	0.040	–0.021	0.032	–0.009	0.055	0.040	–0.030	–0.036	0.045	–0.027
	(0.125)	(0.119)	(0.108)	(0.113)	(0.097)	(0.096)	(0.086)	(0.085)	(0.100)	(0.098)	(0.088)	(0.088)
SPED	–0.035	–0.056	–0.087	–0.103+	0.008	–0.007	–0.027	–0.055	–0.026	–0.015	–0.018	–0.003
	(0.076)	(0.076)	(0.058)	(0.061)	(0.065)	(0.070)	(0.053)	(0.055)	(0.057)	(0.050)	(0.058)	(0.049)
Economically Disadvantaged	0.085	0.002	0.160	0.022	0.112	0.023	0.192*	–0.004	0.258**	–0.008	0.175*	0.003
	(0.119)	(0.109)	(0.108)	(0.082)	(0.089)	(0.080)	(0.078)	(0.062)	(0.089)	(0.040)	(0.086)	(0.064)
Gifted	–0.024	0.032	–0.138	–0.036	–0.016	0.069	–0.110	–0.012	–0.124+	–0.004	–0.109	–0.006
	(0.112)	(0.106)	(0.095)	(0.075)	(0.097)	(0.091)	(0.076)	(0.062)	(0.066)	(0.047)	(0.075)	(0.059)

	Cohort 1, State Sample MAP, Prior Year Spring Baseline				Cohort 1, Full Sample, Prior Year Spring Baseline				Cohort 2, Full Sample, Prior Year Spring Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
Prior Year	–0.079	0.047	–0.063	0.120	–0.023	0.145	–0.048	0.114	–0.017	0.039	–0.084	0.016
ELA MAP	(0.132)	(0.128)	(0.119)	(0.098)	(0.106)	(0.102)	(0.098)	(0.078)	(0.075)	(0.069)	(0.084)	(0.076)
Prior Year	–0.035	0.071	–0.120	0.008	0.010	0.147	–0.086	0.060	–0.118	–0.028	–0.124	–0.028
Math MAP	(0.128)	(0.122)	(0.110)	(0.093)	(0.103)	(0.100)	(0.093)	(0.078)	(0.079)	(0.074)	(0.087)	(0.078)
Prior Year	–0.067	0.010	–0.047	0.052	0.010	0.091	–0.021	0.062	–0.171*	–0.046	–0.118	–0.028
Attendance	(0.062)	(0.067)	(0.050)	(0.067)	(0.054)	(0.060)	(0.051)	(0.055)	(0.075)	(0.080)	(0.077)	(0.083)
Teacher Characteristics												
Proportion	–0.033	–0.099	–0.114	–0.178	–0.002	0.003	–0.068	–0.123	–0.029	0.034	–0.038	0.004
SPED Taught	(0.151)	(0.141)	(0.154)	(0.139)	(0.125)	(0.129)	(0.122)	(0.129)	(0.141)	(0.123)	(0.123)	(0.111)
Proportion ELL	–0.168	–0.115	0.130	0.064	0.015	–0.065	0.113	0.155	0.001	0.014	0.130	0.061
Taught	(0.215)	(0.199)	(0.216)	(0.208)	(0.165)	(0.159)	(0.166)	(0.161)	(0.176)	(0.178)	(0.163)	(0.162)
School Covariates												
Prior Year	–0.141	–0.028	–0.201*	–0.033	–0.156*	–0.016	–0.220**	–0.021	–0.187**	–0.023	–0.187*	–0.049
ELA MAP	(0.086)	(0.079)	(0.088)	(0.066)	(0.065)	(0.060)	(0.063)	(0.053)	(0.065)	(0.055)	(0.074)	(0.067)
Prior Year	–0.111	0.003	–0.188*	–0.020	–0.123+	0.001	–0.201**	–0.009	–0.146*	–0.003	–0.151*	–0.026
Math MAP	(0.089)	(0.080)	(0.091)	(0.069)	(0.068)	(0.062)	(0.066)	(0.057)	(0.058)	(0.051)	(0.063)	(0.056)
Observations	2,739–3,002		2,742–3,013		3,958–4,371		4,045–4,587		3,026–3,764		3,367–4,490	

NOTES: Standard errors are clustered by teacher. Only weighted models include propensity score weights. All covariates are standardized within district. Economically disadvantaged in one district in Program C is an indicator for being eligible for free or reduced-price lunch. UW = unweighted, W = weighted. All variables are standardized.
+ indicates $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table B.12. Covariate Balance on Year 2 Effects with MAP Assessment, Cohort 1 Program A Only, by Baseline Measure

	Fall Baseline				Prior Spring Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
Student Covariates								
Kindergarten	–0.252	–0.211	–0.241	–0.203	–0.252	–0.200	–0.241	–0.210
	(0.161)	(0.163)	(0.168)	(0.188)	(0.161)	(0.170)	(0.168)	(0.186)
Grade 1	0.186	0.064	0.199	0.011	0.186	–0.181	0.199	0.092
	(0.205)	(0.238)	(0.209)	(0.258)	(0.205)	(0.261)	(0.209)	(0.234)
Grade 2	0.194	0.314	0.101	0.344	0.194	0.462*	0.101	0.243
	(0.237)	(0.230)	(0.227)	(0.209)	(0.237)	(0.224)	(0.227)	(0.231)
Grade 3	–0.116	–0.183	0.109	–0.010	–0.116	–0.144	0.109	–0.029
	(0.153)	(0.172)	(0.242)	(0.262)	(0.153)	(0.191)	(0.242)	(0.270)
Grade 4	–0.381**	–0.218*	–0.309*	–0.143	–0.381**	–0.138	–0.309*	–0.093
	(0.129)	(0.107)	(0.119)	(0.110)	(0.129)	(0.105)	(0.119)	(0.110)
Grade 5	0.318	0.140	0.272	0.050	0.318	0.070	0.272	0.027
	(0.256)	(0.277)	(0.256)	(0.292)	(0.256)	(0.290)	(0.256)	(0.302)
Grade 6	0.043	0.091	–0.133	–0.046	0.043	0.116	–0.133	–0.025
	(0.262)	(0.250)	(0.157)	(0.135)	(0.262)	(0.246)	(0.157)	(0.142)
Black	–0.014	–0.112	–0.005	–0.120	–0.014	–0.111	–0.005	–0.056
	(0.092)	(0.100)	(0.092)	(0.109)	(0.092)	(0.114)	(0.092)	(0.100)
Asian	0.120+	0.141*	0.102	0.122+	0.120+	0.136*	0.102	0.114+
	(0.064)	(0.063)	(0.062)	(0.062)	(0.064)	(0.065)	(0.062)	(0.063)
Hispanic	–0.055	0.037	–0.093	0.013	–0.055	–0.006	–0.093	–0.040
	(0.111)	(0.107)	(0.107)	(0.105)	(0.111)	(0.111)	(0.107)	(0.108)
Other	0.078	0.033	0.069	0.019	0.078	0.053	0.069	0.010
	(0.065)	(0.069)	(0.059)	(0.068)	(0.065)	(0.072)	(0.059)	(0.069)
Female	–0.008	–0.002	–0.016	–0.013	–0.008	0.003	–0.016	–0.005
	(0.049)	(0.054)	(0.043)	(0.049)	(0.049)	(0.066)	(0.043)	(0.055)
ELL	–0.005	0.049	–0.072	0.015	–0.005	0.069	–0.072	–0.020
	(0.122)	(0.111)	(0.106)	(0.098)	(0.122)	(0.115)	(0.106)	(0.100)

	Fall Baseline				Prior Spring Baseline			
	ELA (UW)	ELA (W)	Math (UW)	Math (W)	ELA (UW)	ELA (W)	Math (UW)	Math (W)
SPED	0.099	0.027	0.117+	0.039	0.099	0.025	0.117+	0.032
	(0.065)	(0.065)	(0.066)	(0.070)	(0.065)	(0.091)	(0.066)	(0.071)
Economically Disadvantaged	0.060	–0.067	0.099	–0.063	0.060	–0.068	0.099	–0.079
	(0.126)	(0.111)	(0.147)	(0.129)	(0.126)	(0.120)	(0.147)	(0.132)
Gifted	–0.069	0.016	–0.024	0.027	–0.069	0.044	–0.024	0.092
	(0.103)	(0.073)	(0.082)	(0.066)	(0.103)	(0.075)	(0.082)	(0.060)
ELA MAP	–0.014	0.062	0.010	0.033	–0.096	0.067	–0.082	0.114
	(0.099)	(0.090)	(0.093)	(0.093)	(0.127)	(0.112)	(0.114)	(0.101)
Math MAP	–0.014	0.093	–0.047	0.036	0.012	0.174	–0.024	0.169
	(0.108)	(0.097)	(0.098)	(0.095)	(0.121)	(0.109)	(0.121)	(0.116)
Attendance	–0.175*	–0.030	–0.170+	–0.057	–0.175*	0.034	–0.170+	–0.016
	(0.083)	(0.089)	(0.088)	(0.103)	(0.083)	(0.135)	(0.088)	(0.104)
Teacher Characteristics								
Proportion SPED	0.260	0.003	0.362*	0.062	0.260	–0.106	0.362*	0.052
Taught	(0.166)	(0.168)	(0.169)	(0.175)	(0.166)	(0.218)	(0.169)	(0.184)
Proportion ELL	–0.063	0.049	–0.067	0.086	–0.063	–0.007	–0.067	–0.012
Taught	(0.200)	(0.182)	(0.178)	(0.169)	(0.200)	(0.185)	(0.178)	(0.165)
School Covariates								
ELA MAP	–0.106	–0.012	–0.086	0.006	–0.163+	–0.023	–0.162+	–0.015
	(0.076)	(0.065)	(0.083)	(0.075)	(0.083)	(0.077)	(0.087)	(0.077)
Math MAP	–0.120	–0.010	–0.114	0.011	–0.123	0.001	–0.127	0.000
	(0.081)	(0.070)	(0.086)	(0.076)	(0.083)	(0.074)	(0.085)	(0.075)
Observations	3,001–3,653		3,053–3,715		2,203–3,002		2,152–3,013	

NOTES: Standard errors are clustered by teacher. Only weighted models include propensity score weights. All covariates are standardized within district. Economically disadvantaged in one district in Program C is an indicator for being eligible for free or reduced-price lunch. UW = unweighted, W = weighted. All variables are standardized.
+ indicates $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table B.13. Covariate Balance on Retention Effects, by Cohort and Program

	Program A, Cohort 1		Program C, Cohort 1		Program A, Cohort 2		Program C, Cohort 2	
	(UW)	(W)	(UW)	(W)	(UW)	(W)	(UW)	(W)
Classroom Covariates								
Grade	–0.190	–0.016	0.724*	0.672+	–0.444**	0.041	0.478*	0.477*
	(0.116)	(0.114)	(0.317)	(0.390)	(0.121)	(0.122)	(0.195)	(0.229)
Black	0.076	0.009	0.722+	0.801+	–0.020	0.032	0.581	0.591
	(0.131)	(0.137)	(0.420)	(0.469)	(0.139)	(0.150)	(0.404)	(0.398)
Asian	0.221	0.084	0.714	0.534	–0.090	–0.078	0.133	0.259
	(0.163)	(0.179)	(0.685)	(0.717)	(0.143)	(0.167)	(0.408)	(0.410)
Hispanic	0.007	–0.009	–0.077	–0.154	0.123	–0.001	–0.083	–0.193
	(0.132)	(0.134)	(0.202)	(0.249)	(0.137)	(0.152)	(0.227)	(0.212)
Other	–0.104	–0.091	–0.510**	–0.242	–0.020	–0.074	0.146	–0.079
	(0.115)	(0.112)	(0.192)	(0.220)	(0.140)	(0.157)	(0.341)	(0.333)
Female	0.226+	0.073	–0.891*	–0.068	0.281*	0.148	–0.834**	–0.351
	(0.117)	(0.114)	(0.346)	(0.464)	(0.125)	(0.129)	(0.236)	(0.284)
Economically Disadvantaged	0.229*	0.094	–0.287	–0.477	0.407**	0.163	0.358+	0.013
	(0.092)	(0.082)	(0.612)	(0.686)	(0.105)	(0.103)	(0.193)	(0.175)
Prior Year ELA	–0.044	–0.020	–0.992**	–0.252	0.092	0.037	–0.998**	–0.433
Achievement	(0.129)	(0.124)	(0.344)	(0.409)	(0.148)	(0.148)	(0.249)	(0.291)
Prior Year Math	–0.170	–0.121	–0.980**	–0.174	0.029	0.028	–0.938**	–0.408+
Achievement	(0.148)	(0.143)	(0.313)	(0.364)	(0.152)	(0.150)	(0.226)	(0.209)
Prior Year	0.041	0.053	–0.130	0.568	–0.065	0.027	0.154	0.086
Attendance	(0.126)	(0.128)	(0.546)	(0.633)	(0.120)	(0.130)	(0.156)	(0.132)
SPED	–0.265**	–0.116+	0.575	–0.217	–0.335**	–0.097	1.048**	0.481
	(0.078)	(0.062)	(0.435)	(0.499)	(0.095)	(0.082)	(0.290)	(0.346)
ELL	0.154	0.065	1.266*	0.763	0.187	0.023	0.523+	0.247
	(0.127)	(0.126)	(0.560)	(0.599)	(0.133)	(0.145)	(0.280)	(0.322)
School Covariates								
Prior Year ELA	–0.230*	–0.083	–0.183	0.074	–0.259*	–0.084	–0.125	0.028
Achievement	(0.113)	(0.109)	(0.156)	(0.217)	(0.111)	(0.110)	(0.308)	(0.299)

	Program A, Cohort 1		Program C, Cohort 1		Program A, Cohort 2		Program C, Cohort 2	
	(UW)	(W)	(UW)	(W)	(UW)	(W)	(UW)	(W)
Prior Year Math	–0.221+	–0.055	–0.058	0.185	–0.283*	–0.102	0.023	0.099
Achievement	(0.116)	(0.114)	(0.187)	(0.222)	(0.111)	(0.112)	(0.303)	(0.293)
Economically	0.202+	0.063	–0.287	–0.168	0.387**	0.162	0.438+	0.018
Disadvantaged	(0.108)	(0.100)	(0.302)	(0.374)	(0.101)	(0.098)	(0.237)	(0.247)
Black	0.014	–0.027	0.072	0.157	0.006	0.026	–0.033	–0.084
	(0.131)	(0.137)	(0.250)	(0.282)	(0.140)	(0.156)	(0.234)	(0.219)
Hispanic	0.027	–0.011	–0.092	–0.255	0.080	–0.026	0.465+	0.122
	(0.138)	(0.142)	(0.258)	(0.273)	(0.139)	(0.155)	(0.248)	(0.235)
Observations	362–375		96–135		218–320		143–174	

NOTES: All regressions include robust standard errors. Only weighted models include propensity score weights. All covariates are standardized within district. Economically disadvantaged in one district in Program C is an indicator for being eligible for free or reduced-price lunch. UW = unweighted, W = weighted. All variables are standardized. + indicates $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

References

Cooper, Harris, Larry V. Hedges, and Jeffrey C. Valentine, eds., *The Handbook of Research Synthesis and Meta-analysis*, 2nd ed., New York: Russell Sage Foundation, 2009.

Griffin, Beth Ann, Ricardo Sanchez, Matthew Cefalu, Lynsay Ayer, Donna L. Coffman, Brian Garrett Vegetabile, Joseph D. Pane, Lane F. Burgette, and Daniel F. McCaffrey. "Toolkit for Weighting and Analysis of Nonequivalent Groups," Santa Monica, Calif.: RAND Corporation, TL-A570-2, 2020. As of August 31, 2021: https://www.rand.org/pubs/tools/TLA570-2.html

Huguet, Alice, Christopher Joseph Doss, Benjamin K. Master, Fatih Unlu, Jessica L. Sousa, Karen Christianson, and Garrett Baker, *Widening the Pathway: Implementation and Impacts of Alternative Teacher Preparation Programs Across Three Contexts*, Santa Monica, Calif.: RAND Corporation, RR-A281-1, 2021. As of December 16, 2021: https://www.rand.org/pubs/research_reports/RRA281-1.html

Kaufman, Julia H., Benjamin K. Master, Alice Huguet, Paul Youngmin Yoo, Susannah Faxon-Mills, David Schulker, and Geoffrey E. Grimm, *Growing Teachers from Within: Implementation, Impact, and Cost of an Alternative Teacher Preparation Program in Three Urban School Districts*, Santa Monica, Calif.: RAND Corporation, RR-A256-1, 2020. As of August 17, 2021: https://www.rand.org/pubs/research_reports/RRA256-1.html

Ridgeway, G., D. Madigan, and T. Richardson, "Boosting Methodology for Regression Problems," paper presented at the Seventh International Workshop on Artificial Intelligence and Statistics, Fort Lauderdale, Fla., January 3–6, 1999.

Rosenbaum, Paul R., and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70, No. 1, 1983, pp. 41–55.

Rubin, Donald, *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley, 1987.

Schafer, J. L., *Analysis of Incomplete Multivariate Data*, Boca Raton, Fla.: Chapman & Hall/CRC, 1997.