Building and Sustaining Innovative High Schools

Technical Appendixes

Elizabeth D. Steiner, Laura S. Hamilton, John F. Pane, Jonathan Schweig, Laura Stelitano, Joseph D. Pane, Sophie Meyers

SPONSORED BY THE CARNEGIE CORPORATION OF NEW YORK
Preface

The Carnegie Corporation of New York’s (CCNY) Opportunity by Design (ObD) initiative provided support for new, small high schools of choice in several districts across the United States to adopt a set of design principles intended to ensure that students are prepared for college and careers. CCNY engaged the RAND Corporation in 2014 to conduct a comprehensive study of the ObD initiative. This technical appendix accompanies the final report summarizing the methods and findings from this five-year study.

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 report was made possible by a grant from Carnegie Corporation of New York. The statements made and views expressed are solely the responsibility of the authors.

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

Tables ................................................................................................................................. v
Abbreviations ................................................................................................................... vii
Appendix A. Outcomes Analysis, Methods, and Limitations ................................................. 1
  Analytic Methods .......................................................................................................... 1
  Sample ........................................................................................................................... 7
  Measures ....................................................................................................................... 22
  Analysis and Results ..................................................................................................... 28
  Limitations .................................................................................................................... 29
Appendix B. Implementation Analysis, Methods, and Limitations ........................................ 31
  Implementation Methods ............................................................................................. 31
  Analytic Methods ......................................................................................................... 37
  Limitations .................................................................................................................... 39
Appendix C. ObD Design Principles ................................................................................... 40
Appendix D. Student Survey Results .................................................................................. 42
Appendix E. Teacher Survey Results .................................................................................. 48
References ......................................................................................................................... 62
Tables

Table A.1. Demographic and Baseline Achievement Variables Included in the Propensity Score Model .......................................................... 2
Table A.2. ObD School Characteristics, as of Spring 2018 .......................................................... 8
Table A.3. Brooklyn’s Descriptive Statistics Before Propensity Score Weighting ...................... 9
Table A.4. Cleveland’s Descriptive Statistics Before Propensity Score Weighting .................. 10
Table A.5. Denver’s Descriptive Statistics Before Propensity Score Weighting ...................... 10
Table A.6. New York City’s Descriptive Statistics Before Propensity Score Weighting ............ 11
Table A.7. Philadelphia’s Descriptive Statistics Before Propensity Score Weighting ............... 12
Table A.8. Prince George’s County’s Descriptive Statistics Before Propensity Score Weighting .......................................................... 13
Table A.9. Providence’s Descriptive Statistics Before Propensity Score Weighting .................. 13
Table A.10. Brooklyn’s Descriptive Statistics After Propensity Score Weighting .................... 14
Table A.11. Cleveland’s Descriptive Statistics After Propensity Score Weighting .................... 15
Table A.12. Denver’s Descriptive Statistics After Propensity Score Weighting ...................... 15
Table A.13. New York City’s Descriptive Statistics After Propensity Score Weighting .............. 16
Table A.14. Philadelphia’s Descriptive Statistics After Propensity Score Weighting ............... 17
Table A.15. Prince George’s County’s Descriptive Statistics After Propensity Score Weighting .......................................................... 18
Table A.16. Providence’s Descriptive Statistics After Propensity Score Weighting .................. 18
Table A.17. Descriptive Statistics for All Districts Before Propensity Score Weighting for 2018–2019 12th-Graders .................................................. 20
Table A.18. Descriptive Statistics for All Districts After Propensity Score Weighting for 2018–2019 12th-Graders .................................................. 21
Table A.19. Numbers of Schools and Students in MAP Analyses ............................................ 22
Table A.20. Balance Between ObD and VCG Groups on Variables Not Exactly Matched ......... 22
Table A.21. Outcome Measures, by District ........................................................................... 23
Table A.22. Percentages of Students with Data for Outcome Measures, by District ............... 24
Table A.23. Distributions of Outcome Measures, by District ............................................... 25
Table A.24. Graduation Outcome Measure, by District: Percentages of Available Data ........... 27
Table A.25. Distributions of Graduation Outcome Measure, by District .................................. 27
Table A.27. Impact Estimates for Graduation Outcomes, 2018–2019 Academic Year ............ 29
Table A.28. Impact Estimates for MAP Outcomes, 2017–2018 Academic Year ..................... 29
Table A.29. Numbers of Interview Participants, Focus Groups, Observations, and Artifacts, 2015–2018 .................................................................. 32
Table B.2. Teacher Survey Response Rates, by School, 2015–2018 .................................................. 34
Table D.1. Frequency of Tasks Assigned to Students ................................................................. 42
Table D.2. Student Perceptions of Assigned Tasks ................................................................. 42
Table D.3. Student Attitudes Toward Doing Schoolwork ......................................................... 43
Table D.4. Submission Frequency of Students’ Completed Work ............................................. 43
Table D.5. Student Perceptions of Their Teachers in Typical Classroom Experiences .......... 44
Table D.6. Frequency of Students’ Typical Classroom Experiences ......................................... 45
Table D.7. Student Perceptions of Classmates ............................................................................. 45
Table D.8. Extent to Which Students’ Teachers Addressed Social and Emotional Learning Topics in the Classroom ........................................................................................................... 46
Table D.9. 2017–2018 Spring Student Survey Scales ................................................................. 46
Table D.10. Scales Included in Spring 2017–2018 Regression Analyses ...................................... 47
Table E.1. ObD Teachers’ Perceptions of Working Conditions at Their School ......................... 48
Table E.2. Conditions That Obstructed Teachers’ Efforts to Personalize Student Learning .......... 49
Table E.3a. Frequency with Which Teachers Received Student Performance Information........ 50
Table E.3b. Frequency with Which Teachers Received Student Performance Information .......... 51
Table E.4. Extent to Which Teachers Used Student Achievement or Mastery Data .................. 52
Table E.5. Teachers’ Assessment of Their Schools’ Data System .............................................. 54
Table E.6. Extent to Which Teachers Emphasized Personalized Learning Practices ................ 55
Table E.7. Extent to Which Teachers Emphasized Social and Emotional Learning Practices .... 56
Table E.8. Extent to Which Teachers Addressed Social and Emotional Learning Topics in Their Instruction ........................................................................................................................... 57
Table E.9. Teachers’ Assessment of Their Curriculum Materials ............................................ 57
Table E.10. Proportion of School- or District-Provided Curriculum and Instructional Materials That ObD Teachers Used ................................................................. 58
Table E.11. Proportion of Searched-for Resources That ObD Teachers Used to Supplement Their Curriculum and Instructional Materials ......................................................... 58
Table E.12. Proportion of Original Content That ObD Teachers Used to Supplement Their Curriculum and Instructional Materials ......................................................... 58
Table E.13. Time Teachers Spent in a Typical Week Developing or Selecting Instructional Materials** ................................................................................................................................. 59
Table E.14. Frequency with Which ObD Teachers Gave Students Tasks to Monitor or Assess Their Learning .............................................................................................................. 59
Table E.15. Frequency of Teacher Practices in Response to Poor Student Performance on a Task ................................................................................................................................. 60
Table E.16. Extent to Which Descriptions of Personalized Instruction Resembled Teachers’ Practices ............................................................................................................................... 61
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>Advanced Placement</td>
</tr>
<tr>
<td>ATP</td>
<td>American Teacher Panel</td>
</tr>
<tr>
<td>CCNY</td>
<td>Carnegie Corporation of New York</td>
</tr>
<tr>
<td>CEM</td>
<td>Coarsened Exact Matching</td>
</tr>
<tr>
<td>CMO</td>
<td>charter management organization</td>
</tr>
<tr>
<td>ELA</td>
<td>English language arts</td>
</tr>
<tr>
<td>ELL</td>
<td>English language learner</td>
</tr>
<tr>
<td>FRL</td>
<td>free and reduced-price lunch</td>
</tr>
<tr>
<td>LEP</td>
<td>limited English proficiency</td>
</tr>
<tr>
<td>ObD</td>
<td>Opportunity by Design</td>
</tr>
<tr>
<td>PD</td>
<td>professional development</td>
</tr>
<tr>
<td>RIT</td>
<td>Rasch Unit</td>
</tr>
<tr>
<td>VCG</td>
<td>virtual comparison group</td>
</tr>
<tr>
<td>WWC</td>
<td>What Works Clearinghouse</td>
</tr>
</tbody>
</table>
Appendix A. Outcomes Analysis, Methods, and Limitations

Analytic Methods

In this section, we describe the statistical methods we employed for estimating impacts of the Opportunity by Design (ObD) initiative on student outcomes. While all of our models can be described as matched comparison methods, the specifics of our analyses differ depending on the data source. We first describe the analytic methods that were used on outcomes obtained from district administrative data sources. We then describe the methods that were used on outcomes for data obtained from NWEA.

Three different outcomes were obtained from district administrative data: (1) student scores on state mathematics and English language arts (ELA) assessments; (2) behavioral outcomes, including student attendance and suspensions; and (3) college readiness outcomes, which include PSAT and SAT scores, credit accumulation, Advanced Placement (AP) passing rates,\(^1\) and high school graduation rates. We relied on the same two-step approach for all of these analyses. First, we used propensity score weighting to obtain a comparison group that is as similar as possible to the treatment group based on a set of baseline characteristics, and then we ran weighted statistical models, including baseline characteristics as covariates, to estimate program effects. Where an outcome was available in multiple grades (e.g., attendance), the analyses were done separately by grade and then combined using meta-analysis. This procedure was applied separately to data from each district to obtain district-specific impact estimates. Finally, we again applied meta-analysis to synthesize these district effects and estimate an average effect across all districts.

Two different outcomes were obtained from NWEA data: MAP scores on reading and mathematics assessments. These analyses use a different statistical framework—a virtual comparison group—to estimate impacts. This virtual comparison group method is described in more detail later in this chapter.

Matched Comparison Methods for District Administrative Data

To mitigate the threat of selection bias, we employed propensity score methods (e.g., Stuart, 2007) to create a comparison group that is as similar as possible to the treatment group using a set of observed baseline characteristics measured at the end of eighth grade. For ninth-graders, baseline data are taken from the 2017 academic year. For tenth-graders, baseline data are taken

\(^1\) AP tests are scored from 1 to 5, and colleges and universities may have different criteria about whether a specific AP score merits credit for college course work. In some colleges and courses, this occurs with a score of 3 or higher; others use a score of 4 or higher. In this analysis, we defined passing as a score of 3 or higher.
from the 2016 academic year. For 11th-graders, baseline data are taken from the 2015 academic year. For 12th-graders, baseline data are taken from the 2014 academic year. For these analyses, we used a wide range of demographic and prior achievement variables to create the comparison group (Table A.1). We used the TWANG (Toolkit for Weighting and Analysis of Nonequivalent Groups) data analysis package (Ridgeway et al., 2014) to create weights for individuals in the control group so that, on average, they closely matched the matched treatment groups. This R package allows for the estimation of propensity scores and for appraising covariate balance in treatment and comparison groups. All propensity score estimation was conducted within grade and within district (e.g., we estimated propensity scores separately for ninth-grade, tenth-grade, 11th-grade, and 12th-grade students in Brooklyn, N.Y.; Cleveland, Ohio; Denver, Colo.; New York, N.Y.; Philadelphia, Pa.; Prince George’s County, Md.; and Providence, R.I.).

Table A.1:
Demographic and Baseline Achievement Variables Included in the Propensity Score Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Cleveland</th>
<th>Denver</th>
<th>New York</th>
<th>Philadelphia</th>
<th>Prince George’s</th>
<th>Providence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Male</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>English language learner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ELL)/Limited English proficiency</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free or reduced-price lunch Eligible (FRL)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days attended (percentage)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Suspension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA (baseline)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Math (baseline)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Science (baseline)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

NOTE: One of the two ObD schools in New York was a charter school, and it was treated as a separate district for analysis purposes.

After weighting, grade-level estimates of the impact of the implementation of ObD design principles were obtained by comparing the outcomes of treated students and weighted comparison students. The treatment effect was estimated using a generalized least squares model, weighted by the estimated propensity score and using sandwich-estimated standard errors to account for the clustering of students within schools. We used so-called doubly-robust models—
which employ both propensity score weights and covariate adjustment—to obtain estimates of the treatment effect (Bang and Robins, 2005; Hullsiek and Louis, 2002):

\[ y_{igsd} = \beta_0 + \beta_1 T_{gsd} + X'_{igsd} \lambda + e_{igsd} \]  

(1)

where

- \( y_{igsd} \) is the outcome for student \( i \) in grade \( g \) in school \( s \) in district \( d \)
- \( T_{gsd} \) is a treatment indicator (1 = enrolled in an ObD school, 0 = otherwise)
- \( X_{igsd} \) is a vector of baseline student covariates including all of the variables included in the propensity score model (Table A.1)
- \( e_{igsd} \) is a residual term with mean zero and variance \( \sigma^2 \).

\( \hat{\beta}_1 \), the estimate of \( \beta_1 \), is interpreted as a grade-level–specific estimate of ObD impact.

**Meta-Analysis Techniques for District Administrative Data**

We adopted a meta-analytic approach to analyzing outcomes across districts. Broadly speaking, meta-analysis pools the results from individual studies to obtain a summary estimate of effects (Nordmann, Kasenda, and Briel, 2012). Many times, meta-analysis is used to synthesize results from previously conducted studies. However, as noted by Kalaian (2002), meta-analytic methods may also be used to synthesize treatment effects in multisite studies.

The current analysis applies meta-analysis models twice. First, a fixed-effect model was used to meta-analyze the data across grades within each of the study districts. We selected a fixed-effect model because we did not anticipate differences in effect size across grades; grades were conceived of as functionally equivalent (Borenstein, Hedges, and Rothstein, 2007) within districts, and all grades contributed information to a single, common effect estimate. We estimated a grade-level–specific effect in each district (\( \hat{\beta}_1 \) from Equation 1), along with a standard error for that estimate. Then, the following model was employed in each district:

\[ \hat{\beta}_{1g} = \theta + \epsilon_{ig} \]  

(2)

where \( \theta \) is the overall effect and \( \epsilon_{ig} \) is within-grade error, with mean 0 and variance \( \gamma^2 \). An estimate of the weighted mean impact within each district (across \( g \) grades) is given by

\[ \hat{\theta} = \frac{\sum_{i=1}^{g} \frac{1}{v_i} \hat{\beta}_{1i}}{\sum_{i=1}^{g} \frac{1}{v_i}} \]  

(3)

where \( v_i \) is given by the within-grade variance. This is a precision-weighted average across the \( g \) grades, accounting for within-grade sampling error. Grades with greater precision are given more weight in the estimation process (Borenstein, Hedges, and Rothstein, 2007). However, because precision is largely driven by sample size, and sample sizes were generally similar across grade
levels, all grades tended to receive similar weights and, thus, similar influence in these fixed-effect meta-analyses.

Second, we used a random-effect model to meta-analyze the data across the study districts. We selected a random-effect model because program effects were heterogeneous and implementation of the ObD principles differed from site to site. Using the estimated district-level effects from our fixed-effect meta-analysis (\( \hat{\theta} \) in Equation 3), we employed the following model:

\[
\hat{\theta}_d = \mu + \zeta_d + \varepsilon_{id}
\]  

(4)

where \( \mu \) is the mean of all district effects, \( \zeta_d \) is a between-district random effect with mean 0 and variance \( \tau \), and \( \varepsilon_{id} \) is within-district error, with mean 0 and variance \( \vartheta^2 \). An estimate of the weighted mean impact (across \( d \) districts) is given by

\[
\hat{\mu} = \frac{\sum_{i=1}^{d} \frac{1}{v_i} \hat{\theta}_i}{\sum_{i=1}^{d} \frac{1}{v_i}}
\]  

(5)

where \( v_i \) describes the sum of the between-district and the within-district variance. This is essentially a precision-weighted average across the \( d \) districts, accounting for both within-district and between-district sampling error. This is an important distinguishing feature from the fixed-effects model. Whereas in the fixed-effects model, weights are largely driven by sample size, that is not true in the random-effects model. Essentially, the more heterogeneity there is across districts (i.e., the larger \( \tau \) is relative to \( \vartheta^2 \)), the more similarly the districts are weighted in the overall effect estimation. Provided there is meaningful heterogeneity in effect estimates across districts, districts with larger sample sizes will not have outsized influence on the overall effect estimate (\( \hat{\mu} \)). In the case where all of the district-specific effects are relatively similar, \( \tau \) would contribute less to the district weights, and larger districts would be given larger relative weights in the overall effect estimation. We argue that this is a strength of the random-effects framework: If there are large differences in district-specific effects, those differences mean that large districts do not overly influence overall effect estimates. If the district-specific effects are relatively similar, the overall effect estimate “borrows” strength from the districts where the effects are most precisely estimated. All meta-analysis models were estimated in a Microsoft Excel worksheet using formulas provided by Borenstein, Hedges, and Rothstein (2007).

**Matching Method for Virtual Comparison Group (NWEA)**

For each treatment student, NWEA created a virtual comparison group (VCG, see Ma and Cronin, 2009) of up to 51 students from its database. Separate comparison groups were created for the mathematics and reading tests. Our analysis used fall scores as pretests and spring scores
as posttests. NWEA’s standard student and school matching criteria were applied to create the VCG.²

**Requirements for All VCG Matches**

- ObD and VCG students must have valid scores for the pretest and posttest.
- VCG students are not students in the ObD schools.
- VCG students are the same gender and in the same grade as the ObD group students to whom they are matched.
- VCG schools have the same locale classification (e.g., urban, suburban, rural) as the ObD schools, according to the National Center for Educational Statistics Public School Universe Survey.

**Approximate Matching Criteria**

- VCG schools differ from ObD schools by no more than 15 percentage points on the portion of students participating in the national FRL program.
- VCG students scored similarly to the ObD students to whom they were matched on the pretest MAP assessment. Preference is given for students with the exact same pretest score, but this can be expanded to within five points on NWEA’s Rasch Unit (RIT) scale³ if necessary to find matches.
- Time elapsed between the pretest and posttest differs by no more than 18 days between the VCG and ObD students.

**Statistical Estimation Strategy for NWEA MAP Analyses**

NWEA also provided unique identifiers for each VCG student so that we could observe cases where the same VCG student was selected to match more than one ObD student, and we could account for this duplication in our analyses. We use a type of Coarsened Exact Matching (CEM) estimator (Iacus and King, 2012). CEM allows us to analyze a data set with one record per student test event, instead of multiple records for VCG students matched to more than one treated student. It also more closely reflects and capitalizes on the matching algorithm enacted by NWEA.

The basic intuition of the CEM approach is that treated students are matched with control students based on observable similarities across several dimensions together, instead of collapsing the matching space into a univariate distance metric, such as is done with propensity

---

² NWEA first identified all student records that met these criteria, and if there were more than 51, then took a random sample of 51 of those records.

³ NWEA’s RIT scale is a stable equal-interval vertical scale designed to allow items of different difficulty levels to be placed on a common scale. A student’s RIT score indicates the level of question difficulty a given student is capable of answering correctly about 50 percent of the time.
score matching. This method is robust even if a control student is used as a match for multiple treatment students—only the closeness of the match is relevant. The process creates weights that reflect how often control students are repeated and the size of each treated student’s comparison group.

Specifically, treated students all receive a weight of 1, while control students are given a weight equal to the sum of the inverse of the size of their VCG group for each time they are in a treated student’s VCG. Equation 6 shows the definition of these weights.

\[
w_i = \begin{cases} 
1 & \text{if } T_i = 1 \\
\frac{1}{\sum_{j \in VCG_j} |VCG_j|} & \text{if } T_i = 0
\end{cases}
\]

where \(i\) indexes students, \(j\) indexes each VCG group student \(i\) appears in, and \(|VCG_j|\) is the number of VCG students in that group; \(T_i\) is a treatment indicator equal to 1 for ObD students and 0 for VCG students; and \(w_i\) is the weight for student \(i\). For example, consider a control student who appears in two treated students’ VCG groups. The first VCG group she appears in has 50 control students, and the second VCG group she is in has 48 control students. The weight for this control student would be \(\frac{1}{50} + \frac{1}{48} \approx 0.0408\).

After calculating these weights, we reduced the data set to having one observation per student test score. The weights were then applied in a weighted linear regression, as described below. The CEM estimator used here departs slightly from that of Iacus and King (2012), in that matching cells are created around each treated student instead of across all of the data points and, thus, may overlap across treated students; however, the general intuition of the approach is the same.

The dependent variable in the weighted regression is the gain from pretest to posttest in the MAP assessment scale score. We standardized test scores using mean and standard deviations of the pretest scores by grade, so that the pretest scores have a mean of zero and standard deviation of one within each grade level, and posttest scores reflect the standardized growth. We then divided the standardized growth by the number of days elapsed between pretest and posttest to account for variation in the time elapsed and to obtain a standardized measure of growth in achievement per day. We regressed the standardized growth in achievement per day on treatment status and on the following covariates: an indicator of whether the school is district-operated and the school-level percentage of students eligible for FRL, and student-level indicators of grade level and gender. We then scaled the treatment effect back up to a year by multiplying the coefficient on treatment by the average number of elapsed days for the sample (across both treatment and VCG). None of the exactly matched covariates is included in the regression, but they are implicitly controlled for.

We then used a clustering algorithm and degrees of freedom estimators that are robust when there are small numbers of clusters (Pustejovsky and Tipton, 2018). We clustered at the district level, and we used both the treatment and VCG clusters.
Multiple Hypotheses Testing

Inferences about program effectiveness are based on standard null hypothesis tests. However, as the number of these tests increases, the potential to capitalize on chance and find a false positive result (e.g., rejecting the null hypothesis when it is, in fact, true) also increases. Such errors in inference are often referred to as Type I errors. Because we report up to six different regression results and six associated \( p \)-values for each of our outcomes, we adopted a standard corrective measure to mitigate the possibility that our inferences were based on Type I errors. Specifically, following What Works Clearinghouse (WWC) standards, we used the Benjamini-Hochberg method for controlling the false discovery rate (Benjamini and Hochberg, 1995). Basically, this method adjusts the nominal 0.05 alpha level for determining statistical significance to reflect the total number of hypothesis tests that are conducted.

Sample

In this section, we describe the sample of students included in our analyses based on district administrative data. We provide descriptive information about both the unweighted and the weighted samples. Descriptive information about the unweighted sample in each district is useful for understanding the extent to which students enrolled in ObD schools differ from other students in the district with reference to observed characteristics and to gauge the threat of selection on observables. Descriptive information about the weighted sample in each district is useful for understanding the extent to which our propensity score methods were successful at mitigating the threat of selection on observable characteristics and improving the rigorousness of causal inferences that can be drawn from our analyses.

For the NWEA MAP assessments, we provide information on the number of students and schools that are included in the VCG. We then present information about the baseline equivalence of the ObD schools and the VCG.

ObD School Characteristics

Table A.2 presents ObD school characteristics as of spring 2018, which was the final year of the initiative. There were 16 ObD schools in seven districts and one charter management organization (CMO). The five Cohort I schools opened in 2014 and, in the spring of 2018, had been open four years and served students in grades 9 to 12. The five Cohort II schools opened in 2015 and, in the spring of 2018, had been in operation for three years; these schools served grades 9 to 11. The two Cohort III schools, which were both in the same district, had been open two years. One school served grades 9 and 10, and the other served grades 9 to 11. The four Cohort IV schools had been open only one year in spring 2019 and served students in grade 9. Only one district had ObD schools in more than one cohort.
Table A.2.
ObD School Characteristics, as of Spring 2018

<table>
<thead>
<tr>
<th>District/CMO</th>
<th>School</th>
<th>Cohort</th>
<th>Year Opened</th>
<th>Years in Operation</th>
<th>Grade Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>1</td>
<td>2014</td>
<td>4</td>
<td>9–12</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>1</td>
<td>2014</td>
<td>4</td>
<td>9–12</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>1</td>
<td>2014</td>
<td>4</td>
<td>9–12</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>1</td>
<td>2014</td>
<td>4</td>
<td>9–12</td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>1</td>
<td>2014</td>
<td>4</td>
<td>9–12</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>2</td>
<td>2015</td>
<td>3</td>
<td>9–11</td>
</tr>
<tr>
<td>5</td>
<td>G</td>
<td>2</td>
<td>2015</td>
<td>3</td>
<td>9–11</td>
</tr>
<tr>
<td>5</td>
<td>H</td>
<td>2</td>
<td>2015</td>
<td>3</td>
<td>9–11</td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td>2</td>
<td>2015</td>
<td>3</td>
<td>9–11</td>
</tr>
<tr>
<td>6</td>
<td>J</td>
<td>2</td>
<td>2015</td>
<td>3</td>
<td>9–11</td>
</tr>
<tr>
<td>1</td>
<td>K</td>
<td>3</td>
<td>2016</td>
<td>2</td>
<td>9–10</td>
</tr>
<tr>
<td>1</td>
<td>L</td>
<td>3</td>
<td>2016</td>
<td>2</td>
<td>9–11</td>
</tr>
<tr>
<td>1</td>
<td>M</td>
<td>4</td>
<td>2017</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>N</td>
<td>4</td>
<td>2017</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>O</td>
<td>4</td>
<td>2017</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>P</td>
<td>4</td>
<td>2017</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Sample Descriptives for District Administrative Data (2017–2018 Academic Year)

All administrative data were received from local education agencies and processed for analysis at the RAND Corporation. We obtained data on students enrolled in grades 9 to 12 at public schools in six school districts: Cleveland, Denver, New York City, Philadelphia,4 Prince George’s County, and Providence (although Brooklyn Lab is treated as a separate site in this study, the charter school’s data were provided by New York City). In total in the 2017–2018 academic year, there were 3,580 students enrolled in ObD schools across all six districts and 446,801 students from other schools.

The demographic characteristics of these students are presented in Tables A.3 to A.9. All of the variables presented were measured at baseline—that is, prior to enrollment in an ObD school. For students enrolled in the ninth grade in 2017–2018, this baseline reflects data from the 2016–2017 academic year. For students enrolled in tenth grade in 2017–2018, the baseline reflects data from the 2015–2016 academic year, and so on. There is some demographic variability both within districts (comparing ObD students with other students) and across districts. For example, approximately 65 percent of the students enrolled in ObD schools in Cleveland were black, compared with 13 percent in Denver. Nearly two-thirds of the students in Providence were Hispanic, far greater than in Cleveland’s student population. Within-distRICT differences can also be seen for achievement, and students enrolled in ObD schools typically had lower achievement

---

4 Philadelphia data used in this study were derived from data provided by the School District of Philadelphia. ©2015 School District of Philadelphia. All rights reserved.
than their peers. In Cleveland (Table A.4), this is true at every grade level. However, achievement is not directly comparable across districts, because each used different tests, and the scores are reported on different scales.

Table A.3.
Brooklyn’s Descriptive Statistics Before Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
</tr>
<tr>
<td>Male (%)</td>
<td>53.3</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0.8</td>
</tr>
<tr>
<td>Asian</td>
<td>12.3</td>
</tr>
<tr>
<td>Black</td>
<td>23.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>57.4</td>
</tr>
<tr>
<td>Multiple races</td>
<td>0.0</td>
</tr>
<tr>
<td>White</td>
<td>2.5</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>76.3</td>
</tr>
<tr>
<td>ELL (%)</td>
<td>24.7</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>87.5</td>
</tr>
<tr>
<td>Math</td>
<td>271.9</td>
</tr>
<tr>
<td>ELA</td>
<td>291.5</td>
</tr>
<tr>
<td>Total (N)</td>
<td>122</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. * denotes standardized mean difference greater than 0.25 standard deviations.
### Cleveland’s Descriptive Statistics Before Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>52.0</td>
<td>52.4</td>
<td>51.3</td>
<td>50.0</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1.9</td>
<td>1.1</td>
<td>1.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Black</td>
<td>55.9</td>
<td>64.9</td>
<td>67.4</td>
<td>65.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>22.8</td>
<td>16.8</td>
<td>22.3</td>
<td>17.0</td>
</tr>
<tr>
<td>Other</td>
<td>4.1</td>
<td>3.3</td>
<td>0.4</td>
<td>3.3</td>
</tr>
<tr>
<td>White</td>
<td>15.3</td>
<td>13.9</td>
<td>8.4</td>
<td>13.1</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>89.3</td>
<td>89.7</td>
<td>91.6</td>
<td>91.0</td>
</tr>
<tr>
<td>Math</td>
<td>665.1</td>
<td>666.3</td>
<td>654.3</td>
<td>673.1*</td>
</tr>
<tr>
<td>ELA</td>
<td>649.7</td>
<td>655.8</td>
<td>641.8</td>
<td>661.8*</td>
</tr>
<tr>
<td>Total (N)</td>
<td>675</td>
<td>2,650</td>
<td>273</td>
<td>2,616</td>
</tr>
</tbody>
</table>

**NOTES:** For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. For 12th-graders, baseline data are taken from the 2013–2014 academic year. * denotes standardized mean difference greater than 0.25 standard deviations.

### Denver’s Descriptive Statistics Before Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
</tr>
<tr>
<td>Male (%)</td>
<td>73.8</td>
<td>53.4*</td>
<td>71.6</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaskan Native</td>
<td>0.0</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Asian</td>
<td>6.0</td>
<td>2.7</td>
<td>7.5</td>
</tr>
<tr>
<td>Black</td>
<td>17.9</td>
<td>12.1</td>
<td>10.4</td>
</tr>
<tr>
<td>Hispanic</td>
<td>26.2</td>
<td>51.2*</td>
<td>16.4</td>
</tr>
<tr>
<td>Multiple races</td>
<td>7.1</td>
<td>4.4</td>
<td>6.0</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>0.0</td>
<td>0.3</td>
<td>1.5</td>
</tr>
<tr>
<td>White</td>
<td>41.7</td>
<td>17.5*</td>
<td>49.3</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>86.0</td>
<td>79.3*</td>
<td>85.6</td>
</tr>
<tr>
<td>Math</td>
<td>729.4</td>
<td>723.3</td>
<td>744.0</td>
</tr>
<tr>
<td>ELA</td>
<td>749.9</td>
<td>739.7*</td>
<td>755.6</td>
</tr>
<tr>
<td>Total (N)</td>
<td>84</td>
<td>8,276</td>
<td>67</td>
</tr>
</tbody>
</table>

**NOTES:** For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. * denotes standardized mean difference greater than 0.25 standard deviations.
### Table A.6.
**New York City’s Descriptive Statistics Before Propensity Score Weighting**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>68.5</td>
<td>47.7*</td>
<td>67.3</td>
<td>45.3*</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>1.4</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Asian</td>
<td>8.2</td>
<td>14.5</td>
<td>11.9</td>
<td>15.1</td>
</tr>
<tr>
<td>Black</td>
<td>26.0</td>
<td>27.2</td>
<td>28.7</td>
<td>28.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>47.9</td>
<td>40.9</td>
<td>42.6</td>
<td>41.0</td>
</tr>
<tr>
<td>Multiple races</td>
<td>0.7</td>
<td>0.6</td>
<td>3.0</td>
<td>0.6</td>
</tr>
<tr>
<td>White</td>
<td>10.3</td>
<td>12.4</td>
<td>9.9</td>
<td>12.2</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>66.9</td>
<td>70.3</td>
<td>63.6</td>
<td>71.8</td>
</tr>
<tr>
<td>ELL (%)</td>
<td>8.5</td>
<td>13.3</td>
<td>0.0</td>
<td>12.6</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>91.5</td>
<td>90.3</td>
<td>89.7</td>
<td>88.6</td>
</tr>
<tr>
<td>Math</td>
<td>286.9</td>
<td>293.8</td>
<td>302.3</td>
<td>298.5</td>
</tr>
<tr>
<td>ELA</td>
<td>306.7</td>
<td>309.9</td>
<td>311.3</td>
<td>308.2</td>
</tr>
<tr>
<td><strong>Total (N)</strong></td>
<td>146</td>
<td>97,509</td>
<td>101</td>
<td>96,739</td>
</tr>
</tbody>
</table>

**NOTES:** For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. For 12th-graders, baseline data are taken from the 2013–2014 academic year. * denotes standardized mean difference greater than 0.25 standard deviations.
Table A.7.
Philadelphia’s Descriptive Statistics Before Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th></th>
<th>Grade 10</th>
<th></th>
<th>Grade 11</th>
<th></th>
<th>Grade 12</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>50.2</td>
<td>51.9</td>
<td>46.0</td>
<td>51.5</td>
<td>54.4</td>
<td>50.0</td>
<td>50.0</td>
<td>48.4</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaskan Native</td>
<td>1.0</td>
<td>0.1</td>
<td>0.7</td>
<td>0.2</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Asian</td>
<td>1.0</td>
<td>8.3*</td>
<td>0.7</td>
<td>9.8*</td>
<td>0.0</td>
<td>11.2</td>
<td>1.5</td>
<td>11.4*</td>
</tr>
<tr>
<td>Black</td>
<td>46.3</td>
<td>52.7</td>
<td>55.8</td>
<td>54.0</td>
<td>59.6</td>
<td>53.4</td>
<td>66.4</td>
<td>54.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>47.8</td>
<td>21.1*</td>
<td>35.5</td>
<td>20.0*</td>
<td>36.8</td>
<td>18.0*</td>
<td>29.9</td>
<td>17.1*</td>
</tr>
<tr>
<td>Multiple races</td>
<td>2.0</td>
<td>4.5</td>
<td>2.2</td>
<td>3.4</td>
<td>2.9</td>
<td>4.9</td>
<td>0.7</td>
<td>3.3*</td>
</tr>
<tr>
<td>Native Hawaiian or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>White</td>
<td>2.0</td>
<td>13.3*</td>
<td>4.3</td>
<td>12.5*</td>
<td>0.7</td>
<td>12.4*</td>
<td>1.5</td>
<td>13.3*</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>83.2</td>
<td>67.2</td>
<td>69.8</td>
<td>64.1</td>
<td>66.7</td>
<td>60.3</td>
<td>78.2</td>
<td>68.5</td>
</tr>
<tr>
<td>LEP (%)</td>
<td>12.7</td>
<td>9.6</td>
<td>12.5</td>
<td>8.9</td>
<td>11.8</td>
<td>7.7</td>
<td>13.8</td>
<td>7.9</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>87.6</td>
<td>88.6</td>
<td>89.3</td>
<td>90.2</td>
<td>93.3</td>
<td>92.9</td>
<td>94.2</td>
<td>93.7</td>
</tr>
<tr>
<td>Suspension (%)</td>
<td>21.4</td>
<td>18.6</td>
<td>26.0</td>
<td>17.8</td>
<td>23.5</td>
<td>12.0*</td>
<td>24.1</td>
<td>11.9*</td>
</tr>
<tr>
<td>Math</td>
<td>820.0</td>
<td>891.3*</td>
<td>817.0</td>
<td>901.1*</td>
<td>852.4</td>
<td>924.9*</td>
<td>1,201.2</td>
<td>1,391.4</td>
</tr>
<tr>
<td>ELA</td>
<td>910.5</td>
<td>962.4*</td>
<td>902.2</td>
<td>983.5*</td>
<td>909.4</td>
<td>987.8*</td>
<td>1,228.0</td>
<td>1,422.6</td>
</tr>
<tr>
<td>Science</td>
<td>1,097.6</td>
<td>1,163.9*</td>
<td>1,019.4</td>
<td>1,181.4*</td>
<td>1,068.4</td>
<td>1,210.3*</td>
<td>1,067.3</td>
<td>1,218.4</td>
</tr>
<tr>
<td>Total (N)</td>
<td>205</td>
<td>11,723</td>
<td>138</td>
<td>10,526</td>
<td>136</td>
<td>8,690</td>
<td>134</td>
<td>8,142</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. For 12th-graders, baseline data are taken from the 2013–2014 academic year. LEP = limited English proficiency.

* denotes standardized mean difference greater than 0.25 standard deviations.
Table A.8.
Prince George’s County’s Descriptive Statistics Before Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th></th>
<th>Grade 10</th>
<th></th>
<th>Grade 11</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>54.5</td>
<td>52.8</td>
<td>55.3</td>
<td>51.9</td>
<td>58.4</td>
<td>50.1</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaskan Native</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Asian</td>
<td>1.1</td>
<td>2.7</td>
<td>1.1</td>
<td>2.5</td>
<td>1.5</td>
<td>3.3</td>
</tr>
<tr>
<td>Black</td>
<td>5.7</td>
<td>57.8*</td>
<td>7.6</td>
<td>52.5*</td>
<td>13.9</td>
<td>65.9*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>91.6</td>
<td>34.3*</td>
<td>89.5</td>
<td>27.9*</td>
<td>83.2</td>
<td>25.3*</td>
</tr>
<tr>
<td>Multiple races</td>
<td>0.0</td>
<td>1.9</td>
<td>0.0</td>
<td>1.5</td>
<td>0.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Native Hawaiian or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>White</td>
<td>1.6</td>
<td>3.0</td>
<td>1.7</td>
<td>2.8</td>
<td>1.5</td>
<td>3.4</td>
</tr>
<tr>
<td>Math</td>
<td>683.1</td>
<td>707.0*</td>
<td>682.2</td>
<td>717.0*</td>
<td>679.0</td>
<td>716.0*</td>
</tr>
<tr>
<td>ELA</td>
<td>679.4</td>
<td>729.6*</td>
<td>683.5</td>
<td>730.6*</td>
<td>691.7</td>
<td>735.0*</td>
</tr>
<tr>
<td>Total (N)</td>
<td>367</td>
<td>10,356</td>
<td>19</td>
<td>10,384</td>
<td>137</td>
<td>8,185</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. * denotes standardized mean difference greater than 0.25 standard deviations.

Table A.9.
Providence’s Descriptive Statistics Before Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th></th>
<th>Grade 10</th>
<th></th>
<th>Grade 11</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>61.7</td>
<td>55.5</td>
<td>49.4</td>
<td>56.6</td>
<td>41.1</td>
<td>54.4*</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaskan Native</td>
<td>2.0</td>
<td>0.7</td>
<td>2.3</td>
<td>0.9</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Asian</td>
<td>2.0</td>
<td>4.7</td>
<td>1.7</td>
<td>5.3*</td>
<td>2.7</td>
<td>4.5</td>
</tr>
<tr>
<td>Black</td>
<td>13.4</td>
<td>16.4</td>
<td>18.4</td>
<td>15.7</td>
<td>18.8</td>
<td>16.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>68.5</td>
<td>66.2</td>
<td>66.1</td>
<td>66.9</td>
<td>64.3</td>
<td>67.3</td>
</tr>
<tr>
<td>Multiple races</td>
<td>3.4</td>
<td>3.0</td>
<td>5.2</td>
<td>2.8</td>
<td>5.4</td>
<td>2.5</td>
</tr>
<tr>
<td>Native Hawaiian or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>White</td>
<td>10.7</td>
<td>8.8</td>
<td>6.3</td>
<td>8.3</td>
<td>8.0</td>
<td>7.7</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>58.4</td>
<td>61.3</td>
<td>44.8</td>
<td>51.7</td>
<td>35.7</td>
<td>45.8</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>82.9</td>
<td>86.1</td>
<td>89.5</td>
<td>89.8</td>
<td>91.1</td>
<td>92.2</td>
</tr>
<tr>
<td>Suspension (%)</td>
<td>5.6</td>
<td>4.1</td>
<td>16.7</td>
<td>15.1</td>
<td>10.0</td>
<td>11.6</td>
</tr>
<tr>
<td>Math</td>
<td>692.6</td>
<td>688.7</td>
<td>701.8</td>
<td>700.2</td>
<td>707.4</td>
<td>706.4</td>
</tr>
<tr>
<td>ELA</td>
<td>701.3</td>
<td>704.3</td>
<td>710.2</td>
<td>716.6</td>
<td>713.0</td>
<td>724.2*</td>
</tr>
<tr>
<td>Total (N)</td>
<td>149</td>
<td>2,147</td>
<td>174</td>
<td>2,072</td>
<td>112</td>
<td>1,778</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. * denotes standardized mean difference greater than 0.25 standard deviations.
Appraising Balance After Weighting (Administrative Data)

In Tables A.3 to A.9, it is evident that, before we adjust for propensity score, the ObD and comparison students differ significantly in terms of their propensity to attend an ObD school. Tables A.10 to A.16 show the improvement in baseline equivalence after propensity score weights are applied to the comparison pool. For each district and grade level, two columns are displayed: the mean for the ObD schools and the propensity score weighted mean for the comparison schools. Many of the variables with large differences in Tables A.3 to A.9 show dramatic improvements in balance: For example, in Prince George’s County, the proportion of Hispanic students is now nearly equivalent in the ObD and comparison groups. In all districts, for all variables, the standardized mean differences between treatment and control group are less than 0.25 standard deviations, which is often taken as the “cut-off” for determining equivalence in quasi-experimental studies (WWC, 2018). Overall, Tables A.10 to A.16 show that, after weighting, the treatment and control groups are equivalent on all baseline observed covariates.

Table A.10.
Brooklyn’s Descriptive Statistics After Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
</tr>
<tr>
<td>Male (%)</td>
<td>53.3</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0.8</td>
</tr>
<tr>
<td>Asian</td>
<td>12.3</td>
</tr>
<tr>
<td>Black</td>
<td>23.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>57.4</td>
</tr>
<tr>
<td>Multiple races</td>
<td>0.0</td>
</tr>
<tr>
<td>White</td>
<td>2.5</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>76.3</td>
</tr>
<tr>
<td>ELL (%)</td>
<td>24.7</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>87.5</td>
</tr>
<tr>
<td>Math</td>
<td>271.9</td>
</tr>
<tr>
<td>ELA</td>
<td>291.5</td>
</tr>
<tr>
<td>Total (N)</td>
<td>122</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. None of the standardized mean differences is greater than 0.25 standard deviations.
Table A.11.
Cleveland’s Descriptive Statistics After Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>52.0</td>
<td>51.9</td>
<td>51.3</td>
<td>51.6</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1.9</td>
<td>1.6</td>
<td>1.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Black</td>
<td>55.9</td>
<td>57.7</td>
<td>67.4</td>
<td>67.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>22.8</td>
<td>21.6</td>
<td>22.3</td>
<td>20.0</td>
</tr>
<tr>
<td>Other</td>
<td>4.1</td>
<td>3.9</td>
<td>0.4</td>
<td>1.3</td>
</tr>
<tr>
<td>White</td>
<td>15.3</td>
<td>15.1</td>
<td>8.4</td>
<td>9.8</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>89.3</td>
<td>89.5</td>
<td>91.6</td>
<td>91.6</td>
</tr>
<tr>
<td>Math</td>
<td>665.1</td>
<td>664.5</td>
<td>654.3</td>
<td>656.0</td>
</tr>
<tr>
<td>ELA</td>
<td>649.7</td>
<td>650.1</td>
<td>641.8</td>
<td>643.8</td>
</tr>
<tr>
<td>Total (N)</td>
<td>675</td>
<td>2,650</td>
<td>273</td>
<td>2,616</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. For 12th-graders, baseline data are taken from the 2013–2014 academic year. None of the standardized mean differences is greater than 0.25 standard deviations.

Table A.12.
Denver’s Descriptive Statistics After Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
</tr>
<tr>
<td>Male (%)</td>
<td>73.8</td>
<td>70.5</td>
<td>71.6</td>
</tr>
<tr>
<td>Race/Ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaskan Native</td>
<td>0.0</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Asian</td>
<td>6.0</td>
<td>6.0</td>
<td>7.5</td>
</tr>
<tr>
<td>Black</td>
<td>17.9</td>
<td>17.5</td>
<td>10.4</td>
</tr>
<tr>
<td>Hispanic</td>
<td>26.2</td>
<td>26.3</td>
<td>16.4</td>
</tr>
<tr>
<td>Multiple races</td>
<td>7.1</td>
<td>6.3</td>
<td>6.0</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>0.0</td>
<td>0.1</td>
<td>1.5</td>
</tr>
<tr>
<td>White</td>
<td>41.7</td>
<td>41.7</td>
<td>49.3</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>86.0</td>
<td>85.9</td>
<td>85.6</td>
</tr>
<tr>
<td>Math</td>
<td>729.4</td>
<td>728.0</td>
<td>744.0</td>
</tr>
<tr>
<td>ELA</td>
<td>749.9</td>
<td>748.5</td>
<td>755.6</td>
</tr>
<tr>
<td>Total (N)</td>
<td>84</td>
<td>8,276</td>
<td>67</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. None of the standardized mean differences is greater than 0.25 standard deviations.
Table A.13.  
New York City’s Descriptive Statistics After Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>68.5</td>
<td>66.8</td>
<td>67.3</td>
<td>66.4</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaskan Native</td>
<td>1.4</td>
<td>1.6</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Asian</td>
<td>8.2</td>
<td>8.5</td>
<td>11.9</td>
<td>11.7</td>
</tr>
<tr>
<td>Black</td>
<td>26.0</td>
<td>25.7</td>
<td>28.7</td>
<td>29.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>47.9</td>
<td>47.9</td>
<td>42.6</td>
<td>41.9</td>
</tr>
<tr>
<td>Multiple races</td>
<td>0.7</td>
<td>0.8</td>
<td>3.0</td>
<td>1.9</td>
</tr>
<tr>
<td>White</td>
<td>10.3</td>
<td>9.9</td>
<td>9.9</td>
<td>10.8</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>66.9</td>
<td>67.2</td>
<td>63.6</td>
<td>65.6</td>
</tr>
<tr>
<td>ELL (%)</td>
<td>8.5</td>
<td>8.8</td>
<td>0.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>91.5</td>
<td>92.1</td>
<td>89.7</td>
<td>90.2</td>
</tr>
<tr>
<td>Math</td>
<td>286.9</td>
<td>288.6</td>
<td>302.3</td>
<td>303.8</td>
</tr>
<tr>
<td>ELA</td>
<td>306.7</td>
<td>307.3</td>
<td>311.3</td>
<td>311.2</td>
</tr>
<tr>
<td>Total (N)</td>
<td>146</td>
<td>97,509</td>
<td>101</td>
<td>96,739</td>
</tr>
</tbody>
</table>

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. For 12th-graders, baseline data are taken from the 2013–2014 academic year. None of the standardized mean differences is greater than 0.25 standard deviations.
Table A.14.
Philadelphia’s Descriptive Statistics After Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (%)</td>
<td>50.2</td>
<td>52.3</td>
<td>46.0</td>
<td>45.8</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>1.0</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Asian</td>
<td>1.0</td>
<td>2.6</td>
<td>0.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Black</td>
<td>46.3</td>
<td>46.2</td>
<td>55.8</td>
<td>56.4</td>
</tr>
<tr>
<td>Hispanic</td>
<td>47.8</td>
<td>43.8</td>
<td>35.5</td>
<td>34.6</td>
</tr>
<tr>
<td>Multiple races</td>
<td>2.0</td>
<td>2.2</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>White</td>
<td>2.0</td>
<td>4.7</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>83.2</td>
<td>76.7</td>
<td>69.8</td>
<td>70.6</td>
</tr>
<tr>
<td>LEP (%)</td>
<td>12.7</td>
<td>15.6</td>
<td>12.5</td>
<td>13.7</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>87.6</td>
<td>88.0</td>
<td>89.3</td>
<td>89.5</td>
</tr>
<tr>
<td>Suspension (%)</td>
<td>21.4</td>
<td>22.1</td>
<td>26.0</td>
<td>25.8</td>
</tr>
<tr>
<td>Math</td>
<td>820.0</td>
<td>833.6</td>
<td>817.0</td>
<td>818.7</td>
</tr>
<tr>
<td>ELA</td>
<td>910.5</td>
<td>913.6</td>
<td>902.2</td>
<td>902.2</td>
</tr>
<tr>
<td>Science</td>
<td>1,097.6</td>
<td>1,100.4</td>
<td>1,019.4</td>
<td>1,021.7</td>
</tr>
</tbody>
</table>

Total (N) 205 11,723 138 10,526 136 8,690 134 8,142

NOTES: For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. For 12th-graders, baseline data are taken from the 2013–2014 academic year. None of the standardized mean differences is greater than 0.25 standard deviations.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th></th>
<th>Grade 10</th>
<th></th>
<th>Grade 11</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
</tr>
<tr>
<td>Male (%)</td>
<td>54.5</td>
<td>55.4</td>
<td>55.3</td>
<td>55.0</td>
<td>58.4</td>
<td>58.0</td>
</tr>
<tr>
<td>Race/Ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Asian</td>
<td>1.1</td>
<td>0.9</td>
<td>1.1</td>
<td>1.1</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Black</td>
<td>5.7</td>
<td>5.8</td>
<td>7.6</td>
<td>8.1</td>
<td>13.9</td>
<td>14.9</td>
</tr>
<tr>
<td>Hispanic</td>
<td>91.6</td>
<td>91.5</td>
<td>89.5</td>
<td>89.0</td>
<td>83.2</td>
<td>82.4</td>
</tr>
<tr>
<td>Multiple races</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>White</td>
<td>1.6</td>
<td>1.5</td>
<td>1.7</td>
<td>1.5</td>
<td>1.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Math</td>
<td>683.1</td>
<td>682.6</td>
<td>682.2</td>
<td>682.3</td>
<td>679.0</td>
<td>677.7</td>
</tr>
<tr>
<td>ELA</td>
<td>679.4</td>
<td>680.7</td>
<td>683.5</td>
<td>685.5</td>
<td>691.7</td>
<td>693.4</td>
</tr>
<tr>
<td>Total (N)</td>
<td>367</td>
<td>10,356</td>
<td>19</td>
<td>10,384</td>
<td>137</td>
<td>8,185</td>
</tr>
</tbody>
</table>

**NOTES:** For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. None of the standardized mean differences is greater than 0.25 standard deviations.
### Table A.16.
Providence’s Descriptive Statistics After Propensity Score Weighting

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
</tr>
<tr>
<td>Male (%)</td>
<td>61.7</td>
<td>60.2</td>
<td>49.4</td>
</tr>
<tr>
<td>Race/Ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>2.0</td>
<td>1.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Asian</td>
<td>2.0</td>
<td>2.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Black</td>
<td>13.4</td>
<td>14.0</td>
<td>18.4</td>
</tr>
<tr>
<td>Hispanic</td>
<td>68.5</td>
<td>67.1</td>
<td>66.1</td>
</tr>
<tr>
<td>Multiple races</td>
<td>3.4</td>
<td>3.4</td>
<td>5.2</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>White</td>
<td>10.7</td>
<td>10.9</td>
<td>6.3</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>58.4</td>
<td>56.0</td>
<td>44.8</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>82.9</td>
<td>83.1</td>
<td>89.5</td>
</tr>
<tr>
<td>Suspension (%)</td>
<td>5.6</td>
<td>5.0</td>
<td>16.7</td>
</tr>
<tr>
<td>Math</td>
<td>692.6</td>
<td>690.4</td>
<td>701.8</td>
</tr>
<tr>
<td>ELA</td>
<td>701.3</td>
<td>700.1</td>
<td>710.2</td>
</tr>
<tr>
<td><strong>Total (N)</strong></td>
<td>149</td>
<td>2,147</td>
<td>174</td>
</tr>
</tbody>
</table>

**NOTES:** For ninth-graders, baseline data are taken from the 2016–2017 academic year. For tenth-graders, baseline data are taken from the 2015–2016 academic year. For 11th-graders, baseline data are taken from the 2014–2015 academic year. For 12th-graders, baseline data are taken from the 2013–2014 academic year. None of the standardized mean differences is greater than 0.25 standard deviations.

**Sample Descriptives for District Administrative Data (2018–2019 Graduation Data)**

We conducted analyses on graduation using data from students enrolled in 12th grade in the 2018–2019 academic year. The demographic characteristics of these students are presented in Table A.17. As above, all of the variables presented were measured prior to enrollment in an ObD school.
### Table A.17.
Descriptive Statistics for All Districts Before Propensity Score Weighting for 2018–2019 12th-Graders

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cleveland ObD</th>
<th>Denver ObD</th>
<th>New York City ObD</th>
<th>Philadelphia ObD</th>
<th>Providence ObD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (%)</td>
<td>42.1</td>
<td>52.4</td>
<td>71.8</td>
<td>50.5</td>
<td>44.7</td>
</tr>
<tr>
<td>Other</td>
<td>47.8</td>
<td>52.7</td>
<td>52.2*</td>
<td>48.7</td>
<td>52.4*</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Asian</td>
<td>1.7</td>
<td>3.6</td>
<td>1.4</td>
<td>7.0*</td>
<td>0.0</td>
</tr>
<tr>
<td>Black</td>
<td>90.1</td>
<td>7.0</td>
<td>16.0*</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.7</td>
<td>20.2</td>
<td>57.8*</td>
<td>45.1</td>
<td>41.2</td>
</tr>
<tr>
<td>Multiple races</td>
<td>4.2</td>
<td>0.0</td>
<td>0.8</td>
<td>3.1</td>
<td>5.1</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0.0</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>White</td>
<td>5.0</td>
<td>51.2</td>
<td>17.8*</td>
<td>9.9</td>
<td>1.0</td>
</tr>
<tr>
<td>FRL (%)</td>
<td>76.9</td>
<td>72.3</td>
<td>61.2</td>
<td>59.0</td>
<td>40.4</td>
</tr>
<tr>
<td>LEP (%)</td>
<td>4.6</td>
<td>12.9</td>
<td>17.9</td>
<td>7.4*</td>
<td>92.5</td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>91.1</td>
<td>85.1</td>
<td>96.4</td>
<td>93.8</td>
<td>94.1</td>
</tr>
<tr>
<td>Suspension (%)</td>
<td>93.2*</td>
<td>82.8</td>
<td>94.8</td>
<td>94.1</td>
<td>10.6</td>
</tr>
<tr>
<td>Math</td>
<td>702.1</td>
<td>738.0</td>
<td>306.5</td>
<td>300.7</td>
<td>855.0</td>
</tr>
<tr>
<td>ELA</td>
<td>714.4</td>
<td>746.6</td>
<td>299.6</td>
<td>303.2</td>
<td>908.9</td>
</tr>
<tr>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,070.8</td>
</tr>
<tr>
<td>Total (N)</td>
<td>121</td>
<td>2,100</td>
<td>84</td>
<td>8,088</td>
<td>7,581</td>
</tr>
</tbody>
</table>

**NOTES:** For 12th-graders, baseline data are taken from the 2014–2015 academic year. Brooklyn did not have enrolled 12th-graders in 2018–2019. Prince George's County did not provide graduation data.

* denotes standardized mean difference greater than 0.25 standard deviations.

### Appraising Balance After Weighting (2018–2019 Graduation Data)

Table A.18 shows the improvement in baseline equivalence after applying propensity score weights to the comparison pool. As above, for each district, two columns are displayed: the mean for the ObD schools and the propensity score weighted mean for the comparison schools.

Overall, Table A.18 shows that, after weighting, the treatment and control groups are equivalent on all baseline observed covariates.
### Descriptive Statistics for All Districts After Propensity Score Weighting for 2018–2019 12th-Graders

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cleveland</th>
<th>Denver</th>
<th>New York City</th>
<th>Philadelphia</th>
<th>Providence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
<td>Other</td>
<td>ObD</td>
</tr>
<tr>
<td>Male (%)</td>
<td>42.1</td>
<td>45.7</td>
<td>52.4</td>
<td>52.2</td>
<td>71.8</td>
</tr>
<tr>
<td>Race/ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>0.0</td>
<td>0.4</td>
<td>1.4</td>
<td>0.9</td>
<td>0.0</td>
</tr>
<tr>
<td>Asian</td>
<td>1.7</td>
<td>1.7</td>
<td>3.6</td>
<td>3.6</td>
<td>7.0</td>
</tr>
<tr>
<td>Black</td>
<td>90.1</td>
<td>87.3</td>
<td>17.9</td>
<td>17.3</td>
<td>33.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.7</td>
<td>3.7</td>
<td>20.2</td>
<td>22.8</td>
<td>45.1</td>
</tr>
<tr>
<td>Multiple races</td>
<td>7.1</td>
<td>6.7</td>
<td>0.0</td>
<td>0.1</td>
<td>3.1</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Unknown</td>
<td>1.7</td>
<td>1.6</td>
<td>51.2</td>
<td>49.0</td>
<td>9.9</td>
</tr>
<tr>
<td>White</td>
<td>5.0</td>
<td>5.8</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>FRL (%)</td>
<td>76.9</td>
<td>75.1</td>
<td>61.2</td>
<td>65.6</td>
<td>40.4</td>
</tr>
<tr>
<td>LEP (%)</td>
<td>4.6</td>
<td>5.2</td>
<td>17.9</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>Attendance (%)</td>
<td>91.1</td>
<td>91.3</td>
<td>85.1</td>
<td>85.0</td>
<td>96.4</td>
</tr>
<tr>
<td>Suspension (%)</td>
<td>20.9</td>
<td>17.2</td>
<td>10.6</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>702.1</td>
<td>702.4</td>
<td>738.0</td>
<td>737.8</td>
<td>306.5</td>
</tr>
<tr>
<td>ELA</td>
<td>714.4</td>
<td>715.1</td>
<td>746.6</td>
<td>746.3</td>
<td>299.6</td>
</tr>
<tr>
<td>Science</td>
<td>1,070.8</td>
<td>1,082.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (N)</td>
<td>121</td>
<td>2,100</td>
<td>84</td>
<td>8,088</td>
<td>71</td>
</tr>
</tbody>
</table>

**NOTES:** For 12th-graders, baseline data are taken from the 2014–2015 academic year. Brooklyn did not have enrolled 12th-graders in 2018–2019. Prince George's County did not provide graduation data. None of the standardized mean differences is greater than 0.25 standard deviations.

### Numbers of Schools and Students in NWEA MAP Analysis

Table A.19 displays the number of schools and students included in the overall analyses of mathematics and reading. Students had to remain in one of the ObD schools in our sample for the 2017–2018 school year to be included in the analyses. Table A.19 indicates the students’ grade level at the start of the relevant time span.
Table A.19.
Numbers of Schools and Students in MAP Analyses

<table>
<thead>
<tr>
<th>Subject</th>
<th>Group</th>
<th>Number of Schools</th>
<th>Grade 9</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>ObD</td>
<td>14</td>
<td>843</td>
<td>537</td>
<td>351</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>VCG</td>
<td>409</td>
<td>13,784</td>
<td>10,836</td>
<td>8,450</td>
<td>2,947</td>
</tr>
<tr>
<td>Math</td>
<td>ObD</td>
<td>14</td>
<td>869</td>
<td>526</td>
<td>328</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>VCG</td>
<td>430</td>
<td>14,138</td>
<td>11,189</td>
<td>8,401</td>
<td>2,507</td>
</tr>
</tbody>
</table>

Assessment of Balance Between the Treatment Group and the VCG (MAP Analysis)

The VCG is intended to be very similar to the treatment group in terms of students’ observable characteristics prior to treatment. This is true by construction for the criteria that were matched exactly (specifically, the grade level of the student and the urbanicity of the school). For the approximate matching criteria, we examined whether the groups appear to be the same. Table A.20 shows balance on variables that were approximately matched. We present both the unweighted VCG means (after restricting the sample to retain only one observation per VCG student per subject, per year) and the weighted means using the CEM weights described above. We also present the standardized difference, calculated by dividing the difference by the standard deviation of the variable for the pooled sample (treatment and VCG). Both before and after weighting, school-level eligibility for FRL was about 4 percentage points higher in ObD schools than in the schools of the VCG students.

Table A.20.
Balance Between ObD and VCG Groups on Variables Not Exactly Matched

<table>
<thead>
<tr>
<th>Subject</th>
<th>Variable</th>
<th>ObD Mean</th>
<th>Unweighted VCG</th>
<th>Weighted VCG</th>
<th>Standardized Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Difference</td>
<td>Standardized Difference</td>
</tr>
<tr>
<td>Reading</td>
<td>Start RIT</td>
<td>204.34</td>
<td>206.69</td>
<td>–2.35</td>
<td>–0.12</td>
</tr>
<tr>
<td></td>
<td>FRL</td>
<td>93.90</td>
<td>90.30</td>
<td>3.60</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Elapsed</td>
<td>168.84</td>
<td>163.17</td>
<td>5.67</td>
<td>0.10</td>
</tr>
<tr>
<td>Math</td>
<td>Start RIT</td>
<td>212.34</td>
<td>214.97</td>
<td>–2.63</td>
<td>–0.14</td>
</tr>
<tr>
<td></td>
<td>FRL</td>
<td>94.40</td>
<td>90.19</td>
<td>4.22</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Elapsed</td>
<td>171.14</td>
<td>172.70</td>
<td>–1.55</td>
<td>–0.03</td>
</tr>
</tbody>
</table>

NOTE: The unweighted VCG columns show sample characteristics after restricting to one observation per VCG student per subject per year.

Measures

In this section, we describe the outcome measures for our analyses of district administrative data. We provide information about the specific measures used, the proportions of students with available data in each district, and the means and standard deviations of each outcome measure.
Outcome Measures for District Administrative Data (2017–2018 Academic Year)

We examined the ObD impacts on three broad categories of student outcomes: academic outcomes, behavioral outcomes, and college readiness outcomes. The specific measures investigated depended on the data available in each district. All outcome analyses were conducted using standardized variables, and all standardization was conducted within district and within grade level using sample means and standard deviations. For academic outcomes, there was evidence that assessment policy differed across districts—for example, students in some districts all took the same exam in the same grade level. In other districts, students took different exams in different grade years. Decisions about which assessments to include were informed by data availability and comparability across districts. Where possible, we included the assessments with the greatest coverage in each district; in mathematics, we took assessments that focused on algebra. Table A.21 summarizes the available outcomes by district.

Table A.21.
Outcome Measures, by District

<table>
<thead>
<tr>
<th></th>
<th>Cleveland</th>
<th>Denver</th>
<th>New York City</th>
<th>Philadelphia</th>
<th>Prince George’s County</th>
<th>Providence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Academic Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>EOC Algebra (9)</td>
<td>PSAT 8/9 (9)</td>
<td>Regents CC</td>
<td>Keystone Algebra (9)</td>
<td>Algebra 1 (9)</td>
<td>PSAT 8/9 (9)</td>
</tr>
<tr>
<td>ELA</td>
<td>EOC ELA (9)</td>
<td>PSAT 8/9 (9)</td>
<td>Regents CC</td>
<td>Keystone Literature (10)</td>
<td>ELA (11)</td>
<td>PSAT 8/9 (9)</td>
</tr>
<tr>
<td><strong>Behavioral Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
<td>(9,10,11)</td>
</tr>
<tr>
<td>Suspension</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
<td>(9,10,11)</td>
</tr>
<tr>
<td><strong>College Readiness Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSAT</td>
<td>✓ (10)</td>
<td>✓ (10)</td>
<td>✓ (10)</td>
<td>N/A</td>
<td>✓ (10)</td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td>✓ (12)</td>
<td>✓ (11)</td>
<td>✓ (11)</td>
<td>N/A</td>
<td>✓ (12)</td>
<td></td>
</tr>
<tr>
<td>Credits</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
<td>✓ (10,11)</td>
<td></td>
</tr>
<tr>
<td>AP tests passed</td>
<td>N/A</td>
<td>✓ (10,11)</td>
<td>✓ (12)</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Graduation</td>
<td>✓ (12)</td>
<td>N/A</td>
<td>✓ (12)</td>
<td>N/A</td>
<td>✓ (12)</td>
<td></td>
</tr>
</tbody>
</table>

NOTES: Parentheses indicate the specific grade levels for which data were available. Attendance is defined as the proportion of enrolled days attended for each student. Suspension is defined as an indicator of whether a student had been suspended at any time during the academic year. PSAT is the overall PSAT score. SAT is the overall SAT score. Credits is defined as cumulative credits earned. AP tests passed is defined as the number of AP tests with a score of 3 or higher. All data are taken from 2017–2018 administrative data provided by districts. N/A indicates that outcomes were not available for a district. CC = Common Core; EOC = end-of-course. * indicates outcomes used in analysis of the Brooklyn charter school, which was treated as a separate district.

Table A.22 provides the percentage of students with available data for each outcome in the 2017–2018 school year. For some districts and outcomes, the percentage of students with
available outcomes is very low: For example, there are few students in Providence who took math and ELA assessments in ninth grade.

Table A.22.
Percentages of Students with Data for Outcome Measures, by District

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Cleveland</th>
<th>Others</th>
<th>Denver</th>
<th>Others</th>
<th>New York City</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD (% (N))</td>
<td></td>
<td>ObD (% (N))</td>
<td></td>
<td>ObD (% (N))</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>49 (331)</td>
<td>49 (331)</td>
<td>56 (47)</td>
<td>52 (4,325)</td>
<td>72 (105)</td>
<td>47 (46,138)</td>
</tr>
<tr>
<td>ELA</td>
<td>53 (360)</td>
<td>53 (360)</td>
<td>56 (47)</td>
<td>52 (4,325)</td>
<td>30 (44)</td>
<td>6 (5,414)</td>
</tr>
<tr>
<td>Attendance</td>
<td>61 (765)</td>
<td>61 (765)</td>
<td>64 (156)</td>
<td>57 (12,038)</td>
<td>97 (387)</td>
<td>87 (296,933)</td>
</tr>
<tr>
<td>Suspension</td>
<td>49 (615)</td>
<td>49 (615)</td>
<td>64 (156)</td>
<td>57 (12,038)</td>
<td>95 (379)</td>
<td>86 (293,841)</td>
</tr>
<tr>
<td>PSAT</td>
<td>39 (107)</td>
<td>39 (107)</td>
<td>66 (44)</td>
<td>49 (3,401)</td>
<td>76 (77)</td>
<td>58 (56,255)</td>
</tr>
<tr>
<td>SAT</td>
<td>40 (59)</td>
<td>40 (59)</td>
<td>54 (49)</td>
<td>51 (3,142)</td>
<td>82 (58)</td>
<td>72 (56,177)</td>
</tr>
<tr>
<td>Credits</td>
<td>49 (615)</td>
<td>49 (615)</td>
<td>N/A</td>
<td>N/A</td>
<td>95 (379)</td>
<td>86 (293,841)</td>
</tr>
<tr>
<td>AP course</td>
<td>N/A</td>
<td>N/A</td>
<td>8 (13)</td>
<td>9 (1,175)</td>
<td>19 (15)</td>
<td>12 (8,473)</td>
</tr>
<tr>
<td>Graduation</td>
<td>62 (91)</td>
<td>62 (91)</td>
<td>N/A</td>
<td>N/A</td>
<td>100 (79)</td>
<td>100 (69,179)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Philadelphia</th>
<th>Others</th>
<th>Prince George's County</th>
<th>Others</th>
<th>Providence</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ObD (% (N))</td>
<td></td>
<td>ObD (% (N))</td>
<td></td>
<td>ObD (% (N))</td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>34 (69)</td>
<td>32 (3,722)</td>
<td>7 (24)</td>
<td>15 (1,545)</td>
<td>12 (18)</td>
<td>12 (254)</td>
</tr>
<tr>
<td>ELA</td>
<td>46 (63)</td>
<td>34 (3,609)</td>
<td>34 (46)</td>
<td>24 (1,983)</td>
<td>12 (18)</td>
<td>12 (254)</td>
</tr>
<tr>
<td>Attendance</td>
<td>58 (356)</td>
<td>48 (18,941)</td>
<td>N/A</td>
<td>N/A</td>
<td>43 (188)</td>
<td>43 (2,596)</td>
</tr>
<tr>
<td>Suspension</td>
<td>51 (311)</td>
<td>44 (17,192)</td>
<td>N/A</td>
<td>N/A</td>
<td>29 (124)</td>
<td>27 (1,631)</td>
</tr>
<tr>
<td>PSAT</td>
<td>46 (63)</td>
<td>34 (3,622)</td>
<td>N/A</td>
<td>N/A</td>
<td>34 (60)</td>
<td>37 (761)</td>
</tr>
<tr>
<td>SAT</td>
<td>51 (68)</td>
<td>40 (3,247)</td>
<td>N/A</td>
<td>N/A</td>
<td>45 (50)</td>
<td>39 (689)</td>
</tr>
<tr>
<td>Credits</td>
<td>51 (311)</td>
<td>44 (17,192)</td>
<td>N/A</td>
<td>N/A</td>
<td>20 (124)</td>
<td>4 (1,631)</td>
</tr>
<tr>
<td>AP course</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Graduation</td>
<td>61 (82)</td>
<td>49 (4,010)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

NOTES: Brooklyn charter school is not included in this table. Outcome availability is as follows: Math: 47 (39 percent); ELA: 28 (23 percent), Credits: 55 (45 percent); Suspension: 58 (48 percent). N/A indicates that outcomes were not available for a district.

Table A.23 displays the distributions for all available outcome measures in each district.
Table A.23.
Distributions of Outcome Measures, by District

<table>
<thead>
<tr>
<th>Math District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>0.00</td>
<td>-0.18</td>
<td>1.00</td>
<td>-2.37</td>
<td>4.86</td>
</tr>
<tr>
<td>Denver</td>
<td>0.00</td>
<td>-0.11</td>
<td>1.00</td>
<td>-3.21</td>
<td>2.99</td>
</tr>
<tr>
<td>New York City</td>
<td>0.00</td>
<td>0.26</td>
<td>1.00</td>
<td>-4.06</td>
<td>2.25</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.00</td>
<td>-0.16</td>
<td>1.00</td>
<td>-3.25</td>
<td>5.64</td>
</tr>
<tr>
<td>Prince</td>
<td>0.00</td>
<td>-0.11</td>
<td>1.00</td>
<td>-2.49</td>
<td>4.36</td>
</tr>
<tr>
<td>George's</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providence</td>
<td>0.00</td>
<td>-0.02</td>
<td>1.00</td>
<td>-2.55</td>
<td>3.14</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>0.00</td>
<td>0.26</td>
<td>1.00</td>
<td>-4.06</td>
<td>2.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELA District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>0.00</td>
<td>-0.06</td>
<td>1.00</td>
<td>-3.10</td>
<td>3.38</td>
</tr>
<tr>
<td>Denver</td>
<td>0.00</td>
<td>-0.20</td>
<td>1.00</td>
<td>-3.13</td>
<td>2.95</td>
</tr>
<tr>
<td>New York City</td>
<td>0.00</td>
<td>0.26</td>
<td>1.00</td>
<td>-2.57</td>
<td>1.51</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.00</td>
<td>0.03</td>
<td>1.00</td>
<td>-3.09</td>
<td>3.45</td>
</tr>
<tr>
<td>Prince</td>
<td>0.00</td>
<td>0.09</td>
<td>1.00</td>
<td>-1.93</td>
<td>3.11</td>
</tr>
<tr>
<td>George's</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providence</td>
<td>0.00</td>
<td>-0.14</td>
<td>1.00</td>
<td>-2.41</td>
<td>3.20</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>0.00</td>
<td>0.26</td>
<td>1.00</td>
<td>-2.56</td>
<td>1.51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attendance District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>0.92</td>
<td>0.95</td>
<td>0.11</td>
<td>0.012</td>
<td>1.00</td>
</tr>
<tr>
<td>Denver</td>
<td>0.74</td>
<td>0.80</td>
<td>0.23</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>New York City</td>
<td>0.86</td>
<td>0.94</td>
<td>0.21</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.89</td>
<td>0.95</td>
<td>0.16</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Prince</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>George's</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Providence</td>
<td>0.83</td>
<td>0.90</td>
<td>0.19</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>0.86</td>
<td>0.95</td>
<td>0.22</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Suspension District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>0.24</td>
<td>0.00</td>
<td>0.43</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Denver</td>
<td>0.07</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>New York City</td>
<td>0.05</td>
<td>0.00</td>
<td>0.22</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.10</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Prince</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>George's</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Providence</td>
<td>0.08</td>
<td>0.00</td>
<td>0.27</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>0.06</td>
<td>0.00</td>
<td>0.24</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
**Table A.23—Continued**

### PSAT

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>0.00</td>
<td>−0.26</td>
<td>1.00</td>
<td>−3.60</td>
<td>4.15</td>
</tr>
<tr>
<td>Denver</td>
<td>0.00</td>
<td>−0.14</td>
<td>1.00</td>
<td>−3.07</td>
<td>3.21</td>
</tr>
<tr>
<td>New York City</td>
<td>0.00</td>
<td>−0.15</td>
<td>1.00</td>
<td>−3.13</td>
<td>3.26</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.00</td>
<td>−0.17</td>
<td>1.00</td>
<td>−2.80</td>
<td>3.74</td>
</tr>
<tr>
<td>Prince George's</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Providence</td>
<td>0.00</td>
<td>−0.25</td>
<td>1.00</td>
<td>−2.80</td>
<td>4.16</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### SAT

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>0.00</td>
<td>−0.20</td>
<td>1.00</td>
<td>−3.04</td>
<td>4.36</td>
</tr>
<tr>
<td>Denver</td>
<td>0.00</td>
<td>−0.15</td>
<td>1.00</td>
<td>−2.89</td>
<td>3.05</td>
</tr>
<tr>
<td>New York City</td>
<td>0.00</td>
<td>−0.15</td>
<td>1.00</td>
<td>−2.81</td>
<td>2.89</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.00</td>
<td>−0.16</td>
<td>1.00</td>
<td>−2.64</td>
<td>3.51</td>
</tr>
<tr>
<td>Prince George's</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Providence</td>
<td>0.00</td>
<td>−0.23</td>
<td>1.00</td>
<td>−2.63</td>
<td>4.05</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### Credits

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>3.27</td>
<td>4.00</td>
<td>2.38</td>
<td>0.00</td>
<td>48.00</td>
</tr>
<tr>
<td>Denver</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>New York City</td>
<td>11.49</td>
<td>13.00</td>
<td>4.45</td>
<td>0.00</td>
<td>89.38</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>3.65</td>
<td>4.00</td>
<td>1.25</td>
<td>0.00</td>
<td>15.00</td>
</tr>
<tr>
<td>Prince George's</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Providence</td>
<td>11.39</td>
<td>11.00</td>
<td>6.79</td>
<td>0.00</td>
<td>30.50</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>11.32</td>
<td>13.00</td>
<td>4.67</td>
<td>0.00</td>
<td>89.38</td>
</tr>
</tbody>
</table>

### AP Test

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denver</td>
<td>1.63</td>
<td>1.00</td>
<td>0.98</td>
<td>1.00</td>
<td>6.00</td>
</tr>
<tr>
<td>New York City</td>
<td>1.95</td>
<td>2.00</td>
<td>1.19</td>
<td>1.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Prince George's</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Providence</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
### Table A.23—Continued

<table>
<thead>
<tr>
<th></th>
<th>Graduation</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>SD</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Cleveland</td>
<td>0.87</td>
<td>1.00</td>
<td>0.34</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Denver</td>
<td>New York City</td>
<td>0.77</td>
<td>1.00</td>
<td>0.42</td>
<td>0.00</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.79</td>
<td>1.00</td>
<td>0.41</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Prince George's</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Providence</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Brooklyn Lab</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**NOTES:** SD = standard deviation. N/A indicates that outcomes were not available for a district.

### Outcome Measures for 2018–2019 Graduation Analyses

Table A.24 provides the percentage of students with available graduation data for the 2018–2019 school year in districts where we had graduation data.

**Table A.24.**

**Graduation Outcome Measure, by District: Percentages of Available Data**

<table>
<thead>
<tr>
<th>District</th>
<th>ObD % (N)</th>
<th>Others % (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>54 (65)</td>
<td>45 (954)</td>
</tr>
<tr>
<td>Denver</td>
<td>25 (21)</td>
<td>15 (1,238)</td>
</tr>
<tr>
<td>New York City</td>
<td>69 (49)</td>
<td>47 (34,347)</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>60 (58)</td>
<td>47 (3,565)</td>
</tr>
<tr>
<td>Providence</td>
<td>100 (47)</td>
<td>100 (677)</td>
</tr>
</tbody>
</table>

Table A.25 displays the distributions for graduation data in each district for 2018–2019.

**Table A.25.**

**Distributions of Graduation Outcome Measure, by District**

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>0.89</td>
<td>1.00</td>
<td>0.32</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Denver</td>
<td>0.59</td>
<td>1.00</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>New York City</td>
<td>0.71</td>
<td>1.00</td>
<td>0.45</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>0.89</td>
<td>1.00</td>
<td>0.31</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Providence</td>
<td>0.83</td>
<td>1.00</td>
<td>0.37</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Analysis and Results

Table A.26 shows the results for all district outcomes for the 2017–2018 academic year.

Table A.26.
Impact Estimates for District Outcomes, 2017–2018 Academic Year

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Overall</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>−0.170*</td>
<td>−0.173**</td>
<td>0.016</td>
<td>−0.275***</td>
<td>−0.146***</td>
<td>−0.524***</td>
<td>−0.001</td>
<td>−0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.081</td>
<td>0.062</td>
<td>0.082</td>
<td>0.019</td>
<td>0.025</td>
<td>0.016</td>
<td>0.068</td>
</tr>
<tr>
<td>ELA</td>
<td>−0.050</td>
<td>−0.119**</td>
<td>−0.073</td>
<td>0.021</td>
<td>0.005</td>
<td>−0.049</td>
<td>0.083</td>
<td>−0.155**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.029</td>
<td>0.044</td>
<td>0.110</td>
<td>0.045</td>
<td>0.028</td>
<td>0.046</td>
<td>0.193</td>
</tr>
<tr>
<td>Attendance</td>
<td>−0.027</td>
<td>0.460***</td>
<td>0.069</td>
<td>−0.150***</td>
<td>−0.209***</td>
<td>−0.075***</td>
<td>−0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.038</td>
<td>0.088</td>
<td>0.040</td>
<td>0.013</td>
<td>0.031</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Suspension</td>
<td>−0.029</td>
<td>−0.258***</td>
<td>0.107*</td>
<td>−0.171***</td>
<td>0.085**</td>
<td>−0.039***</td>
<td></td>
<td>0.132**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.031</td>
<td>0.049</td>
<td>0.019</td>
<td>0.032</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>PSAT</td>
<td>−0.030</td>
<td>−0.024</td>
<td>−0.049</td>
<td>−0.100**</td>
<td>0.038*</td>
<td></td>
<td>−0.029</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.029</td>
<td>0.035</td>
<td>0.041</td>
<td>0.031</td>
<td>0.018</td>
<td></td>
<td>0.045</td>
</tr>
<tr>
<td>SAT</td>
<td>−0.076**</td>
<td>−0.090**</td>
<td>−0.013</td>
<td>−0.175***</td>
<td>−0.043**</td>
<td>−0.052</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.025</td>
<td>0.033</td>
<td>0.097</td>
<td>0.047</td>
<td>0.013</td>
<td></td>
<td>0.049</td>
</tr>
<tr>
<td>Credits</td>
<td>−0.272**</td>
<td>0.099*</td>
<td>−1.655***</td>
<td>0.060**</td>
<td>0.004</td>
<td>−0.121</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.104</td>
<td>0.045</td>
<td>0.116</td>
<td>0.022</td>
<td>0.011</td>
<td></td>
<td>0.104</td>
</tr>
<tr>
<td>AP course</td>
<td>0.003</td>
<td></td>
<td>0.244</td>
<td>−0.193***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.217</td>
<td></td>
<td>0.144</td>
<td>0.029</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduation</td>
<td>0.018</td>
<td>0.011</td>
<td></td>
<td>0.182***</td>
<td>−0.153**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.106</td>
<td>0.054</td>
<td></td>
<td>0.022</td>
<td>0.054</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: *** p < 0.001, ** p < 0.01, * p < 0.05 before correction for multiple hypothesis tests. Standard errors are presented beneath effect size estimates.

Table A.27 shows the results for graduation outcomes for the 2018–2019 academic year, and Table A.28 shows those for MAP outcomes for the 2017–2018 academic year.
Table A.27.
Impact Estimates for Graduation Outcomes, 2018–2019 Academic Year

<table>
<thead>
<tr>
<th>District</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.007</td>
<td>0.106</td>
</tr>
<tr>
<td>1</td>
<td>0.229*</td>
<td>0.090</td>
</tr>
<tr>
<td>2</td>
<td>0.054</td>
<td>0.147</td>
</tr>
<tr>
<td>4</td>
<td>−0.428*</td>
<td>0.120</td>
</tr>
<tr>
<td>5</td>
<td>0.198***</td>
<td>0.012</td>
</tr>
<tr>
<td>7</td>
<td>−0.086</td>
<td>0.082</td>
</tr>
</tbody>
</table>

NOTE: *** p < 0.001, ** p < 0.01, * p < 0.05 before correction for multiple hypothesis tests.

Table A.28.
Impact Estimates for MAP Outcomes, 2017–2018 Academic Year

<table>
<thead>
<tr>
<th>District</th>
<th>Mathematics Estimate</th>
<th>Mathematics Standard Error</th>
<th>Reading Estimate</th>
<th>Reading Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.053</td>
<td>0.047</td>
<td>0.041</td>
<td>0.069</td>
</tr>
<tr>
<td>1</td>
<td>−0.185</td>
<td>0.088</td>
<td>−0.844***</td>
<td>0.067</td>
</tr>
<tr>
<td>2</td>
<td>0.011</td>
<td>0.080</td>
<td>−0.056</td>
<td>0.174</td>
</tr>
<tr>
<td>3</td>
<td>0.112***</td>
<td>0.012</td>
<td>0.036**</td>
<td>0.011</td>
</tr>
<tr>
<td>5</td>
<td>0.032</td>
<td>0.048</td>
<td>0.107</td>
<td>0.075</td>
</tr>
<tr>
<td>6</td>
<td>0.105**</td>
<td>0.030</td>
<td>0.010</td>
<td>0.024</td>
</tr>
<tr>
<td>7</td>
<td>0.031</td>
<td>0.028</td>
<td>0.156***</td>
<td>0.028</td>
</tr>
</tbody>
</table>

NOTE: *** p < 0.001, ** p < 0.01, * p < 0.05 before correction for multiple hypothesis tests.

Limitations

Several limitations to the impact analyses are acknowledged in the report. First, the outcome measures are broad and may not be sensitive to the ObD model. Second, we expect that the methods mitigate the threat of selection bias but do not remove it entirely. Here, we note several other limitations that are more technical in nature. First, for some outcomes and in some districts, there was a large amount of missing data, and it was not uncommon to encounter situations where over half of the sample was missing either outcome data or the necessary baseline data to be included in the analyses. Our analytic methods account for baseline variable missingness by incorporating information about missingness into the estimation of propensity scores. We also conducted a series of analyses to explore whether there were systematic differences across the ObD and comparison groups in terms of patterns of missingness. These analyses did not suggest that there were systematic differences. However, there is a possibility that data missingness is associated with unmeasured or unobserved student characteristics, and, to the extent that these
unobserved characteristics are associated with both the treatment and the outcome, our causal inferences may be threatened by attrition bias.

Second, we used a random-effect model to meta-analyze the data even though the number of districts was relatively small. Additionally, one of the districts is substantially larger than the others, and, relatedly, the regression estimates in this district tend to have smaller standard errors. In a fixed-effect model, this district then becomes highly influential on the overall effect estimate, since the effect estimate weights are based only on within-district variance—smaller within-district variance produces higher weights. Our decision to use a random-effect model was motivated by the fact that we believe that there was substantial heterogeneity in effects across districts, and that it was unreasonable to assume that all of the districts were functionally identical (Borenstein, Hedges, and Rothstein, 2007). However, with a small number of studies, it is possible that the between-study variance components are imprecisely estimated, which would, in turn, potentially impact the effect estimates and the associated standard errors.

Third, we used cluster-robust standard errors to account for student clustering within schools. The number of schools in each district ranges from over 500 (New York City) to just over ten (Providence). Four of the districts had more than 50 schools. The cluster-robust estimator we employed is based on asymptotic theory, and, for small numbers of clusters, these standard errors might not be correctly specified. Given that the overall effect estimates are generally close to zero and not statistically significant, we report these standard errors and interpret results cautiously. Importantly, while small sample corrections are available, little is known about their performance with many sparsely populated clusters or when used in conjunction with propensity score methods.

Finally, it should be noted that our propensity score estimation approach includes only student-level variables and does not account for the clustering of students in schools or for peer effects. In this way, our propensity estimation approach emulates a random experiment where individual students are randomly assigned to attend ObD schools or to attend other schools in the district. However, as mentioned above, outcomes analyses do account for student clustering within schools (through the use of robust standard errors). Such a design addresses the research question of the effect of enrollment in an ObD school versus another school in the district. However, it does not address school-level effects, such as how implementing ObD practices influenced school-level outcomes. An attempt to address the latter question would have suffered from poor statistical power.
Appendix B. Implementation Analysis, Methods, and Limitations

Implementation Methods

The analyses of implementation of the ObD schools were designed to examine the features of each school’s model, the ways that educators were implementing those features across the four years of the study, the challenges and facilitators associated with implementation, and the contextual factors and supports in each district. The analyses produced information that can be aggregated across schools and districts while also being sensitive to the unique features of each school’s approach. We describe each of our implementation data collection approaches below. Numbers of interview and focus group participants are summarized in Table B.1; survey response rates are summarized in Tables B.2 and B.3. In the last two years of the study (2017 and 2018), we administered selected teacher survey questions to RAND’s American Teacher Panel (ATP) to collect nationally representative data on the prevalence of the instructional practices, supports, and barriers common to ObD schools.

Annual Fall Interviews with District, Springpoint, and Carnegie Corporation of New York Staff

We conducted one-hour telephone interviews with key staff at Carnegie Corporation of New York (CCNY), Springpoint, each ObD district, and intermediary organizations in two districts between fall 2015 and winter 2018, as shown in Table B.1. Interviewers followed semistructured interview protocols to ensure consistency in the questions asked and coverage of important content while also allowing for respondents to elaborate or offer unsolicited input. The interviews helped us gather information about district context, the ways in which the district supported the ObD schools, future plans for implementation, challenges, and successes.

5 The discussion of methods in this section is adapted from Steiner et al., 2017.
Table B.1
Numbers of Interview Participants, Focus Groups, Observations, and Artifacts, 2015–2018

<table>
<thead>
<tr>
<th>Data Collection Method</th>
<th>Source</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviews(^a)</td>
<td>CCNY and Springpoint staff</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>District and intermediary leaders</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>School leaders(^b)</td>
<td>11</td>
<td>13</td>
<td>21</td>
<td>16</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Teachers(^c)</td>
<td>20</td>
<td>46</td>
<td>50</td>
<td>64</td>
<td>180</td>
</tr>
<tr>
<td>Focus groups(^d)</td>
<td>Students (groups/participants)</td>
<td>5/28</td>
<td>10/65</td>
<td>12/69</td>
<td>16/98</td>
<td>43/260</td>
</tr>
<tr>
<td></td>
<td>Parents (groups/participants)</td>
<td>—</td>
<td>4/14</td>
<td>4/17</td>
<td>—</td>
<td>8/31</td>
</tr>
<tr>
<td>Observations(^e)</td>
<td>Classroom</td>
<td>21</td>
<td>42</td>
<td>52</td>
<td>—</td>
<td>116</td>
</tr>
<tr>
<td>Artifacts(^f)</td>
<td>Assignments, assessment reports</td>
<td>17</td>
<td>30</td>
<td>20</td>
<td>—</td>
<td>67</td>
</tr>
</tbody>
</table>

NOTES:  
\(^a\) The total number of interviews represents the count of interviews, not individuals; we interviewed some of the same individuals across years.  
\(^b\) The school leader interview \(N\) is greater than the number of ObD schools in 2015, 2016, and 2017 because in those years, we requested permission to interview up to two leaders in each school. We interviewed all principals, and, in several schools, we also interviewed a second school leader (e.g., vice principal, mastery specialist, design fellow, campus coordinator).  
\(^c\) Includes artifact and design interviews from 2015 through 2017.  
\(^d\) Not all schools provided permission to conduct parent focus groups. We did not conduct parent focus groups in spring 2015 or 2018.  
\(^e\) We did not conduct classroom observations or collect artifacts in spring 2018.  
\(^f\) We conducted two-day, in-person visits at each ObD school in spring 2015, 2016, and 2017, and one-day, in-person visits at each school in spring 2018. The purposes of the site visits were to gather in-depth information about implementation of the school model and instructional practices and to solicit student and parent perspectives. During each visit, we interviewed the school principal. In study years two and three (2015–2017), we interviewed another school leader in a position to provide insight on implementation of the design principles. At each school, we selected four teachers to participate in 45-minute interviews. If no second school leader was available in 2015–2017, we interviewed a fifth teacher. We also conducted one-hour focus groups with six to eight students. In spring 2018, total teaching staff sizes in these schools ranged from about ten in Cohort IV schools to about 20 in Cohort I schools. 2018 student enrollment ranged from about 100 students in Cohort IV schools to about 400 in Cohort I schools.

In 2015, 2016, and 2017, two teachers at each school participated in artifact-based interviews, and two or three (depending on the school) participated in school design interviews. The artifact-based interviews were intended to capture evidence regarding some important
aspects of instructional practice. Teachers were asked to bring class assignments, assessment criteria or rubrics, and examples of student work to the artifact interviews. In 2018, we did not conduct separate artifact interviews or collect instructional artifacts; instead, we incorporated specific questions about a recent assignment into the design interview. In the school design interviews, we asked teachers to describe aspects of the school design, along with perceived challenges and facilitators. In 2015–2017, we also conducted 10- to 15-minute observations of four to six classrooms where mathematics or ELA instruction was taking place. In 2016 and 2017, we conducted one-hour focus groups with four to eight parents in schools in which the principal gave permission for the groups to take place. Counts of interviews and focus groups conducted, classrooms observed, and artifacts collected are shown in Table B.1.

We selected teachers for the interviews to ensure variability in years of teaching experience, subjects taught, and grade level, if applicable. A school administrator selected students for the focus groups so that the group would include students with a mix of ages, interests, and learning levels, as well as students of both genders. We used semistructured interview and focus group protocols to promote consistency in the questions asked across schools and to ensure coverage of important content while also allowing for respondents to elaborate or offer unsolicited input. The classroom observation protocol was open-ended to allow observers to capture the diversity of instructional approaches and classroom arrangements. The protocol captured classroom conditions, such as student-to-adult ratios, presence of technology, type and content of instruction, teacher and student interactions, and the nature of student groupings.

**Annual Spring Teacher Surveys**

Teachers of core academic content areas (i.e., mathematics, ELA, social studies, and science) were invited to participate in web-based surveys in the spring of each study year, 2015–2018. The surveys gathered systematic information about teachers’ perceptions about various aspects of the school models, including professional training and support, access to resources, the quality of instructional and curriculum materials, use of different models of classroom instruction, use of technology in the classroom, use of data to assess student progress, and obstacles to implementation. The teacher survey took approximately 30 minutes to complete. Although many of the survey items were developed specifically for this study, several were adapted from other RAND surveys (including those used in Pane et al., 2015) or from surveys developed by the University of Chicago Consortium on School Research (1999). Teacher survey response rates are displayed in Table B.2.
### Table B.2.
Teacher Survey Response Rates, by School, 2015–2018

<table>
<thead>
<tr>
<th>Cohort</th>
<th>School</th>
<th>2015 N</th>
<th>Response Rate (%)</th>
<th>2016 N</th>
<th>Response Rate (%)</th>
<th>2017 N</th>
<th>Response Rate (%)</th>
<th>2018 N</th>
<th>Response Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>J</td>
<td>4</td>
<td>50</td>
<td>9</td>
<td>75</td>
<td>11</td>
<td>69</td>
<td>13</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>5</td>
<td>56</td>
<td>11</td>
<td>100</td>
<td>14</td>
<td>67</td>
<td>13</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>6</td>
<td>86</td>
<td>7</td>
<td>78</td>
<td>7</td>
<td>58</td>
<td>18</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>5</td>
<td>83</td>
<td>5</td>
<td>63</td>
<td>10</td>
<td>67</td>
<td>8</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4</td>
<td>80</td>
<td>6</td>
<td>86</td>
<td>9</td>
<td>75</td>
<td>9</td>
<td>75</td>
</tr>
<tr>
<td>II</td>
<td>K</td>
<td>—</td>
<td>—</td>
<td>4</td>
<td>80</td>
<td>3</td>
<td>33</td>
<td>13</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>—</td>
<td>—</td>
<td>4</td>
<td>80</td>
<td>9</td>
<td>82</td>
<td>13</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>—</td>
<td>—</td>
<td>3</td>
<td>60</td>
<td>9</td>
<td>64</td>
<td>16</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>—</td>
<td>—</td>
<td>5</td>
<td>83</td>
<td>6</td>
<td>75</td>
<td>13</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>—</td>
<td>—</td>
<td>7</td>
<td>100</td>
<td>4</td>
<td>40</td>
<td>11</td>
<td>92</td>
</tr>
<tr>
<td>III</td>
<td>F</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>12</td>
<td>100</td>
<td>17</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>5</td>
<td>71</td>
<td>19</td>
<td>74</td>
</tr>
<tr>
<td>IV</td>
<td>E</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>9</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>7</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>10</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>5</td>
<td>56</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>24</td>
<td>69</td>
<td>61</td>
<td>81</td>
<td>99</td>
<td>67</td>
<td>189</td>
<td>81</td>
</tr>
</tbody>
</table>

**NOTE:** Cohort II schools opened in fall 2015, Cohort III schools opened in fall 2016, and Cohort IV schools opened in fall 2017.

### Annual Fall and Spring Student Surveys

Students were invited to participate in a brief (20- to 30-minute) online survey in each fall and spring of the study (fall 2014 through spring 2018). The fall survey included questions about study habits, attitudes toward learning, and goals for high school and beyond. The spring survey included the questions asked in the fall, along with additional questions about students’ perceptions of their school and classroom environments. We offered the survey in a variety of languages (e.g., English, Spanish, French, Arabic) to all students after consulting with district staff about students’ language needs. As shown in Table B.3, in 2017–2018, fall response rates among students with consent\(^6\) ranged from 71 percent to 96 percent, with an overall response rate of 86 percent. In total 3,143 students were eligible to participate, and 2,695 students participated. Numbers of eligible students in each school ranged from 89 to 313. Spring response rates among students with consent ranged from 44 percent to 97 percent, with an overall response rate of 85 percent.

---

\(^6\) In most schools, we were able to obtain passive consent from parents to allow their children to participate and therefore had consent from most students (zero to ten students per school were opted out of the survey by their parents). One school required active parental consent; therefore, for that school we had consent from the parents of fewer students.
percent. In total, 2,261 students were eligible to participate, and 1,918 students participated. Numbers of eligible students in each school ranged from 85 to 296.

As with the teacher surveys, we developed many of the items specifically for this study, but the surveys also included original or modified versions of items from the University of Chicago Consortium on School Research’s *My Voice, My School* student survey (2015); the High School Survey of Student Engagement, developed by the Center for Evaluation, Policy, and Research at Indiana University (2013); and the RAND survey of student perceptions of personalized learning practices (Pane et al., 2015).
Table B.3.
Student Survey Response Rates, by School, 2014–2015 to 2017–2018

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fall</td>
<td>Spring</td>
<td>Fall</td>
<td>Spring</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>N (%)</td>
<td>RR</td>
<td>N (%)</td>
<td>RR</td>
</tr>
<tr>
<td>I</td>
<td>J</td>
<td>91</td>
<td>83</td>
<td>76</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>109</td>
<td>92</td>
<td>97</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>54</td>
<td>53</td>
<td>78</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>105</td>
<td>87</td>
<td>87</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>78</td>
<td>79</td>
<td>71</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>88</td>
<td>90</td>
<td>82</td>
<td>88</td>
</tr>
<tr>
<td>II</td>
<td>K</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>III</td>
<td>F</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>121</td>
<td>95</td>
<td>109</td>
<td>83</td>
</tr>
<tr>
<td>IV</td>
<td>E</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>437</td>
<td>79</td>
<td>409</td>
<td>81</td>
</tr>
</tbody>
</table>

NOTE: Response rates (RRs) reported are among students with consent. Cohort I schools opened in fall 2014, Cohort II schools opened in fall 2015, Cohort III schools opened in fall 2016, and Cohort IV schools opened in fall 2017.
Annual Collection of Artifacts

We asked CCNY, Springpoint, district, and school interview participants to provide us with artifacts relevant to understanding ObD implementation in the first three years of the study. Examples of such artifacts included materials that CCNY used to assess district enabling conditions; design documents, such as work plans, school handbooks, or competency maps; instructional materials, such as rubrics or lesson plans; and other items, such as materials from professional development (PD) sessions. We reviewed these documents to inform our understanding of the school designs, contexts, and instructional practices.

American Teacher Panel

In spring 2018, we administered questions from the ObD teacher survey to RAND’s ATP to collect nationally representative data on the prevalence of the instructional practices, supports, and barriers common to ObD schools. In spring 2018, we administered the entire ObD teacher survey to 1,818 teachers who taught at least one core academic subject (mathematics, ELA, science, or social studies) in at least one high school grade (9–12) to be consistent with the ObD teacher sample. The 2018 analytic sample contained 995 teachers who fully completed the survey and 14 teachers who partially completed the survey, for a total of 1,009 teachers (56-percent response rate).

Analytic Methods

Qualitative Data

We analyzed the qualitative data using Dedoose software, which allowed us to code for common themes across data collection sites and across sources (e.g., teachers, district administrators). The analysis of the interview and focus group data proceeded in several steps. First, interview notes were compared with the audio recording and cleaned to serve as a near-transcript of the conversation. We engaged in member checking as appropriate to ensure data accuracy. The cleaned interview notes were then loaded into Dedoose and coded using a thematic codebook developed by the evaluation team to align with the ten design principles. Once the thematic coding was complete, we conducted a second round of coding, analyzing the data according to questions of interest (e.g., to what extent are schools implementing mastery-based progression?). In this stage, we used an inductive coding process (i.e., codes were derived from the data rather than from a structured codebook) to develop responses to the questions of interest. Coders double-coded selected transcripts and discussed coding decisions throughout the process to ensure reliability. Finally, we summarized implementation of each design principle across schools. We also summarized district, CCNY, and Springpoint perspectives on implementation.
**ObD Survey Data**

We analyzed the quantitative data from the surveys of ObD teachers and students using statistical software (SAS and R). We calculated item means and frequencies and conducted exploratory factor analyses for the student and teacher surveys to assess the appropriateness of combining individual items into multi-item scales.

We investigated whether there were systematic differences in students’ experiences with mastery-based, personalized learning and PYD practices across grade levels by analyzing student survey responses by grade level. We used four multi-item scales that captured students’ experiences with these practices as outcomes in four generalized least squares models analyzing whether there were differences by grade level while holding all other variables in the model constant (see Table D.10 for scale names and constituent items). The outcome of interest in each of the four models was the relevant multi-item scale, and the covariates were indicators for grade level (ninth-graders were the reference level) and each school. We controlled for school to better understand whether the descriptive student grade-level differences were driven by differences in school-level factors.

To identify whether a subset of schools stood out as having higher or lower reported implementation relative to the others, we analyzed the 2018 teacher survey data by school to discern whether teachers’ responses at any school were systematically higher or lower than those at other schools. We first dichotomized responses for each item into “high” and “low” responses and then sorted from highest agreement to lowest agreement. To dichotomize, for agreement questions, we added the percentage of teachers at a given school who indicated that they agreed or strongly agreed with a question item. When questions used a different response scale, we added the percentage of teachers at a given school who answered a question positively—examples of positive responses would include “used data to a moderate extent/used data to a large extent” or “about half/more than half/all of the materials.” We wanted to understand which schools had the highest (and lowest) agreement on each question, relative to the others. We therefore classified a school as having highest agreement if the school-level average was one of the three highest compared with other schools; we classified a school as having lowest agreement if the school-level agreement was one of the three lowest compared with other schools. We then looked at whether the same schools had the highest or lowest agreement across several questions. One school stood out from this analysis as having higher reported implementation of the design principles relative to the others. This school is the subject of the case study presented in Chapter Three.

**American Teacher Panel Data**

The ATP is a nationally representative web-based survey administered by RAND to a randomly selected sample of full-time public school teachers. The ATP was weighted to represent the United States using various baseline and demographic characteristics, including
subject, school level, region, size, minority percentage, geographic locale, and FRL eligibility. For additional information on weighting, please refer to Robbins and Grant (2020). We analyzed the ATP data using statistical software (R). We calculated item means and frequencies and conducted exploratory factor analyses for the student and teacher surveys to assess the appropriateness of combining individual items into multi-item scales. We compared ATP teachers’ responses with ObD teachers’ responses by conducting t-tests on the item means. We conducted tests of significance on differences between the item means rather than on the categorical results to better capture differences throughout the full distribution of responses.

Limitations

The implementation data are drawn from a variety of sources and provide a rich picture of ObD model implementation. At the same time, readers should keep in mind the limitations of the data sources. In particular, the survey and interview data rely on the self-reports of stakeholders who voluntarily participated. We have no independent means of verifying the accuracy of their responses. Where response rates are lower, particularly for the teacher survey in some schools and some years, responses may not accurately represent the perceptions of the entire teaching staff, limiting generalizability. Moreover, although the interview data are crucial for providing richness and context, the numbers of interview participants are small in many cases, and the teachers and students who participated in the interviews and focus groups are not representative samples of the full populations of teachers and students in the ObD schools.

Our implementation data collection efforts were subject to time and resource constraints, and we were not able to collect data about every aspect of the schools’ operations. The lack of information about such topics as ObD teachers’ working conditions (e.g., class sizes, schedules, or compensation), or how these conditions were different from other high schools in the ObD districts or high schools nationally, limits our understanding of the contexts in which the design principles were implemented. These and other working conditions could affect the time teachers spend on key aspects of the school model, such as curriculum development, data analysis, or mastery grading.

Finally, we were not able to uncover any meaningful insights about the extent to which differences in implementation might relate to differences in student outcomes. We did not have the statistical power to analyze these relationships using regression analyses or other quantitative methods, and our visual inspection of the data did not produce any meaningful results. The absence of substantive findings could be attributable to the small number of ObD schools and the lack of variability on some implementation measures. In addition, the quantitative analyses did not identify strong overall effects on any student outcomes, and the absence of some outcomes in a few districts made it difficult to identify any patterns.
Appendix C. ObD Design Principles

- **Design Principle 1**: A high-performing secondary school has a *clear mission and coherent culture* that is evidenced by a clearly defined purpose, goals, and school culture. The school’s mission and culture is embodied in all aspects of school design.\(^7\)

- **Design Principle 2**: A high-performing secondary school *prioritizes mastery of rigorous standards* aligned to college and career readiness and has curriculum that enables all students to meet rigorous standards, multiple opportunities for students to show mastery through performance-based assessments, and student advancement based on demonstration of mastery of knowledge and skills.

- **Design Principle 3**: A high-performing secondary school *personalizes learning to meet student needs* such that instruction is offered in a variety of learning modalities; is linked to students’ strengths and learning goals; is data-driven, with real-time feedback for students and teachers; and incorporates embedded, performance-based formative assessments. Technology is used effectively to facilitate anytime, anywhere learning.

- **Design Principle 4**: A high-performing secondary school *maintains an effective human capital strategy* aligned with the school model and priorities. The human capital strategy includes consistent, high-quality systems for sourcing and selecting teachers and staff; individualized PD that cultivates teachers’ strengths and meets school needs and priorities, including use of blended learning; fair and equitable teacher evaluation; leadership development opportunities; and a leadership pipeline.

- **Design Principle 5**: A high-performing secondary school *develops and deploys collective strengths* such that teaching in teams strengthens instructional design and delivery and enables professional growth. School designs should include mechanisms that promote opportunities for innovation and initiative among teachers and staff. Differentiated roles for adults (e.g., multiple “teacher” roles) enable effective implementation of the school model.

- **Design Principle 6**: A high-performing secondary school *remains porous and connected* such that it cultivates and maintains effective partnerships with organizations that enrich student learning and increase access to community resources and supports. The school should also participate in a network of schools that share knowledge and assets.

- **Design Principle 7**: A high-performing secondary school *integrates PYD to optimize student engagement and effort* in a way that fosters caring, consistent student-adult relationships that communicate high expectations for student learning and behavior;

\(^7\) The definitions of the design principles were developed by CCNY.
allows adults to communicate clear expectations for student competencies and standards of performance; and provides opportunities for students to contribute to the school environment and have a voice in decisions. The school also encourages student responsibility for meeting learning and personal goals; openness to and encouragement of family participation; and integration of community participation, assets, and culture.

- **Design Principle 8:** A high-performing secondary school *empowers and supports students through key transitions* (i.e., into and beyond high school) such that explicit linkages between future academic and career pathways and current learning and activities are apparent, and there is transparency regarding student status and progress toward graduation for students and parents/guardians.

- **Design Principle 9:** A high-performing secondary school *manages school operations efficiently and effectively* such that time, people, and technology are used purposefully to optimize teachers’ ability to support student learning; all elements of school design are organized to maximize efficient use of resources; scheduling is flexible and customized; there are clear operational performance goals and accountability mechanisms; and basic tasks are automated whenever possible.

- **Design Principle 10:** A high-performing secondary school *continuously improves its operations and model* such that performance data and analytics are used to improve curriculum and instruction, and there is regular review and revision of school operations and model to increase effectiveness.
Appendix D. Student Survey Results

Table D.1. Frequency of Tasks Assigned to Students

<table>
<thead>
<tr>
<th>Type of Task</th>
<th>Percentage of Students Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never or Hardly Ever</td>
</tr>
<tr>
<td>Memory-based or repetitive tasks, like solving a page of equations, filling in the blanks, or taking a short quiz</td>
<td>9</td>
</tr>
<tr>
<td>Basic comprehension tasks, like answering short questions or writing a summary based on what you learned</td>
<td>5</td>
</tr>
<tr>
<td>Connecting or applying what you learned to real-world problems</td>
<td>10</td>
</tr>
<tr>
<td>Projects where you worked with other students over more than one class period</td>
<td>7</td>
</tr>
</tbody>
</table>

NOTES: Question text: Think about all the assignments and tasks you work on in class and out of class. How often do your teachers assign each of the following types of tasks? Response options: Never or hardly ever; Sometimes (1–3 times per month); Often (1–3 times per week); Daily or almost daily. Percentages may not sum to 100 percent because of rounding. N = 2,275–2,281.

Table D.2. Student Perceptions of Assigned Tasks

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage of Students Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td>The assignments the teachers give have a clear purpose; I know how they help me to learn what I’m supposed to learn.</td>
<td>5</td>
</tr>
<tr>
<td>The instructions the teachers give about what I need to do for each assignment are clear.</td>
<td>4</td>
</tr>
<tr>
<td>The way the teachers assess my work is fair.</td>
<td>4</td>
</tr>
<tr>
<td>Based on the feedback teachers give me on my work, I know how to improve in future assignments.</td>
<td>5</td>
</tr>
</tbody>
</table>

NOTES: Question text: Again, think about all the assignments and tasks you work on in class and out of class. How much do you agree with the following statements? Response options: Strongly disagree; Disagree; Agree; Strongly agree. Percentages may not sum to 100 percent because of rounding. N = 2,275–2,280.
### Table D.3.
**Student Attitudes Toward Doing Schoolwork**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage of Students Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td>I pay attention and try not to get distracted when I’m doing schoolwork.</td>
<td>Strongly Disagree</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>I make time to do schoolwork outside of class time.</td>
<td>6</td>
</tr>
<tr>
<td>I try to do well on my schoolwork even when it isn’t interesting to me.</td>
<td>4</td>
</tr>
</tbody>
</table>

NOTES: Question text: How much do you agree with the following statements? Response options: Strongly disagree; Disagree; Agree; Strongly agree. Percentages may not sum to 100 percent because of rounding. $N = 2,276–2,277$.

### Table D.4.
**Submission Frequency of Students’ Completed Work**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage of Students Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete all the classwork and activities you are assigned</td>
<td>Never</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Submit your work by the due date</td>
<td>3</td>
</tr>
<tr>
<td>Submit your work with enough time to get feedback and make changes before it is due</td>
<td>5</td>
</tr>
<tr>
<td>Revise or make changes to your work when given the opportunity to improve on it</td>
<td>3</td>
</tr>
</tbody>
</table>

NOTES: Question text: How often do you...? Response options: Never; Once in a while; Usually; Always. Percentages may not sum to 100 percent because of rounding. $N = 2,281–2,283$. 
Table D.5.
Student Perceptions of Their Teachers in Typical Classroom Experiences

<table>
<thead>
<tr>
<th>Classroom Experience</th>
<th>Percentage of Students Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td>I like the way we learn in my classes.</td>
<td>Totally Untrue  Mostly Untrue  Somewhat Untrue  Mostly True  Totally True</td>
</tr>
<tr>
<td>My classes really make me think.</td>
<td>6  8  27  42  17</td>
</tr>
<tr>
<td>The feedback that I get on my schoolwork helps me understand how to improve.</td>
<td>4  6  25  43  21</td>
</tr>
<tr>
<td>Students get to decide how activities are done in my classes.</td>
<td>9  12  31  35  14</td>
</tr>
<tr>
<td>My teachers don't let students give up when the work gets hard.</td>
<td>5  8  23  40  24</td>
</tr>
<tr>
<td>My teachers want students to explain our answers—why we think what we think.</td>
<td>3  5  20  44  28</td>
</tr>
<tr>
<td>My teachers check to make sure students understand what we are learning.</td>
<td>3  6  22  43  27</td>
</tr>
<tr>
<td>My teachers respect my ideas and suggestions.</td>
<td>4  6  23  43  25</td>
</tr>
<tr>
<td>In my classes, we learn a lot almost every day.</td>
<td>4  7  24  43  23</td>
</tr>
<tr>
<td>The material I am learning in my classes is interesting.</td>
<td>5  8  28  41  19</td>
</tr>
<tr>
<td>The activities and assignments I work on are challenging, but not so difficult I can't complete them.</td>
<td>3  7  26  44  20</td>
</tr>
<tr>
<td>My teachers take into account my interests and experiences when deciding what I will work on.</td>
<td>5  8  26  42  18</td>
</tr>
<tr>
<td>I am able to work on activities and assignments that fit my learning needs.</td>
<td>4  6  24  46  20</td>
</tr>
<tr>
<td>My classmates and I have opportunities to work together and give each other feedback.</td>
<td>3  6  22  45  24</td>
</tr>
<tr>
<td>My teachers and I work together to set personal goals for my own learning.</td>
<td>5  8  26  40  20</td>
</tr>
<tr>
<td>My teachers connect what we are learning to the world outside of school.</td>
<td>4  8  25  42  21</td>
</tr>
<tr>
<td>My teachers help us understand why it's important to learn what we are learning.</td>
<td>4  6  24  42  24</td>
</tr>
<tr>
<td>My instructors have high expectations for my academic performance.</td>
<td>4  5  21  44  27</td>
</tr>
</tbody>
</table>

NOTES: Question text: The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all of your teachers this year, and choose the response that best matches your typical experience. Response options: Totally true; Mostly true; Somewhat untrue; Mostly true; Totally true. Percentages may not sum to 100 percent because of rounding. N = 2,252–2,270.
### Table D.6.
**Frequency of Students’ Typical Classroom Experiences**

<table>
<thead>
<tr>
<th>Classroom Experience</th>
<th>Percentage of Students Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have opportunities to choose which instructional materials (such as books or computer software) I use in class.</td>
<td>Never: 7  Rarely: 11  Sometimes: 34  Most of the Time: 31  Always: 18</td>
</tr>
<tr>
<td>I have opportunities to choose which topics I focus on in class.</td>
<td>Never: 7  Rarely: 13  Sometimes: 36  Most of the Time: 30  Always: 14</td>
</tr>
<tr>
<td>I am required to show that I understand a topic before I move onto a new topic.</td>
<td>Never: 4  Rarely: 8  Sometimes: 33  Most of the Time: 36  Always: 19</td>
</tr>
<tr>
<td>I work on different topics or skills than what my classmates are working on at the same time.</td>
<td>Never: 6  Rarely: 11  Sometimes: 36  Most of the Time: 32  Always: 16</td>
</tr>
<tr>
<td>I am given the chance to work through instructional material at a faster or slower pace than other students in this class.</td>
<td>Never: 5  Rarely: 8  Sometimes: 34  Most of the Time: 35  Always: 19</td>
</tr>
<tr>
<td>When I am working on an assignment or activity, I know what the goals of the assignment or activity are.</td>
<td>Never: 3  Rarely: 6  Sometimes: 32  Most of the Time: 38  Always: 21</td>
</tr>
<tr>
<td>I keep track of my learning progress using technology (for example, by using an online gradebook or portfolio).</td>
<td>Never: 4  Rarely: 7  Sometimes: 30  Most of the Time: 37  Always: 23</td>
</tr>
<tr>
<td>I have opportunities to review or practice new material until I really understand it.</td>
<td>Never: 3  Rarely: 7  Sometimes: 30  Most of the Time: 37  Always: 23</td>
</tr>
</tbody>
</table>

**NOTES:** Question text: The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all of your classes this year, and mark the response that indicates your typical experience. Response options: Never; Rarely; Sometimes; Most of the time; Always. Percentages may not sum to 100 percent because of rounding. $N = 2,256–2,263$.

### Table D.7.
**Student Perceptions of Classmates**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage of Students Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td>My classmates usually behave the way my teacher wants them to.</td>
<td>Not at All: 12  A Little: 46  Somewhat: 26  A Lot: 16</td>
</tr>
<tr>
<td>Students help each other understand the material in class.</td>
<td>Not at All: 6  A Little: 48  Somewhat: 22  A Lot: 24</td>
</tr>
<tr>
<td>When someone makes a mistake, students try to help him or her.</td>
<td>Not at All: 8  A Little: 46  Somewhat: 22  A Lot: 24</td>
</tr>
<tr>
<td>My classmates think it’s important to really understand what we’re learning.</td>
<td>Not at All: 8  A Little: 47  Somewhat: 20  A Lot: 24</td>
</tr>
</tbody>
</table>

**NOTES:** Question text: How much do you agree with the following statements about your classmates? Response options: Not at all; A little; Somewhat; A lot. Percentages may not sum to 100 percent because of rounding. $N = 2,264–2,267$. 

---

45
### Table D.8.
**Extent to Which Students’ Teachers Addressed Social and Emotional Learning Topics in the Classroom**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Not Addressed</th>
<th>To a Small Extent</th>
<th>To a Moderate Extent</th>
<th>To a Large Extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborating with other students</td>
<td>5</td>
<td>19</td>
<td>52</td>
<td>25</td>
</tr>
<tr>
<td>Understanding and managing emotions</td>
<td>6</td>
<td>21</td>
<td>53</td>
<td>20</td>
</tr>
<tr>
<td>Handling stress</td>
<td>9</td>
<td>23</td>
<td>48</td>
<td>20</td>
</tr>
<tr>
<td>Persisting through challenging tasks</td>
<td>5</td>
<td>20</td>
<td>52</td>
<td>23</td>
</tr>
<tr>
<td>Making responsible decisions</td>
<td>5</td>
<td>17</td>
<td>52</td>
<td>26</td>
</tr>
<tr>
<td>Establishing and maintaining positive relationships</td>
<td>6</td>
<td>20</td>
<td>51</td>
<td>23</td>
</tr>
<tr>
<td>Feeling and showing empathy for others</td>
<td>7</td>
<td>19</td>
<td>50</td>
<td>24</td>
</tr>
<tr>
<td>Learning mindsets (e.g., growth mindset, sense of purpose and belonging)</td>
<td>6</td>
<td>17</td>
<td>52</td>
<td>27</td>
</tr>
<tr>
<td>Communicating their thoughts and emotions</td>
<td>7</td>
<td>19</td>
<td>51</td>
<td>23</td>
</tr>
<tr>
<td>Developing a sense of identity</td>
<td>6</td>
<td>17</td>
<td>50</td>
<td>26</td>
</tr>
</tbody>
</table>

NOTES: Question text: Please indicate to what extent your teachers addressed each of the following topics in your classes. Response options: Not addressed; To a small extent; To a moderate extent; To a large extent. Percentages may not sum to 100 percent because of rounding. N = 2,245–2,256.

### Table D.9.
**2017–2018 Spring Student Survey Scales**

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>SCL Practices (N = 2,246)</th>
<th>Personalized Instructional Practices (N = 2,235)</th>
<th>Choice (N = 2,251)</th>
<th>Mastery (N = 2,252)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.456**</td>
<td>3.629**</td>
<td>3.329**</td>
<td>3.595**</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.067)</td>
<td>(0.076)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>Tenth grade</td>
<td>0.028</td>
<td>0.014</td>
<td>−0.006</td>
<td>−0.013</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.046)</td>
<td>(0.053)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>11th grade</td>
<td>0.127</td>
<td>0.107</td>
<td>0.070</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.052)</td>
<td>(0.059)</td>
<td>(0.553)</td>
</tr>
<tr>
<td>12th grade</td>
<td>0.245**</td>
<td>0.201**</td>
<td>0.260**</td>
<td>0.176**</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.069)</td>
<td>(0.073)</td>
<td>(0.064)</td>
</tr>
</tbody>
</table>

NOTES: **p < 0.01. SCL = student-centered learning. Models are fit using a generalized linear model; we report the estimate (standard error). Scales (the outcome) were made by averaging responses to questions for each construct. All four models included school as a fixed effect, and grade-level estimates are relative to ninth-graders.
### Measure Name: SCL Practices (Underlying Likert Scale 1–5)

**Question:** The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all your teachers this year, and choose the response that best matches your typical experience.

- I like the way we learn in my classes.
- Students get to decide how activities are done in my classes.
- My teachers take into account my interests and experiences when deciding what I will work on.
- My classmates and I have opportunities to work together and give each other feedback.

### Measure Name: Personalized Instructional Practices (Underlying Likert Scale 1–5)

**Question:** The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all your teachers this year, and choose the response that best matches your typical experience.

- The feedback that I get on my schoolwork helps me understand how to improve.
- My teachers don't let students give up when the work gets hard.
- My teachers want students to explain out answers—why we think what we think.
- My teachers check to make sure students understand what we are learning.
- My teachers respect my ideas and suggestions.
- In my classes, we learn a lot almost every day.
- The material I am learning in my classes is interesting.
- My teachers and I work together to set personal goals for my own learning.
- My teachers connect what we are learning to the world outside of school.
- My teachers help us understand why it's important to learn what we are learning.

### Measure Name: Choice (Underlying Likert Scale 1–5)

**Question:** The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all of your classes this year, and mark the response that indicates your typical experience.

- I have opportunities to choose which instructional materials (such as books or computer software) I use in class.
- I have opportunities to choose which topics I focus on in class.
- I work on different topics or skills than what my classmates are working on at the same time.

### Measure Name: Mastery (Underlying Likert Scale 1–5)

**Question:** The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all of your classes this year, and mark the response that indicates your typical experience.

- I am required to show that I understand a topic before I move onto a new topic.
- I am given the chance to work through instructional material at a faster or slower pace than other students in this class.
- When I am working on an assignment or activity, I know what the goals of the assignment or activity are.
- I keep track of my learning progress using technology (for example, by using an online gradebook or portfolio).
- I have opportunities to review or practice new material until I really understand it.

**NOTE:** SCL = student-centered learning.
## Appendix E. Teacher Survey Results

Table E.1.

ObD Teachers’ Perceptions of Working Conditions at Their School

<table>
<thead>
<tr>
<th>Working Condition</th>
<th>Percentage of ObD Teachers Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly Disagree</td>
</tr>
<tr>
<td>Teachers at my school engage in regular, productive conversations with one another about how to improve instruction.</td>
<td>4</td>
</tr>
<tr>
<td>Teachers at my school are highly focused on improving student learning.</td>
<td>1</td>
</tr>
<tr>
<td>Teachers at my school have the skills needed to foster meaningful student learning.</td>
<td>2</td>
</tr>
<tr>
<td>Teachers at my school support each other in their efforts to improve student learning.</td>
<td>2</td>
</tr>
<tr>
<td>Teachers at my school really believe every child can learn and be college and career ready.</td>
<td>2</td>
</tr>
<tr>
<td>Teachers at my school are fully prepared to teach based on the Common Core State Standards, Next Generation Science Standards or other relevant standards.</td>
<td>1</td>
</tr>
<tr>
<td>I feel respected as a colleague by staff members at my school.</td>
<td>2</td>
</tr>
<tr>
<td>If I had concerns about my school, I would feel comfortable raising them with administrators at the school.</td>
<td>12</td>
</tr>
<tr>
<td>Administrators at my school are highly supportive of teachers.</td>
<td>15</td>
</tr>
<tr>
<td>Administrators at my school are highly focused on student learning.</td>
<td>13</td>
</tr>
<tr>
<td>Administrators at my school trust teachers to make decisions about their own instruction.</td>
<td>12</td>
</tr>
<tr>
<td>Students in this school respect one another.</td>
<td>8</td>
</tr>
<tr>
<td>Students in this school respect the school staff.</td>
<td>9</td>
</tr>
<tr>
<td>Students in this school are motivated to achieve.</td>
<td>9</td>
</tr>
</tbody>
</table>

**NOTES:** Question text: Please rate your level of agreement with each of the following statements about your school. Response options: Strongly disagree; Disagree; Agree; Strongly agree. Percentages may not sum to 100 percent because of rounding. \( N = 187–188. \)
<table>
<thead>
<tr>
<th>Condition</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N/A</td>
<td>Not an Obstacle</td>
</tr>
<tr>
<td>Too much diversity in achievement levels among my students</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Lack of flexibility in the curriculum I am required to teach (i.e., need to teach specific material in a specific time frame)</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Pressure to cover specific material as a result of state or district standards or testing requirements</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Excessive amounts of time I need to spend developing content that meets individual students’ needs</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>Inadequate opportunities to participate in professional development related to personalizing learning</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>Inadequate data to help me personalize instruction for students</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>Lack of high-quality content or materials</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>High levels of student <em>absenteeism</em> **</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>High levels of student <em>disciplinary problems</em> **</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Scheduling constraints</td>
<td>0</td>
<td>25</td>
</tr>
</tbody>
</table>

NOTES: * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01. Our national teacher data are based on results from the ATP. Question text: Please indicate the extent to which each of the following conditions is an obstacle to your efforts to personalize students’ learning to address their individual learning needs and interests. If the condition does not exist in your school, please mark N/A—condition does not exist in my school. Response options: N/A—condition does not exist in my school; Condition exists but is not an obstacle; Condition exists and is a minor obstacle; Condition exists and is a major obstacle. ObD N = 92–178; ATP N = 1,005–1,006.
### Table E.3a.
Frequency with Which Teachers Received Student Performance Information

<table>
<thead>
<tr>
<th>Type of Student Performance Information</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never</td>
<td>Once a Year</td>
</tr>
<tr>
<td>Scores on district or state assessments in mathematics**</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Scores on district or state assessments in language arts**</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Scores on district or state assessments in science</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>Scores on district or state assessments in subjects other than mathematics, language arts, or science**</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>Information about student performance on specific concepts or skills**</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>Assessment data that are built into curriculum software**</td>
<td>31</td>
<td>7</td>
</tr>
<tr>
<td>Identification of specific students who need extra assistance**</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Identification of specific students who have achieved mastery**</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Youth development outcomes (for example, student behavior, attitudes, or motivation)**</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>Identification of specific students who are at risk of dropping out or not progressing to the next grade**</td>
<td>13</td>
<td>12</td>
</tr>
</tbody>
</table>

NOTES: * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01.

Question text: In general, how frequently do you receive the following types of information about the performance of your students? Response options: Never; Once a year; A few times per year; Approximately monthly; A few times per month; Approximately weekly; A few times per week; At least daily. (Percentages for the final four response options are shown in Table E.3b.) ObD $N = 182–183$; ATP $N = 1,002–1,004$. 

50
### Table E.3b.  
Frequency with Which Teachers Received Student Performance Information

<table>
<thead>
<tr>
<th>Type of Student Performance Information</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Few Times per Month (%)</td>
<td>Approx. Weekly (%)</td>
</tr>
<tr>
<td>Scores on district or state assessments in <strong>mathematics</strong></td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Scores on district or state assessments in <strong>language arts</strong></td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Scores on district or state assessments in <strong>science</strong></td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Scores on district or state assessments in subjects other than mathematics, language arts, or science**</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Information about student performance on specific concepts or skills**</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Assessment data that is built into curriculum software**</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Identification of specific students who need extra assistance**</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Identification of specific students who have achieved mastery**</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>Youth development outcomes (for example, student behavior, attitudes, or motivation)**</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Identification of specific students who are at risk of dropping out or not progressing to the next grade**</td>
<td>12</td>
<td>10</td>
</tr>
</tbody>
</table>

**Notes**:  
* indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01.

Question text: In general, how frequently do you receive the following types of information about the performance of your students? Response options: Never; Once a year; A few times per year; Approximately monthly; A few times per month; Approximately weekly; A few times per week; At least daily. (Percentages for the first four response options are shown in Table E.3a.) ObD $N = 182–183$; ATP $N = 1,002–1,004$. 

51
Table E.4.
Extent to Which Teachers Used Student Achievement or Mastery Data

<table>
<thead>
<tr>
<th>Purpose of Data Use</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I Don’t Do This Without Data</td>
<td>Small Extent</td>
</tr>
<tr>
<td>Tailoring the pace of instruction to individual students’ needs**</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Tailoring the content of instruction to individual students’ needs**</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Tailoring the instructional strategies to individual students’ needs**</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Developing recommendations for tutoring or other educational support services for particular students**</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>Assigning or reassigning students to groups within my class(es)**</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Allowing students to skip units or lessons if they’ve demonstrated mastery of the content in some other way**</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>Allowing students to skip courses or grades if they’ve demonstrated mastery of the content in some other way**</td>
<td>65</td>
<td>3</td>
</tr>
<tr>
<td>Recommending students for extended learning opportunities (for example, extended-day programs, Saturday classes, or an extended school year)**</td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td>Identifying topics requiring more or less emphasis in instruction**</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Identifying areas where I need to strengthen my content knowledge or teaching skills**</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Reflecting on and discussing teaching and learning with other teachers**</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Reflecting on and discussing learning with my students**</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>
**Table 1: Use of Data for Specific Purposes**

<table>
<thead>
<tr>
<th>Purpose of Data Use</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I Don’t Do This</td>
<td>I Do This Without Data</td>
</tr>
<tr>
<td>Providing college/career advice or guidance**</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Revising our school’s strategic goals or educational plan**</td>
<td>39</td>
<td>11</td>
</tr>
</tbody>
</table>

NOTES: * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01. Our national teacher data are based on results from the ATP. Question text: This school year (2017–2018), to what extent have you used student achievement or mastery data for each of the following purposes? (Consider data provided by instructional software, interim assessments or quizzes, unit or end of course tests, state accountability tests, district benchmark or interim tests, the MAP tests and other standardized tests.) If the activity is something that you don’t do (for example, if you never tailor the pace of instruction), please mark I don’t do this. Response options: I don’t do this; I do this but do not use data for it; Used data to a small extent; Used data to a moderate extent; Used data to a large extent. ObD $N = 180–182$; ATP $N = 999–1,000$. 

53
Table E.5.
Teachers’ Assessment of Their Schools’ Data System

<table>
<thead>
<tr>
<th>Statement</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Strongly Disagree</strong></td>
<td><strong>Disagree</strong></td>
</tr>
<tr>
<td>I have access to high-quality data that help me adapt the pace, content,</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>or strategies of instruction to meet students’ needs.**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can use the school’s data system to easily produce the views or reports</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>that I need.*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Our school’s data system provides real-time data that are actionable.**</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>I have the necessary skills and experience to use data to guide my</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>instruction.**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Our school’s data system and assessments enable me to make good</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>decisions about mastery-based progression for individual students (i.e.,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>using data to determine whether students have mastered a set of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>competencies and should move on to new material).**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Our school’s data system and assessments provide adequate information</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>toward specific learning objectives.**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students have access to information from data systems that track their</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>progress on particular tasks, skills, or for the course overall.**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students regularly review data on their own progress using the school’s</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>data system.**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTES: * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01.
Our national teacher data are based on results from the ATP. Question text: Please rate your level of agreement with each of the following statements. Response options: Strongly disagree; Disagree; Agree; Strongly agree. ObD N = 179–181; ATP N = 999–1,000.
### Table E.6.
**Extent to Which Teachers Emphasized Personalized Learning Practices**

<table>
<thead>
<tr>
<th>Practice</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None</td>
<td>Small Extent</td>
</tr>
<tr>
<td>Students have opportunities to choose what instructional materials (such as books or computer software) they use in class.**</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>Students have opportunities to choose what topics they focus on in class.**</td>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>I frequently adapt course content to meet students’ needs by providing additional assignments, resources, and activities for remediation or enrichment.**</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>I clearly present the goal or objective for each assignment.</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>I have adopted strategies that allow students to keep track of their own learning progress.**</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>I require students to demonstrate mastery of a topic before they can move onto a new topic.**</td>
<td>9</td>
<td>28</td>
</tr>
<tr>
<td>Different students work on different topics or skills at the same time.**</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>I give students the chance to work through instructional material at a faster or slower pace than other students in this class.**</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>When students are working on an assignment or activity, they know what the goals of the assignment or activity are.</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Students keep track of their own learning progress using technology (for example, by using an online gradebook or portfolio).**</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Students have opportunities to review or practice new material until they fully understand it.**</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>When students are working independently, I require them to get through a certain amount of material even if they are working at their own pace.**</td>
<td>5</td>
<td>18</td>
</tr>
</tbody>
</table>

55
Table E.6—Continued

<table>
<thead>
<tr>
<th>Practice</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None (%</td>
<td>Small Extent (%)</td>
</tr>
<tr>
<td>I frequently regroup students for instruction to address changing</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>learning needs and interests.**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students are able to access instructional materials both in and outside</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>of the classroom.**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I provide a variety of materials or instructional approaches to</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>accommodate individual needs and interests.**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I connect what students are learning with experiences they have</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>throughout the rest of the school day or outside of school.*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTES: * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01. Our national teacher data are based on results from the ATP. Question text: Teachers take a variety of approaches to personalizing, or customizing, learning opportunities for their students. Please indicate the extent to which you emphasize the following practices related to personalization. Response options: Have not emphasized [none]; Emphasized to a small extent; Emphasized to a moderate extent; Emphasized to a large extent. ObD N = 180–181; ATP N = 998–999.

Table E.7.
Extent to Which Teachers Emphasized Social and Emotional Learning Practices

<table>
<thead>
<tr>
<th>Practice</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None (%</td>
<td>Small Extent (%)</td>
</tr>
<tr>
<td>I connect academic content to themes related to social and emotional</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>learning (e.g., by discussing resilience in the context of a character in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a book or emphasizing perseverance or time management in solving math</td>
<td></td>
<td></td>
</tr>
<tr>
<td>problems).**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I greet each student by name when they come to class.**</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>I actively establish one-on-one relationships with each of my students.**</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I give students opportunities to display their nonacademic talents in</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>class.**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTES: * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01. Our national teacher data are based on results from the ATP. Question text: Teachers take a variety of approaches to supporting students’ social and emotional growth and developing relationships with students. Please indicate the extent to which you emphasize the following practices related to social and emotional learning. Response options: Have not emphasized [none]; Emphasized to a small extent; Emphasized to a moderate extent; Emphasized to a large extent. ObD N = 180–181; ATP N = 998.
### Table E.8.
**Extent to Which Teachers Addressed Social and Emotional Learning Topics in Their Instruction**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None</td>
<td>Small Extent</td>
</tr>
<tr>
<td>Collaborating with other students**</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Understanding and managing emotions**</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Handling stress</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Persisting through challenging tasks**</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Making responsible decisions*</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Establishing and maintaining positive relationships*</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Feeling and showing empathy for others**</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Learning mindsets (e.g., growth mindset, sense of purpose and belonging)**</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Communicating their thoughts and emotions**</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Developing a sense of identity</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

**NOTE:** * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01.

Our national teacher data are based on results from the ATP. Question text: Please indicate to what extent you address each of the following topics in your instruction. Response options: Not addressed [none]; To a small extent; To a moderate extent; To a large extent. ObD N = 178–180; ATP N = 998.

### Table E.9.
**Teachers’ Assessment of Their Curriculum Materials**

<table>
<thead>
<tr>
<th>I have adequate access to curriculum materials (whether technology- or non–technology-based) that . . .</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly Disagree</td>
<td>Disagree</td>
</tr>
<tr>
<td>are of high quality.</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>address the learning needs of all of my students.</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>contribute to my efforts to promote college and career readiness.</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>are accessible to students outside of the school building and outside of the regular school day.</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>support mastery-based instruction (i.e., help me to determine whether students have mastered a set of competencies and should move on to new material)*</td>
<td>6</td>
<td>21</td>
</tr>
</tbody>
</table>

**NOTE:** * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01.

Our national teacher data are based on results from the ATP. Question text: Please rate your level of agreement with the following statements. Response options: Strongly disagree; Disagree; Agree; Strongly agree. ObD N = 180; ATP N = 998.
### Table E.10.
Proportion of School- or District-Provided Curriculum and Instructional Materials That ObD Teachers Used

<table>
<thead>
<tr>
<th>Response Option</th>
<th>Percentage of ObD Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>None; I don’t receive any materials from my school or district</td>
<td>16</td>
</tr>
<tr>
<td>Fewer than half</td>
<td>31</td>
</tr>
<tr>
<td>About half</td>
<td>22</td>
</tr>
<tr>
<td>More than half</td>
<td>21</td>
</tr>
<tr>
<td>All of the materials</td>
<td>10</td>
</tr>
</tbody>
</table>

**NOTES:** Question text: Approximately what proportion of the curriculum and instructional materials you use (including textbooks, software, assignments, and other materials needed to teach a lesson) was provided to you by your school or district? Percentages may not sum to 100 percent because of rounding. $N = 180$.

### Table E.11.
Proportion of Searched-for Resources That ObD Teachers Used to Supplement Their Curriculum and Instructional Materials

<table>
<thead>
<tr>
<th>Response Option</th>
<th>Percentage of ObD Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>None; I don’t search for supplemental materials</td>
<td>2</td>
</tr>
<tr>
<td>Fewer than half</td>
<td>29</td>
</tr>
<tr>
<td>About half</td>
<td>32</td>
</tr>
<tr>
<td>More than half</td>
<td>27</td>
</tr>
<tr>
<td>All of the materials</td>
<td>11</td>
</tr>
</tbody>
</table>

**NOTES:** Question text: Approximately what proportion of the curriculum and instructional materials you use consists of existing material that you searched for (e.g., from websites) to supplement the curriculum that was provided to you? Percentages may not sum to 100 percent because of rounding. $N = 178$.

### Table E.12.
Proportion of Original Content That ObD Teachers Used to Supplement Their Curriculum and Instructional Materials

<table>
<thead>
<tr>
<th>Response Option</th>
<th>Percentage of ObD Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>None; I don’t create original material</td>
<td>1</td>
</tr>
<tr>
<td>Fewer than half</td>
<td>27</td>
</tr>
<tr>
<td>About half</td>
<td>19</td>
</tr>
<tr>
<td>More than half</td>
<td>39</td>
</tr>
<tr>
<td>All of the materials</td>
<td>14</td>
</tr>
</tbody>
</table>

**NOTES:** Question text: Approximately what proportion of the curriculum and instructional materials you use consists of original material you created individually or with your school colleagues to supplement the curriculum that was provided to you? Percentages may not sum to 100 percent because of rounding. $N = 178$. 
### Table E.13.
**Time Teachers Spent in a Typical Week Developing or Selecting Instructional Materials**

<table>
<thead>
<tr>
<th>Response Option</th>
<th>Percentage of ObD Teachers</th>
<th>Percentage of Teachers Nationally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 1 hour</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>1–2 hours</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>3–4 hours</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td>5–6 hours</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>7–8 hours</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>9–10 hours</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>More than 10 hours</td>
<td>12</td>
<td>6</td>
</tr>
</tbody>
</table>

NOTES: Average amount of time spent by ObD teachers in a typical week = 4.4 hours; average amount of time spent by teachers nationally in a typical week = 3.6 hours. ** indicates statistically significant difference in average amount of time spent at alpha < 0.01. Our national teacher data are based on results from the ATP. Question text: Roughly how many hours do you spend developing and/or selecting instructional materials in a typical week? Percentages may not sum to 100 percent because of rounding. ObD N = 180; ATP N = 996.

### Table E.14.
**Frequency with Which ObD Teachers Gave Students Tasks to Monitor or Assess Their Learning**

<table>
<thead>
<tr>
<th>Task</th>
<th>Percentage of ObD Teachers Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never or Hardly Ever</td>
</tr>
<tr>
<td>Practice for building fluency of procedural skills (e.g., often via worksheets, quizzes, or drill-and-practice software)*</td>
<td>9</td>
</tr>
<tr>
<td>Basic comprehension or recall of facts (e.g., short answer questions, summary)</td>
<td>6</td>
</tr>
<tr>
<td>Application of knowledge or skills with real-world connections</td>
<td>1</td>
</tr>
<tr>
<td>Culminating projects and performance tasks that provide opportunities for deeper learning, application of skills and content, and multiple formats for demonstrating student mastery</td>
<td>1</td>
</tr>
</tbody>
</table>

NOTES: * indicates statistically significant difference in item means at alpha < 0.05. Question text: How often do you give each of the following types of tasks to students to monitor or assess their learning? Response options: Never or hardly ever; Sometimes (1–3 times per month); Often (1–3 times per week); Daily or almost daily. Percentages may not sum to 100 percent because of rounding. N = 180.
Table E.15.
Frequency of Teacher Practices in Response to Poor Student Performance on a Task

<table>
<thead>
<tr>
<th>Practice</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never or Hardly Ever</td>
<td>Sometimes</td>
</tr>
<tr>
<td>I moved students along to learn other topics and skills or to do other tasks.</td>
<td>13</td>
<td>41</td>
</tr>
<tr>
<td>I gave students a task that is similar in complexity to work on, to reinforce the content or skills.</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>I gave students a simpler task that covers the same content or skill.**</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>I retaught or reviewed the content or skills.*</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>I reviewed the student work, and the way it was assessed, with students.</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>I worked step-by-step with students as they revised the task or worked on similar tasks.**</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>I provided students with samples or models of finished work for the task they were assigned.**</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>I placed students in groups to receive support with learning the content or skills.**</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>I pointed students to materials or resources to practice or review on their own.**</td>
<td>6</td>
<td>25</td>
</tr>
</tbody>
</table>

**NOTES:** *, indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01.

Our national teacher data are based on results from the ATP. Question text: How often did each of the following practices occur when students did not perform well on a task (e.g., quizzes, short-answer questions, extended projects, etc.)? Response options: Never or hardly ever; Sometimes; Often; Always or almost always. ObD \( N = 177–179 \); ATP \( N = 994–996 \).
### Table E.16.
**Extent to Which Descriptions of Personalized Instruction Resembled Teachers’ Practices**

<table>
<thead>
<tr>
<th>Description</th>
<th>Percentage of ObD Teachers Who Responded</th>
<th>Percentage of Teachers Nationally Who Responded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Never or Hardly Ever</td>
<td>Sometimes</td>
</tr>
<tr>
<td>When starting on a new topic or competency, I first identify students’ prior knowledge and skills with a diagnostic assessment or task.*</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>Students attempt a task that is assessed for mastery when I believe they have a good chance at success on it (when they have demonstrated the knowledge/skills through other activities or tasks already).</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Students receive the task that will be assessed for mastery at the beginning of the unit, and they work on it in pieces throughout the unit.**</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Tasks that are assessed for mastery closely resemble tasks that students have already seen or experienced.</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Tasks that are assessed for mastery require students to apply knowledge and skills they have learned to a new problem or context.*</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Tasks that are assessed for mastery of a given competency differ in difficulty, depending on the student's ability level.**</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>Mastery is assessed as what students can accomplish mostly independently (i.e., without extensive peer support or scaffolds).*</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Students have multiple opportunities, throughout a unit or throughout the year, to demonstrate mastery of certain content and skills.**</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Students can make multiple attempts at a given task that counts toward mastery.**</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Students are considered to have achieved mastery when they have completed the expected number of tasks.</td>
<td>14</td>
<td>31</td>
</tr>
<tr>
<td>Student work counts toward mastery only when it reaches an adequate performance level.</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>Students are considered to have achieved mastery only when they have demonstrated the knowledge or skills consistently (i.e., multiple times, on different tasks, in multiple ways).</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>

**NOTES:** * indicates statistically significant difference in item means at alpha < 0.05; ** indicates statistically significant difference in item means at alpha < 0.01. Our national teacher data are based on results from the ATP. Question text: Please rate the extent to which each of the following descriptions resemble your instructional practices. Response options: Never or hardly ever; Sometimes; Often; Always or almost always. ObD N = 163–166; ATP N = 810–811.
References


https://www.rand.org/pubs/tools/TL136z1.html

https://www.rand.org/pubs/research_reports/RR3104.html


WWC—See What Works Clearinghouse.