Building and Sustaining Innovative High Schools

Findings from the Opportunity by Design Study

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

RAND CORPORATION

For more information on this publication, visit www.rand.org/t/RRA322-3

Published by the RAND Corporation, Santa Monica, Calif.

Photo credits: Cover, SolStock/Getty Images; page 3, Monkey Business/Adobe Stock; page 11, Monkey Business/Adobe Stock; page 18, FatCamera/Getty Images; page 19, SDI Productions/Getty Images; page 21, FatCamera/Getty Images; page 32, kali9/Getty Images; page 33, sorapop/Adobe Stock; page 35, Monkey Business/Adobe Stock; page 40, goodluz/Adobe Stock; page 50, kali9/Getty Images; page 55, Monkey Business/Adobe Stock.

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 final report summarizes the methods and findings from this five-year study and is intended to provide lessons and implementation guidance for the field.

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.

Table of Contents

Figures

Tables

Summary

Study Context

- This report focuses on estimating the effects of the Opportunity by Design (ObD) initiative on student academic achievement, attendance, and suspension, and on describing implementation of the three key design principles in the final year of the initiative.
- The ObD initiative provided funding and technical assistance to 16 small schools of choice across seven large, urban districts over four years.
- The three key design principles focused on innovative instructional practices: personalization of learning, mastery-based instruction, and positive youth development (PYD).

Key Takeaways

- This research did not find much evidence that ObD's principles-based design process and supports led to more-effective schools, but limitations of the study design and available data may not have adequately captured ObD's effects.
- Mastery-based learning was conceptualized as deep knowledge of content and skills but in practice entailed offering students multiple attempts to demonstrate mastery.
- Personalization entailed accommodating students' interests and did not typically involve such practices as providing students with extensive choices in content or materials.
- PYD emphasized skills to support student academic achievement and positive behavior, but these skills were not explicitly taught or assessed in most schools.
- A technical assistance provider—Springpoint—played a unique role by providing ObD school leaders with tailored support, but the time limit on Springpoint support posed a challenge to sustainability in schools and districts that experienced leadership turnover.
- Enablers of implementation in ObD schools included alignment of school and district grading policies in some districts and autonomy from district curriculum and professional development (PD) requirements.
- Barriers to implementation in ObD schools included limited district support for selecting or developing curriculum and PD materials and inflexible district policies.
- Compared with a national sample of high school teachers, ObD teachers reported more extensive use of practices related to the three key design principles of mastery-based learning, personalization, and PYD.

Introduction

High schools across the United States are responsible for ensuring that their students develop the skills, knowledge, and dispositions that will prepare them to thrive in college and careers and in their lives. High schools are not only responsible for teaching students the academic content they will need in their lives after high school, but they are also responsible for helping students develop a variety of inter- and intrapersonal competencies that are necessary for postsecondary success (Aspen Institute National Commission on Social, Emotional, and Academic Development, 2019).

However, high schools face substantial challenges and are working to address students' academic, social, and emotional learning needs in the face of large disparities in the readiness of incoming students (National Assessment of Educational Progress, undated). Many high schools in high-poverty areas struggle to prepare students for graduation within four years (Balfanz, 2009). Although the ObD initiative concluded prior to the coronavirus disease 2019 pandemic, pandemic-related school building closures and disruption to instruction have created new hurdles for high schools to overcome as they work to support equitable access to high-quality supports for learning (Hamilton et al., 2020). Educators therefore need strategies to promote students' academic, social, and emotional development in a way that supports equity of opportunities and outcomes (Aspen Institute National Commission on Social, Emotional, and Academic Development, 2019). The findings from this analysis, which discuss instructional practices that may be well suited to support students working independently, may be particularly relevant.

This report focuses on estimating the effects of the ObD initiative on student academic achievement, attendance, and suspension and on describing implementation of the three key design principles in the final year of the initiative.

The ObD Initiative

The Carnegie Corporation of New York (CCNY) launched the ObD initiative in 2013 to support the design and creation of a network of small high schools of choice in large, urban districts in the United States. The ultimate goal of the initiative was to help students graduate from high school within four years with the academic and social and emotional skills needed for postsecondary success. The ObD initiative was based on the premise that promising high school reform strategies need to be integrated into a comprehensive school design and accompanied by appropriate, sustained levels of financial, policy, and implementation supports (Hamilton and Mackinnon, 2013). Thus, the designs of the ObD schools focused on ten design principles, developed by CCNY, that addressed key aspects of school culture, instruction, and operations (see Box S.1 and Appendix C).

The initiative was designed to provide funding and technical support to the schools during a design year prior to opening the school and during the first two years of implementation (although many schools opted to receive technical support beyond the second year of implementation), with the expectation that the schools would design their models to be self-sustaining on public funds. CCNY selected the ObD districts from the pool of applicants based on the presence of a series of enabling conditions, such as district leader support, financial stability, standards for college and career readiness, and support from community partners (Hamilton and Mackinnon, 2013).

As schools of choice, the ObD schools were ones that students opted to attend and that were open to students throughout the districts. There were a total of 16 ObD schools across seven large, urban districts: New York, N.Y.; Denver, Colo.; Providence, R.I.; Prince George's County, Md.; Cleveland, Ohio; Philadelphia, Pa.; and a small charter school organization in Brooklyn, N.Y. Fifteen of the 16 ObD schools were located in traditional school districts, and one was a charter school. Although we describe the participating districts as "ObD districts" in this report and discuss district-level effects, not all high schools in these districts were ObD schools. The number of ObD schools in a given district ranged from one to seven.

The first ObD schools opened in fall 2014, and the last opened in fall 2017. All but one ObD school opened with ninth grade only and added one grade level each year; one school opened with ninth and tenth grades. The ObD schools were designed to be smaller than large, comprehensive high schools—about 400 students per school at full enrollment.

The ObD initiative was different from many other high school reform initiatives in three ways. First, the school models were expected to incorporate the ten design principles holistically and to continuously refine the model over time. The design principles were intended to serve as guidelines and described desired instructional practices and school conditions rather than a well-defined intervention or model. Second, ObD school principals were supported by a dedicated, external technical assistance organization: Springpoint. Third, the design principles encouraged the ObD schools to adopt innovative instructional practices that had the potential to help underprepared students catch up and graduate in four years with the academic and social and emotional skills necessary for postsecondary success.

Following research that suggests that comprehensive school redesign efforts are most successful when embedded in broader systems of support (e.g., Bloom, Thompson, and Unterman, 2010), CCNY considered the presence of a series of enabling conditions (e.g., district leader support, standards for college and career readiness) and support from community partners (Hamilton and Mackinnon, 2013) when selecting grantees. CCNY also established Springpoint: Partners in School Design, to provide technical assistance to ObD school and district leaders in designing, implementing, and sustaining innovative high schools. The ObD schools were also supported by other external partners (e.g., technical assistance or curriculum providers), intermediary organizations, and the districts in which they were located.

This report emphasizes the three design principles that focus on adopting innovative instructional practices: *prioritizing mastery*, *personalizing learning*, and *PYD*, which we refer to as *key design principles*. We chose this focus because CCNY highlighted these principles as foundational to the schools' models and instructional approaches based on schools' experiences in the first two years of the initiative. We also touch on several of the design principles more closely related to school operations (e.g., hiring, data use) because they support high-quality instructional practices.

Although there is not yet widespread consensus about how to define each of these approaches, mastery-based practices tend to include clear learning targets that support deep understanding and application of knowledge by allowing students to move through content at their own pace and attempt a task or assignment multiple times (Gross and DeArmond,

BOX S.1 **ObD Design Principles**

A high-performing high school . . .

1. has a clear mission and coherent culture
2. prioritizes mastery of rigorous standards aligned to college and career readiness
3. personalizes student learning to meet student needs
4. maintains an effective human capital strategy aligned with school model and priorities
5. develops and deploys collective strengths
6. remains porous and connected
7. integrates PYD to optimize student engagement and effort
8. empowers and supports students through key transitions into and beyond high school
9. manages school operations efficiently and effectively
10. continuously improves its operations and model (CCNY, 2014).

BOX S.2

Key ObD Design Principles Addressed in This Report

- **Prioritizing mastery:** Students demonstrate deep understanding of clearly defined, rigorous competencies.
- **Personalizing learning:** Student learning experiences are tailored to individual learning needs and interests.
- **Positive youth development:** Students have a voice in their learning and access to experiences and relationships that help them develop the skills and mindsets to succeed (CCNY, 2017).

2018; Pane et al., 2017; Steele et al., 2014; Sturgis, 2012). Personalizing learning often involves tailoring learning experiences to students' individual learning needs and interests. Personalized and mastery-based approaches frequently intersect with PYD. All three include efforts to provide students with more control over their learning and access to experiences that will support development of social and emotional and academic skills (Gross, Tuchman, and Patrick, 2018). PYD is characterized by a focus on helping students develop their social and emotional skills in addition to academic skills and by a school model that prioritizes supportive relationships and an engaging and supportive environment (Taylor et al., 2017). Box S.2 provides the definitions that CCNY used in the context of ObD.

Although there is some evidence that personalized, mastery-based, and PYD approaches can improve student academic, social, and emotional outcomes in some contexts (e.g., Pane et al., 2017; Deming, 2017), they can be challenging to implement and risk resulting in inequitable opportunities for students (Gross and DeArmond, 2018; Pane et al., 2017; Steele et al., 2014; Marsh et al., 2018; Schwartz et al., 2020). For example, many schools lack high-quality curriculum materials that are appropriate to personalized and mastery-based learning environments and that integrate social and emotional learning skills, and teacher-developed materials can vary greatly in quality. In many schools, the criteria for mastery of the material can vary by teacher and are sometimes defined as completion of tasks rather than demonstration of deep understanding of the skills and content.

Research Approach

We began a five-year study of the ObD initiative in June 2014. We monitored implementation over four years (the 2014–2015 school year through the 2017–2018 school year) and collected student outcomes for those years and one additional year for the 16 ObD schools. The study addressed four RQs:

1. How did student outcomes in ObD schools compare with outcomes of similar students in other schools in the same districts or across the United States?
2. How did teachers and other school staff implement the key design principles, and what factors might have facilitated or hindered implementation?
3. What system-level conditions supported or hindered implementation?
4. How did ObD teachers' practices and perceptions of implementation enablers and challenges compare with those of a nationally representative sample of high school teachers?

This report builds on a prior report (Steiner et al., 2017) that focused on implementation of the ten design principles after the first two years of the initiative and presented interim findings related to

RQs 1 and 2. The prior report included surveys of and interviews or focus groups with school leaders, teachers, and students, along with observations of classrooms and analysis of instructional artifacts in the ten ObD schools that were open at the time. The report did not present findings related to student outcomes or compare ObD teachers' practices and perceptions with those of high school teachers nationally (RQs 1 and 4). This report fills those gaps. Box S.3 summarizes key findings from the earlier report.

The analyses that we present in this report focus on data collected in the fourth year of the study (school year 2017–2018). We discuss the impact of attending an ObD school on academic, behavioral, and college readiness outcomes. We also describe implementation of the three key design principles, examine local contextual conditions in the ObD districts, and situate ObD teachers' reports of their instructional practices and supports in a national context.

Academic outcomes are measured by mathematics and English language arts (ELA) assessments; behavioral outcomes are measured by attendance and suspension rates; and college readiness outcomes are measured by PSAT and SAT scores, credit accumulation, Advanced Placement® tests passed, and graduation rates. Each analysis compared the ObD students with a comparison group of non-OBD students, matched on prior achievement and demographic characteristics. The comparison group consisted of either students in other district schools, or nationwide takers of an assessment also taken by ObD students. We estimated effects of ObD within each district and combined those to obtain an estimate of the initiative's overall effect. The implementation findings are based on surveys of teachers and students; interviews with CCNY, Springpoint, and district staff; and school visits in which we interviewed school leaders and teachers, conducted focus groups with students, observed classrooms, and collected instructional artifacts (e.g., culminating assignments). The discussion of the national context is based on RAND's American Teacher Panel (ATP), a nationally representative survey of teachers.

Although the ObD schools were expected to design their models around a common set of ten design principles, each school's model was unique and designed for its local context. This diversity of school models makes it difficult to interpret variation in results across schools. Even though this report presents the

BOX S.3

ObD After Two Years of Implementation

After two years of implementation of these complex reforms, the ten ObD schools demonstrated progress in defining and implementing their school models but also faced numerous challenges. Students in all schools reported positive perceptions of school culture, and, in general, teachers believed that many aspects of implementation, including their PD experiences, had improved from the first year of the initiative to the second. District and school leaders reported that the coaching and technical assistance provided by Springpoint was a key implementation support and helped spread best practices and align district and school systems in a way that could facilitate the spread of innovation.

However, efforts to personalize instruction and implement mastery-based instructional approaches were often inconsistent and limited by varying access to data, external pressure to advance students at a certain pace, and the significant time required to create appropriate instructional materials. Integration of PYD with mastery-based and personalized instructional approaches was nascent, and, in general, school staff described PYD as separate from academics. Human capital challenges, such as persistent teacher vacancies, limited availability of high-quality instructional materials, and inconsistent teacher PD exacerbated these challenges. Although teachers reportedly enjoyed the autonomy to create their own materials, they struggled to find the time to do so and wished for more district support to guide their choices. Most schools had yet to develop clear systems for data-driven improvement and instead responded to feedback and addressed issues as they arose.

final findings from this research, we urge readers to keep in mind that—in the fourth year of this study, on which our analysis focuses—four of the 16 ObD schools (or 25 percent) were in their first year of implementation while others had been in operation for two, three, or four years. The outcomes study relies on relatively sparse student outcomes data and methods that are vulnerable to selection bias. Apparent differences between ObD and comparison samples might not have been caused by ObD. Furthermore, in limiting our focus to the three key design principles, we could have overlooked some aspects of implementation that may have influenced effects. The data on implementation are limited by their self-reported nature and relatively small sample sizes, which are themselves limited by the size of the schools.

Nevertheless, the implementation data provide detailed examples of the design principles in the ObD schools and rich descriptions of implementation facilitators and challenges. This report fills two key gaps in the prior report (Steiner et al., 2017) by including ObD's effects on student outcomes and data from a national comparison group of high school teachers. We hope the discussion in this report will be of interest to schools that have adopted, or are considering adopting, mastery-based, personalized, or PYD practices; the funder; and educators and policymakers who are implementing, or considering implementing, similar reforms.

ObD Impacts on Student Outcomes

Focusing on estimated effects of the ObD initiative overall, the study found no statistically significant evidence of improved student performance as captured by 12 measures of academic, behavioral, and college readiness outcomes. We found statistically significant negative effects for the initiative on credit accumulation and SAT scores. The credit accumulation result appears to have been driven by a strong negative effect in one district. The SAT estimates were more uniformly negative across districts.

Estimates of district-specific academic effects tended to be negative for state assessments of standards attainment and positive for achievement growth measured by MAP. Such a pattern could emerge from implementation of personalized and mastery-based learning if ObD schools placed relatively greater emphasis on student growth, even if off grade level, and less emphasis on attainment of grade-level standards—or if MAP was better able to capture the concepts and skills that students improved on than were state assessments.

Estimated district-specific effects on attendance and suspension were mixed, with a similar number of positive and negative estimates. However, there is suggestive evidence that attendance and suspension may be key antecedents of improved credit accumulation and graduation. Positive effects appeared in credit accumulation or graduation in only three districts, the same three districts that showed positive effects on at least one of these behavioral outcomes. Otherwise, effects on college readiness outcomes were mixed, with negative effects more prevalent than positive.

Implementation of the Key Design Principles

In most schools, mastery-based learning was conceptualized as deep knowledge of content and skills. In practice, it entailed offering students multiple attempts to demonstrate mastery. Staff and students at half of the ObD schools described mastery as assessing whether a student deeply understood a topic and could apply the knowledge in novel contexts. In the other schools, staff and students described mastery in terms of completion of tasks or assignments. Regardless of how the school conceptualized mastery, the most common mastery-based practice was providing students with multiple attempts to demonstrate mastery. Teachers, leaders, and students described three broad strategies: providing students with many opportunities to complete and turn in assignments; providing students with multiple

opportunities to show improvement on assignments; and structuring assignments so that the same competency was taught in multiple units throughout the year—a technique known as "spiraling."

Personalization entailed accommodating students' interests but did not typically involve other personalization practices, such as providing students with extensive choices in content or materials. In all schools, personalizing learning entailed tailoring instructional practices, materials, and topics to accommodate students' interests. Other personalization practices, such as offering extensive choices of content, topics, or instructional materials to students, were uncommon in most schools. Teachers reported that most assignments were not personalized; typically, all students received the same assignment. Teachers described two strategies for personalizing assignments—offering the assignment at different levels of difficulty and providing a limited set of choices of topic or mode of presentation (e.g., oral presentation, written report). Few teachers reported using both strategies. Teachers found using these strategies to be time-consuming and therefore challenging to implement.

PYD emphasized skills to support student academic achievement and positive behavior, but these skills were not explicitly taught or assessed in most schools. In most schools, PYD encompassed structures and systems to help students build social and emotional skills that could help improve their academic achievement and positive school culture. In most schools, PYD approaches involved connecting academic content to development of social and emotional skills and connecting mastery-based learning and personalization to PYD. In about half of the ObD schools, academic lessons incorporated social and emotional competencies that were assessed using a rubric. But, in most schools, teachers reported teaching key social and emotional learning (SEL) skills implicitly (e.g., by modeling the skill or behavior). Few teachers described providing students with specific guidance about how to enact key SEL skills or assessing such skills as part of the course. The extent to which schools were successfully teaching SEL skills has implications for how well prepared students will be to achieve their postsecondary goals.

Local Contextual Conditions

Springpoint played a unique role in supporting ObD principals, but leadership turnover combined with the time-limited nature of Springpoint's assistance reportedly hindered sustainability in some schools. The external support provided by Springpoint is not common to most comprehensive school-reform efforts. CCNY designed the initiative to provide schools with three years of Springpoint support, which generally focused on building the capacity of the founding principal and teachers and select district staff. In some cases, Springpoint did not have the resources to provide the same level of support to new principals or district staff. Principals in several districts expressed frustration that, due to turnover, the leaders in their district had a limited understanding of the ObD design principles and lacked Springpoint support to help them become more familiar. According to these principals, the combination of these conditions often limited the relevance and usefulness of district supports. In other schools, however, principals were able to repurpose funds to cover the costs of additional Springpoint support, so the challenges associated with the time limit were not universal.

Enablers of implementation included district support for innovation and autonomy from district curriculum and district PD requirements. Principals described varying levels of district support for ObD schools, and those who described stronger district-level support were members of strong district-wide networks of innovative schools. All the ObD schools had the autonomy to adopt or develop curricula that were aligned to their school model, and staff at most schools used a blend of materials, combining those they developed or sourced themselves with others provided by external partners or the district. Staff at most ObD schools also had the flexibility to develop their own PD or work with external providers. This autonomy and flexibility were key enablers of implementation because

they allowed the ObD schools to select materials and PD that were tailored to their school models.

Implementation challenges included misalignment between school and district grading systems, limited district support for selecting or developing curriculum and PD materials, and inflexible district policies. ObD school and district grading policies were often misaligned, which hindered successful implementation of mastery at the school level and created extra work and confusion for students and families. The challenge created by misalignment between mastery-based grading and external systems is not unique to ObD and has come up in other studies of schools that use personalized and mastery-based instructional approaches. Staff in most ObD schools reported that they had the autonomy to select or design curriculum and PD materials, but most wished for some help from their districts. Principals perceived district policies related to hiring teachers and following district scheduling requirements to hamper implementation. Some districts granted the ObD schools a temporary reprieve from such policies in the early years of the initiative. The principals who benefited found the reprieve helpful in the first year but reported that the policies emerged as constraints in subsequent years.

ObD in a National Context

Compared with a national sample of high school teachers, ObD teachers reported more-extensive use of practices related to the three key design principles. We compared survey data from a nationally representative sample of teachers collected by RAND's ATP with data collected from ObD teachers. ObD teachers reported emphasizing mastery-based, personalized, and PYD practices in their instruction to a greater extent than teachers nationally. ObD teachers also reported more extensive use of student data to inform their instruction and more-positive opinions about data access and school data system quality. Although these summary findings mask some variability in practices and supports within both the ObD and the ATP groups, on average, the differences were consistent and sometimes large. We cannot make any claims regarding the reasons for these differences, but it is possible that they stem, at least in part, from the resources and guidance provided to ObD teachers as part of the initiative.

Implications

Many of the component strategies adopted by ObD schools are popular and are being adopted more broadly but are difficult to implement well. To be successful, school staff and students may need different resources and supports than are available at other high schools. We discuss lessons drawn from ObD implementation that may help other schools that use mastery-based, personalized, or PYD approaches manage similar challenges and implement similar reforms in a high-quality way. The implications from this study can also be informative for researchers and others who are interested in understanding how high schools enact innovative practices and how those practices influence students' learning.

Although teachers in ObD schools reported more extensive use of personalized, mastery-based, and PYD instructional approaches, our analyses of student outcomes did not show positive effects for the initiative overall. In their survey responses, ObD teachers reported emphasizing mastery-based, personalized, and PYD practices in their instruction and using student data to inform their instruction to a greater extent than teachers nationally. However, we did not find any evidence that these changes in instructional practice were linked to changes in student outcomes. We were not able to quantitatively investigate the relationship between implementation and outcomes, but our interview data suggest that one possible explanation for the lack of improvement in student outcomes could be that the ObD schools were still working to implement the key design principles in a deep and integrated way. Staff in most schools reported implementing only one or

two of the instructional practices that characterized each of the three key design principles. In addition, teachers in most schools reported that they were not explicitly teaching or assessing the social and emotional skills that would help students succeed in personalized and mastery-based environments. The key design principles are complex, and it is not surprising that staff in most schools focused on a few aspects of the design principles, but it is possible that this relatively limited implementation was sufficient to be different from practices reported by high school teachers nationally—but not sufficient to have a measurable effect on student outcomes.

Mastery-based instruction requires data that typically are not available in traditional high school classrooms. Mastery-based approaches are a marked departure from how most high schools typically assess and promote students. Teachers in the ObD schools used multiple sources of data—including conversations with students and student achievement data—to inform instructional decisions. ObD teachers also reportedly used such data—and implemented mastery-based practices—to a greater extent than high school teachers nationally. These findings suggest that teachers in mastery-based schools need multiple sources of high-quality, accessible data about student mastery. Principals and district staff should work with teachers in mastery-based schools to ensure that they can access the data they need easily and frequently and ensure that assignments and assessments provide detailed information about student mastery.

District leaders, external support providers, principals, and teachers should consider working together to select and/or develop high-quality curricula suitable to mastery-based and personalized learning environments. A high level of autonomy from district curriculum requirements allowed the ObD schools to select or develop curricula consistent with mastery and personalization but was not often accompanied by district support to identify suitable materials and assess their quality. Many teachers reported developing at least a portion of their curriculum materials and wished for more help from experts to do so. The time teachers spent on these tasks reportedly left them little time to focus on other things, such as personalizing lessons for all students. Teachers also worried that the materials they developed were not challenging enough to prepare students for college and the workforce. Districts and school principals should consider working with external support providers, such as Springpoint, and teachers to provide some resources for curriculum development. Expert external organizations similar to Springpoint could play a valuable role as convener and facilitator or recommend suitable materials based on their experience.

Such resources as adaptable curriculum materials and PD could help teachers address the needs of students who struggle to achieve mastery. In the ObD schools, where most students perform well below grade level, teachers must be skilled at using information about students' mastery of concepts and skills to help those who struggle to achieve mastery. Although most ObD teachers reported adjusting the supports they provided depending on the student, they used a limited set of strategies and reported that doing so was time-consuming. Thus, teachers in mastery-based schools need a number of resources—such as easily adapted curriculum materials and assistance from expert organizations, such as Springpoint—to help them support students who did not achieve mastery.

PYD and related social and emotional competencies may help students succeed in mastery-based systems. Most ObD schools initially did not emphasize PYD and related social and emotional competencies in a way that was integrated with academic instruction. In the final year of the initiative, many ObD teachers reportedly emphasized social and emotional competencies in their instruction to a greater extent than high school teachers nationally. The extent to which schools focus on skills like communication, critical thinking, and collaboration can have implications for how well prepared students will be to achieve their academic and postsecondary goals. Mastery-based schools should consider including competencies for SEL skills in mastery frameworks, providing teachers with research-based curricula and assessments aligned to those competencies, and targeting training to help teachers implement them. Future schools could also consider ways to help students adjust to the expectations of a mastery-based environment.

Schools and districts should develop strategies to mitigate the negative effects of principal turnover. Many ObD schools experienced a change in principal leadership during the four years of the study, and in most of these schools it was perceived as a barrier to successful implementation of the design principles. Districts, funders, and external support

providers could help ease principal transitions by planning for changes in school leadership over the course of a reform by working with principals to develop school operations manuals and providing support targeted to new principals. And, when changes do occur, new principals could develop clear systems and channels for communicating changes to the school design or policies.

Support providers for schools that are engaged in complex reforms should bring a continuous improvement lens. CCNY expected that the ObD schools' models and support needs would evolve over time. Springpoint also evolved over time to respond to schools' changing needs. Support providers and developers of schools engaged in complex reforms should communicate their expectation that schools will experiment and devise ways to collect and disseminate lessons learned. They should also be ready to adapt and help schools engaged in complex reforms develop a continuous improvement process and strategies to communicate changes clearly and consistently.

Complex reforms require a dedicated, aligned system of supports that includes district staff, principals, and teachers. Support from an external organization, such as Springpoint, is not a feature common to most comprehensive school-reform efforts. Nearly all ObD school leaders reported that they valued the expert, tailored support Springpoint provided and said it supplemented the supports they received from their districts. School leaders who were supported by intermediary organizations were similarly positive about those supports, but many wished that teachers could be consistently included. Districts, funders, and external support providers should consider that PD needs may increase rather than decrease over time and work together to ensure that staff at all levels receive continued expert implementation support. An aligned system of continuous support that includes district leaders, school leaders, and teachers could benefit schools engaged in complex reforms.

As in most studies of broad high school initiatives, measurement limitations may have hampered our ability to capture the full breadth of potential ObD effects. Although the study measured a broad set of outcomes, academic measures were limited to mathematics and ELA assessments that do not reflect the breadth of subject matter that high schools cover. Behavioral measures were limited to attendance and suspension rates, which may not be very sensitive to the full range of PYD skills that ObD schools sought to impart; college readiness measures were confined to predictive high school outcomes rather than direct measures of college admission, persistence, or graduation. Ideally, researchers would need to follow students into their postsecondary lives to document post–high school performance. When this type of long-term study is infeasible, researchers should explore opportunities to incorporate a wider variety of measures (e.g., grade-point average). These measures were not available in the present research owing to lack of consistent metrics across participating districts. This limitation points to the need for more common, standardized indicators of postsecondary readiness and success to enable educators and researchers to generate evidence-based guidance for high schools.

Conclusion

This study did not find much evidence that ObD's principles-based design process and supports led to schools that produced strong academic outcomes for students, but it is possible that the limitations of the research may have hindered our ability to capture the initiative's effects. It is important to note that the ObD schools faced challenges that are commonly faced by many schools, and the strategies undertaken by the ObD schools are popular and are likely to be adopted by other schools. Thus, the lessons drawn from ObD implementation are valuable for helping other schools manage similar challenges and implement similar reforms in a high-quality way. The implementation findings from this report can also be informative for researchers and others who are interested in understanding how high schools enact innovative practices and how those practices influence students' learning.

Abbreviations

AP	Advanced Placement
ATP	American Teacher Panel
CASEL	Collaborative for Academic, Social, and Emotional Learning
CCNY	Carnegie Corporation of New York
ELA	English language arts
ObD	Opportunity by Design
PD	professional development
PL	personalized learning
PYD	positive youth development
RQ	research question
SEL	social and emotional learning

CHAPTER ONE

Introduction

In this report, we examine the implementation and outcomes of an innovative school initiative—Opportunity by Design (ObD)—to improve teaching and learning at the high school level. This report focuses on estimating the effects of the ObD initiative on student academic achievement, attendance, and suspension, and on describing implementation in the final year of the four-year ObD initiative. This report builds on a prior report (Steiner et al., 2017) that described implementation after the second year of the initiative.

ObD was motivated by the unique set of challenges faced by high schools in the United States. High schools are responsible for ensuring that their students develop the skills, knowledge, and dispositions that will prepare them to thrive in college and careers and in their lives. Students enter high school with varying levels of academic and social and emotional needs. Assessment data suggest that many American high school students do not achieve proficiency in academic performance and reveal sizable achievement gaps among students from different racial/ethnic and socioeconomic groups (National Assessment of Educational Progress, undated). High schools are charged with ensuring that students graduate within four years, but many high schools in high-poverty areas struggle to prepare students for graduation within four years (Balfanz, 2009). These findings suggest a continued need for high schools to adopt new strategies to accelerate student learning.

Although policymakers prioritize high schools' contributions to academic achievement, as evidenced by the design of state testing and accountability systems that place significant weight on test scores, students also need to develop a variety of inter- and intrapersonal competencies to succeed in their chosen postsecondary pathways (Aspen Institute National Commission on Social, Emotional, and Academic Development, 2019). Surveys of employers and postsecondary institutions reveal a mismatch between the competencies they would like high school graduates to have and the extent to which graduates display these competencies (Committee for Children and Collaborative for Academic, Social, and Emotional Learning [CASEL], 2016).

Many high schools are addressing these aspects of youth development in the face of large disparities in the readiness of incoming students. Although the ObD initiative concluded prior to the COVID-19 pandemic, the resulting school building closures and disruption to instruction have created new hurdles for high schools to overcome as they work to support equitable access to high-quality supports for learning (Hamilton et al., 2020). Educators therefore need strategies and resources to promote students' academic, social, and emotional development in a way that supports equity of opportunities and outcomes (Aspen Institute National Commission on Social, Emotional, and Academic Development, 2019).

The Opportunity by Design Initiative

To help historically marginalized students graduate from high school with the academic and life skills necessary for postsecondary success, CCNY launched the ObD initiative in 2013 to support the design and creation of a network of small high schools of choice that focus on ten design principles (see Box 1). The ObD initiative was based on the premise that promising high school reform strategies need to be integrated into a comprehensive school design and accompanied by appropriate, sustained levels of financial, policy, and implementation supports (Hamilton and Mackinnon, 2013).

As we describe in the prior report (Steiner et al., 2017), ObD districts were selected through a competitive process. CCNY aimed to select districts in which a set of district-level enabling conditions that

were likely to enable ObD implementation were in place. CCNY drew on school-reform implementation research to identify these conditions, which were district leader support for ObD, financial stability, a commitment to a portfolio strategy,[1] a commitment to innovation, technology quantity and quality, and an emphasis on ambitious college- and career-ready standards (CCNY, 2014).

The ObD districts—Cleveland, Ohio; Denver, Colo.; New York, N.Y.; Philadelphia, Pa.; Prince George's County, Md., Providence, R.I.; and a small charter school organization in Brooklyn, N.Y.—were large, urban districts. The ObD districts were expected to design school models consistent with the district's overall school improvement plan and to convene school design teams consisting of the school leader, teachers, community partners, and district staff, such as curriculum writers and technology support providers. School design teams were responsible for developing ObD school structures and systems that incorporated the design principles in a way that would meet the needs of the student population and district context. CCNY hoped that the design principles, if fully implemented, would result in a school that functioned differently from a traditional high school.

[1] A portfolio strategy prioritizes a city- or community-wide emphasis on developing a diverse portfolio of high-quality public schools designed to cater to community needs and supports parents in choosing the best school for their children (see Center on Reinventing Public Education, 2018, for more information on this subject).

BOX 1

ObD Design Principles

A high-performing high school . . .

1. has a clear mission and coherent culture
2. prioritizes mastery of rigorous standards aligned to college and career readiness
3. personalizes student learning to meet student needs
4. maintains an effective human capital strategy aligned with school model and priorities
5. develops and deploys collective strengths
6. remains porous and connected
7. integrates positive youth development (PYD) to optimize student engagement and effort
8. empowers and supports students through key transitions into and beyond high school
9. manages school operations efficiently and effectively
10. continuously improves its operations and model (CCNY, 2014).

Snapshot of the ObD Initiative

With the ObD initiative, the Carnegie Corporation of New York (CCNY) aimed to incorporate multiple research-based best practices for high school reform in a single major funding initiative. ObD was intended to test whether these best practices could be holistically combined, with expert external support, into a comprehensive school model that could recuperate and accelerate student learning. The ultimate goal of the initiative was to help students graduate from high school within four years with the academic and social and emotional skills needed for postsecondary success. The ObD initiative had three key features:

1. **ten design principles drawn from research on best practices for high school reform to serve as the foundation of the school models**
2. **a design year and two years of implementation support from Springpoint, a newly launched technical assistance partner also funded by CCNY. Springpoint provided principal leadership coaching and site visits followed by feedback, among other supports**
3. **ObD districts were chosen based on the presence of enabling conditions that could support reform, such as buy-in and flexibility from district administration.**

ObD districts were expected to support the schools' development at a policy level (e.g., create or change policies that support mastery-based learning) and with resources to design the school models (e.g., space to work, hiring a school principal), as well as provide support for specific design activities (e.g., vetting or developing curriculum materials and assessments, providing consulting or coaching support for the design team). CCNY worked to ensure that the ObD schools were free from district policies that could present challenges to implementation, such as a mandated curriculum or professional development (PD), seniority-based hiring rules, and grading structures. CCNY negotiated with participating districts to waive many of these requirements in the first year or two of the initiative.

The initiative was intended to test whether the design principles could be implemented holistically to help schools use innovative instructional methods, organizational and staffing structures, and partnerships to help historically marginalized students prepare for life after high school. The initiative provided funding and technical support to the schools during a design year prior to opening the school and during the first two years of implementation (although some schools chose to receive additional support during the third or fourth years of implementation)—with the expectation that the schools would design their models to be self-sustaining on public funds.

Students opted to attend the ObD schools, which were open to students throughout the district, because they were schools of choice. There were a total of 16 ObD schools across seven large, urban districts. The size of the ObD schools—about 400 students per school at full enrollment—was intentionally small relative to most comprehensive high schools in the districts. The number of ObD schools in a given district ranged from one to seven. All but one ObD school opened with ninth grade only and added one grade level each year; one school opened with ninth and tenth grades. The first ObD school opened in fall 2014, and the last opened in fall 2017; all 16 ObD schools were included in the analysis.

Following research that comprehensive school redesign initiatives, such as ObD, are often most successful when they are embedded within larger systems of support (Bloom, Thompson, and Unterman, 2010; Bloom and Unterman, 2012; Le and Frankfort, 2011) and coincide with enabling conditions that can support innovation, CCNY aimed to select grantees based on the presence of a series of enabling conditions (e.g., district leader support, financial stability, standards for college and career readiness), as well as support from community partners (Hamilton and Mackinnon, 2013).

However, in implementing comprehensive reforms, school districts often grapple with tension between providing schools with the autonomy and flexibility they need to innovate and acknowledging constraints present in the district context (e.g., end-of-year grade-level tests, union contracts, curriculum requirements) (Gross and DeArmond, 2018; Pane et al., 2017; Steiner et al., 2017). Recognizing this challenge, CCNY designed the initiative to provide ObD principals with technical assistance from an external provider. CCNY therefore launched Springpoint: Partners in School Design, a national school design organization, to support ObD school and district leaders in designing, establishing, and sustaining innovative high schools. Springpoint focused on supporting ObD school leaders. School

leaders, in turn, were expected to design and deliver PD for teachers. ObD school leaders were supported by other external partners (e.g., technical assistance or curriculum providers) and by the districts in which they were located. In two districts, the ObD schools were supported by intermediary organizations that provided PD and coaching to school leaders, vetted and recommended curriculum materials, and served as another community of practice for school leaders, as shown in the ObD theory of change in Figure 1.1.

The ObD Design Principles

Although the schools participating in the ObD initiative were not expected to implement specific programs (e.g., specific online curricula), they were required to design school models that incorporated the ten design principles in a way that was consistent with the school mission and local context. This occurred with support from Springpoint, other external partners, and district leaders during a design year prior to opening the school. CCNY's design principles were intended to serve as guidelines and describe desired instructional practices and school conditions rather than a well-defined intervention or model. Each school design team had the flexibility to design a model that would work best with the local context, students, and goals and was expected to continuously refine the model over time.

The ten design principles, which are listed in Box 1, were intended to focus on the foundational aspects of high schools. Five design principles were related to the schools' culture and instructional approaches: clear mission and coherent culture, prioritizing mastery of rigorous standards, personalized learning (PL) to meet student needs, PYD, and empowering and supporting students through key transitions. The other five design principles were more focused on school operations and management: human capital, collective strengths, remaining porous and connected, effective and efficient management of school operations, and continuous improvement. Detailed descriptions of the design principles can be found in Appendix C.

The three design principles that focused on instructional approaches reflect key aspects of innovative PL approaches. PL approaches are intended to support all students' social, emotional, and academic development; promote equity; and accelerate learning by personalizing, or individualizing, instruction to address each student's strengths, needs, and interests. A number of recent initiatives, such as the XQ initiative, the state of Maine's proficiency-based learning initiative, and the Barr Foundation's

FIGURE 1.1 **ObD Theory of Change**

"Doing High School Differently" initiative, have focused on redesigning high schools to focus on personalizing instruction for students (Barnum, 2018; Hamilton, 2018; XQ, undated). The 2016 White House Convening for Next Generation High Schools highlighted personalization of learning, competency-based progression, and innovative uses of education technologies as strategies that can support educators' efforts to prepare all their students for college and careers (U.S. Department of Education, undated). Similarly, in 2017, the Center for American Progress proposed personalization of learning as a crucial feature of redesigned high schools that can better prepare students for success in life after graduation (Jerald, Campbell, and Roth, 2017).

This report emphasizes these three instructionally focused design principles: *prioritizing mastery*, *personalizing learning*, and *PYD*, which we refer to as *key design principles* throughout the report. CCNY highlighted these key principles as foundational to the schools' models and instructional approaches based on schools' experiences in the first two years of the initiative. We also touch on several of the design principles more closely related to school operations (e.g., hiring, data use) because they support high-quality instructional practices.

Although there is not yet widespread consensus about the instructional approaches that define these key design principles, the literature identifies some elements that generally characterize each of them (Aspen Institute National Commission on Social, Emotional, and Academic Development, 2019; Pane et al., 2017; Steele et al., 2014). Box 2 provides the definitions that CCNY used in the context of ObD.

In many conceptualizations, mastery-based practices[2] include clearly defined learning targets designed to support deep understanding and application of knowledge by allowing students to move through content at their own pace and attempt a task or assignment multiple times (Gross and DeArmond, 2018; Pane et al., 2017; Sturgis, 2012). Personalizing learning often involves practices such as providing students with choice in content or topic and are designed to tailor learning experiences to students' individual learning needs and interests.

Personalized and mastery-based approaches frequently intersect with PYD in that they include efforts to provide students with more control over their learning and access to experiences that will support their social and emotional development in addition to academic skills (Gross, Tuchman, and Patrick, 2018). A focus on academic as well as social and emotional development (with the latter sometimes referred to as *PYD*, or, more recently, *social and emotional learning* [SEL]) has been shown to promote both short-term outcomes, including academic achievement and improved behaviors, and longer-term outcomes, including workforce readiness and success in postsecondary education (Deming, 2017; Grant et al., 2017; Jackson et al., 2020; Mahoney, Durlak, and Weissberg, 2018; Taylor et al., 2017). PYD can be especially important in high school, when students are preparing for college or work (Nagaoka et al., 2015; Steiner et al., 2017). In addition to focusing on building positive social and emotional skills, schools that adopt PYD broadly aim to take a positive, constructive (in contrast to deficit-based) approach to develop youths' strengths and interests and build engaging and supportive environments throughout school systems, policies, and instruction (Taylor et al., 2017). Some research has suggested that successful schools enact a coherent set of policies and practices that emphasize high-quality, rigorous instruction and incorporate PYD for all students (Lake, Hill, and Maas, 2015).

In practice, the ObD schools adopted a variety of strategies related to each of the three key design principles. The three key design principles were intended to intersect and be mutually reinforcing—PYD can

BOX 2

Key ObD Design Principles Addressed in This Report

- **Prioritizing mastery:** Students demonstrate deep understanding of clearly defined, rigorous competencies.
- **Personalizing learning:** Student learning experiences are tailored to individual learning needs and interests.
- **Positive youth development (PYD):** Students have a voice in their learning and access to experiences and relationships that help them develop the skills and mindsets to succeed (CCNY, 2017).

[2] Mastery-based practices are sometimes described as proficiency-based or competency-based practices and generally include similar instructional practices (Steele et al., 2014).

support effective personalization and mastery-based approaches—and schools were charged with finding the right balance among them. Figure 1.2 presents examples of the practices we documented in this research.

Although there is some evidence that personalized, mastery-based, and PYD approaches can improve student academic, social, and emotional outcomes in some contexts, the evidence reveals several important implementation challenges. Several studies of personalized and mastery-based instructional approaches in non-ObD schools with a variety of characteristics (e.g., school size, urbanicity, grade level) have documented a common set of barriers to effective implementation (Gross and DeArmond, 2018; Pane et al., 2017; Steele et al., 2014). For example, many schools lack curriculum materials that are appropriate to personalized and mastery-based learning environments. As a result, teachers often develop their own curriculum materials, which may vary greatly in quality. In addition, schools often have difficulty finding or developing high-quality assessments that are appropriate for mastery-based learning environments and face policy barriers to implementation of mastery, such as seat time requirements or curriculum pacing requirements. In many schools, the criteria for mastery of the material can vary by teacher; that is, sometimes mastery is defined as completion of tasks rather than demonstration of deep understanding of the skills and content. These challenges can sometimes result in inequitable opportunities for students. Studies of SEL interventions that aim to promote aspects of PYD have identified some similar challenges, including inadequate access to high-quality instructional guidance and materials, along with lack of a common language or definition of SEL (Marsh et al., 2018; Schwartz et al., 2020).

FIGURE 1.2 **Examples of Mastery-Based Practices, Personalization Practices, and PYD Practices in ObD Schools**

Mastery-based practices include

- different students work on different topics of skills at the same time
- students have the chance to work through instructional material at a faster or slower pace than other students in the class
- students have opportunities to review or practice new material until they fully understand it
- students have multiple opportunities, throughout a unit or throughout the year, to demonstrate mastery of certain content and skills
- students can make multiple attempts at tasks that count toward mastery.

Personalization practices include

- tailoring the pace or content of instruction to individual students' needs
- tailoring instructional strategies to individual students' needs
- adapting course content to meet students' needs by providing additional assignments, resources, and activities for remediation or enrichment
- providing students with opportunities to choose what instructional materials (such as books or computer software) they use in class
- providing a variety of materials or instructional approaches to accommodate individual students' needs and interests.

PYD practices include

- connecting academic content to themes related to SEL
- addressing topics that can help students improve their academic achievement, such as time management, collaborating with other students, persisting through challenging tasks, and making responsible decisions during instruction
- implementing strategies to promote positive student behavior
- helping students develop agency, or a sense of control over their learning
- encouraging staff to build positive relationships with students and get to know their learning needs and interests to support personalization.

NOTE: This list of practices appears in another report that relies on the American Teacher Panel (ATP) data collected as part of the ObD project (Steiner, Doss, and Hamilton, 2020).

Research Approach

In this final report, we aim to present a holistic picture of the ObD initiative to provide guidance and lessons for the field about designing and sustaining innovative high schools.

We began a five-year study of the ObD initiative in June 2014. We monitored implementation over four years (the 2014–2015 school year through the 2017–2018 school year) and collected student outcomes for those years and one additional year. The study was designed to address four research questions (RQs) that examine students' academic and behavioral outcomes and implementation of the schools' models.

1. How did student outcomes in ObD schools compare with outcomes of similar students in other schools in the same districts and across the United States?
2. How did teachers and other school staff implement the key design principles, and what factors might have facilitated or hindered implementation?
3. What system-level conditions supported or hindered implementation?
4. How did ObD teachers' practices and perceptions of implementation enablers and challenges compare with those of a nationally representative sample of high school teachers?

This report builds on a prior report (Steiner et al., 2017) that focused on implementation of the ten design principles after the first two years of the initiative and presented interim findings related to RQs 1 and 2. The prior report included surveys of and interviews or focus groups with school leaders, teachers, and students, along with observations of classrooms and analysis of instructional artifacts in the ten ObD schools that were open at the time. The prior report did not present findings related to student outcomes or compare ObD teachers' practices and perceptions with those of high school teachers nationally (RQs 1 and 4). This report fills those gaps. We examine U.S. high school teachers' reported use of mastery-based and personalized instructional practices in more detail in a separate report (Steiner, Doss, and Hamilton, 2020). Box 3 summarizes key findings from the earlier report.

We examined three categories of outcomes to assess the impact of attending an ObD school: academic outcomes measured by mathematics and English language arts (ELA) assessments; behavioral outcomes measured by attendance and suspension rates; and college readiness outcomes measured by PSAT and SAT scores, credit accumulation, and graduation rates.

Throughout the research, we gathered information about implementation of the schools' models, district context, and challenges and facilitators in all the ObD schools. We interviewed CCNY, Springpoint, and district staff; collected instructional artifacts (e.g., culminating assignments, assessment criteria); and surveyed teachers and students. We also visited each school to interview school leaders and teachers, conduct focus groups with students, and observe classrooms. We conducted a nationally representative survey of teachers, using RAND's ATP, to understand how ObD teachers' reports of their instructional practices compare with those of teachers nationally.

The analyses that we present in this report focus on data collected in the fourth year of the study (school year 2017–2018). At that time, the 16 ObD schools were at various stages of implementation. Some had been in operation for four years and some for only one year. For outcomes analyses, the 2017–2018 data are supplemented with examination of graduation rates in 2018–2019 to include five newer schools that did not enroll seniors until that year. Regardless of experience, all the schools were working to continuously improve their implementation of the design principles. Thus, the implementation findings presented in this report represent multiple years of iteration and experience for some schools and a single year for others.

The ObD Schools

The 16 ObD schools in this analysis were located in seven large, urban districts and include one urban charter school; they served large proportions of minority students from low-income families. The distribution of schools across districts and cohorts is shown in Table 1.1. Half of the schools were reorganizations of existing schools. Students opted to attend the ObD schools, which were open to students throughout the districts. All but one ObD school opened with ninth grade only and added one grade level each year; one school opened with ninth and tenth grades. Most schools served primarily high-minority, high-poverty students, and two

BOX 3

Implementation Successes and Challenges After Two Years of the ObD Initiative

After the first two years of implementation, the ObD schools demonstrated progress but also faced numerous challenges. Across schools, staff and students reported that implementation of the design principles was high across schools and improved over time. Students in all schools reported positive perceptions of school culture, and teachers' opinions of their PD experiences improved over time. Teachers in Cohort II schools reported more extensive adoption of mastery-based and personalized instructional practices, and fewer obstacles to adopting such practices, than teachers in Cohort I schools. District and school leaders reported that the coaching and technical assistance provided by Springpoint supported implementation and helped spread best practices and align systems in a way that could help facilitate the spread of innovation.

However, at this early stage, schools' efforts to personalize instruction and implement mastery-based approaches were often inconsistent and limited by varying access to data, external pressure to advance students at a certain pace, and the significant time required to create instructional materials. Schools were still working to integrate PYD approaches with mastery-based and personalized instructional practices, and, in general, school staff described PYD as separate from academics.

Human capital challenges and limited availability of high-quality instructional materials exacerbated these challenges. Persistent teacher vacancies limited collaboration and strained teacher capacity, and principals reported difficulties finding and retaining qualified, experienced teachers. Although teachers' reports of their PD experiences improved over time, the quality of teacher PD reportedly varied across schools. Although teachers reported that they had the autonomy to create their own materials, they struggled to find the time to do so and received limited district support. Most schools had yet to develop clear systems for data-driven improvement, instead responding to feedback and addressing issues as they arose.

schools primarily served English language learners. Approximate enrollment across all 16 schools was 3,580 students; most schools had about 100 students per grade level. More information about the demographic characteristics of these schools can be found in Chapter Two (Table 2.1).

In the 2017–2018 school year, five schools had been in operation for four years, five schools for three years, two schools for two years, and four schools for one year. We refer to these groups of schools as follows:

- **Cohort I:** Five schools in their fourth year of operation served students in grades 9–12.
- **Cohort II:** Five schools in their third year of operation served students in grades 9–11.
- **Cohort III:** Two schools in their second year of operation; one school served students in grades 9–10 and the other school served students in grades 9–11.
- **Cohort IV:** Four schools in their first year of operation served students in grade 9.

Outcomes Data Sources and Sample

Our student outcomes analyses relied mainly on administrative data provided by the participating school districts. Each district provided data from ObD schools and other district schools.[3] These were supplemented with additional achievement data from a national testing company. We received baseline achievement and demographic data, as well as outcomes data in three categories: academic outcomes measured by mathematics and ELA assessments; behavioral outcomes measured by attendance and

[3] Data from the Brooklyn-based charter management organization were provided by the New York City Department of Education.

Table 1.1. Number of Schools, by District and Cohort

Cohort	Brooklyn Charter School	Cleveland	Denver	New York	Philadelphia	Prince George's	Providence
I		2		1	2		
II			1			2	2
III		2					
IV	1	3					

suspension rates; and college readiness outcomes measured by PSAT and SAT scores, credit accumulation, Advanced Placement® (AP) tests passed, and graduation rates.

The district data sample included students enrolled in grades 9–12. In total, there were 3,580 students enrolled in ObD schools and 446,801 comparison students from other schools. In all but one of the ObD sites, additional mathematics and reading assessment data were obtained from a national testing company, covering 1,870 ObD students and 36,235 comparison students from other schools.

Chapter Two and Appendix A contain additional details about the sample and data, along with information about the analytic methods we employed.

Implementation Data Sources and Sample

We drew on a number of data sources to understand implementation. We conducted interviews with CCNY and Springpoint staff and key leaders in each district. We surveyed ObD teachers and students and administered the ObD teacher survey (with a few revisions to omit ObD-specific references) to a nationally representative sample of high school teachers via RAND's ATP. We also visited each ObD school to interview a sample of school leaders and teachers and conduct a focus group with students. The numbers of interview participants, focus groups, and survey response rates in 2017–2018 are summarized in Table 1.2. Additional information about each of these data collection methods and our analytic approach is in Appendix B.

The implementation findings presented in this report comprise a synthesis of the implementation data. We rely heavily on teacher and student survey data because those sources are the most representative of teachers' and students' attitudes and perceptions. We also rely on the interviews with principals, teachers, and district staff, which, although less representative than the surveys, provide in-depth information about key aspects of implementation that can help clarify patterns in the survey data and illuminate comparisons among schools and districts. We triangulated these sources with student focus group data as applicable.

Limitations

Even though this report represents the final presentation of findings from this research, we urge readers to recognize that several of the ObD schools were still relatively early in their implementation, while others had been in operation for three or four years. The outcomes study relies on relatively sparse student outcomes data and methods that are vulnerable to selection bias. Apparent differences between ObD and comparison samples might not have been caused by ObD. Limitations of the outcomes study are discussed in greater detail in Chapter Two.

Designing and launching new high schools is a highly complex endeavor and takes time. The ObD schools were expected to implement the design principles in a way that was consistent with their school contexts. Furthermore, the ObD schools were charged with implementing innovative instructional practices for which there were few exemplars. Thus, we should not expect all of the design principles to be implemented in a comprehensive way across schools. Readers should also keep in mind that the data on implementation are limited by their self-reported nature and relatively small sample sizes, which are themselves limited by the size of the schools. In addition, our efforts to collect data about implementation were subject to time and resource constraints, and we were not able to collect information about every aspect of the schools' operations. We did not, for example, collect data about ObD teachers' working conditions, such as class sizes, schedules, or compensation, or how these conditions were different from other high schools in the ObD districts or high schools nationally. These, and other, working conditions could affect the time teachers spend on key aspects of the school model, such as

Table 1.2. Numbers of Interview Participants, Focus Groups, and Teacher Survey Response Rates, 2017–2018

Data Collection Method	Source	Number of Participants	Response Rate (%) Among Initial Sample (if applicable)	Range of Response Rates Across Schools (%) (if applicable)
Interviews	CCNY and Springpoint staff	4	—	—
	District and intermediary leaders	15	—	—
	School leaders	16	—	—
	Teachers	64[a]	—	—
Focus groups[b]	Students	98	—	—
Surveys	Teachers	61	81	60–100
	Students[c]	1,070	88	70–100
	American Teacher Panel [d]	1,009	56	—

NOTES: [a] In two schools, we interviewed a PL coach and a mastery specialist, both of whom are classified as teachers because they instructed students in addition to their leadership responsibilities.

[b] Across schools, we held 16 focus groups.

[c] Student survey response rate is given among students who gave consent.

[d] RAND's ATP is a nationally representative panel of public school teachers who provide their input on a wide range of education topics. Information about the ATP is available on RAND's Education and Labor webpage (RAND Education and Labor, 2020).

curriculum development, data analysis, or mastery grading. Readers are encouraged to review a more detailed discussion of implementation study methods and limitations in Appendix B.

We also recognize that schools across the United States have begun adopting many of these practices even when they do not explicitly espouse personalized and mastery-based approaches. The ObD schools were among the early adopters of these innovative instructional practices, and while we would expect to see them implemented to a greater extent in the ObD schools than in traditional schools, there is not necessarily a clear distinction between what constitutes "typical" practices and what constitutes practices aligned with the design principles. This report fills a gap in the prior report by including data from a national comparison group. Nevertheless, the discussion of implementation in this report is largely descriptive. Despite these limitations, these data allow us to provide detailed examples of the design principles in the ObD schools and rich descriptions of implementation facilitators and challenges and to examine areas of similarities and differences across schools and districts with a view to sharing lessons for the field.

We discuss student outcomes in Chapter Two and implementation of the three key design principles during the fourth year of the study in Chapter Three. In Chapter Four, we discuss key contextual factors relating to implementation, including district enabling conditions and supports, as well as the role of Springpoint and other external partners. In Chapter Five, we frame the ObD results in a national context and compare ObD teachers' reports about their instructional practices and supports with those of teachers nationally, using data gathered from RAND's ATP. In Chapter Six, we focus on implications and lessons for the field. We hope the discussion in this report will be of interest to schools that employ mastery-based, personalized, and PYD approaches; to the funder; and to educators and policymakers who are implementing, or considering implementing, similar reforms.

CHAPTER TWO

Student Outcomes

Summary

We examined three categories of outcomes to assess the impact of attending an ObD school: academic outcomes, measured by mathematics and ELA assessments; behavioral outcomes, measured by attendance and suspension rates; and college readiness outcomes, measured by PSAT and SAT scores, credit accumulation, AP tests passed, and graduation rates. The available data differed across districts and were, in some cases, sparse. We therefore selected a set of common measures to enable estimation of overall effects of ObD and supplemented the district-provided data with additional achievement measures. Each analysis compared the ObD students with a comparison group of non-ObD students, matched on prior achievement and demographic characteristics. For district data, the comparison group consisted of students in non-ObD schools in the same district; for supplemental achievement data, the comparison group was drawn from a national database of students who took the same assessments. In addition to making the ObD and comparison samples as similar as possible, we statistically adjusted for any remaining differences between them and then estimated the effects for each grade level within each district. We then used meta-analysis to produce district-specific estimates and to combine those into estimates of the overall effect of ObD on that outcome.

KEY TAKEAWAYS

- Focusing on estimated effects of the ObD initiative overall, the study found no statistically significant evidence of improved student performance as captured by 12 measures of academic, behavioral, and college readiness outcomes.
- Estimates of district-specific academic effects tended to be negative for state assessments of standards attainment and positive for achievement growth measured by MAP.
- Estimated district-specific effects on attendance and suspension were mixed, with a similar number of positive and negative estimates. However, there is suggestive evidence that attendance and suspension may be key antecedents of improved credit accumulation and graduation.

This chapter addresses RQ 1: How did student outcomes in ObD schools compare with outcomes of similar students in other schools in the same districts and across the United States?

Data

Our student outcomes analyses relied mainly on administrative data provided by the participating school districts. The districts provided data for ObD schools, and for other schools in the district for comparison. In most districts, the comparison sample included all of the other high schools in the district; however, two districts restricted the comparison sample to a subset of schools they deemed as serving student populations most comparable to the ObD sample. The available data were widely diverse across districts, and from those we sought a set of common measures to enable estimation of overall effects of ObD. For some outcomes, we focused on data from a single grade across districts; for others, we focused on the grade for which data were most prevalent in each district. The same focus was applied to the data from comparison schools within the same district. We supplemented the district-provided data with some additional achievement measures, described below.

Although we received data from the 2014–2015 through 2018–2019 academic years, we generally focus on the 2017–2018 academic year for the following reasons. Selecting a more recent year enables capturing the operation of the ObD schools as long as possible and maximizes the number of students who have available college readiness indicators. However, 2018–2019 data were considerably sparser than those for 2017–2018 because some districts discontinued providing data, making 2017–2018 the richest data set. We made one exception by also including graduation data for 2018–2019, because that enabled obtaining the metric in two additional districts that had not yet enrolled seniors in 2017–2018.

Achievement Data

For student achievement measures, the districts provided state standardized assessment results. Many states tested students in mathematics and ELA only once during the high school years, but the tested grades varied across states. Tested grades also varied because some states use end-of-course exams for courses that are not taken during a specific grade, and some of the courses and tests are retaken in a subsequent year if failed on the initial attempt. From among this diverse set of assessments and grade levels, we selected the most common mathematics and ELA assessment in each district. For mathematics, these were end-of-course algebra exams or the PSAT 8/9 mathematics exam, all administered in ninth grade. For ELA, these were end-of-course ELA exams or the PSAT 8/9 verbal exam administered in ninth grade, a literature end-of-course exam administered in tenth grade, and a general ELA exam administered in 11th grade. Our methods for dealing with this variety of measures are discussed in the methods section of this chapter. All districts provided mathematics and ELA assessment data through at least the 2017–2018 academic year.

Recognizing the evaluation challenges posed by this sparse and disparate set of measures, we sought to supplement them with a common achievement measure across districts that could be used in multiple grades and that might be more sensitive than state assessments to off-grade-level achievement growth that may occur as a result of personalized, mastery-based instructional approaches. To that end, we asked ObD schools to administer the NWEA MAP® Growth mathematics and reading interim assessments to students in grades 9 to 11 in the fall and spring of each academic year. MAP is an online adaptive test that adjusts the consecutive difficulty of questions in response to an individual student's answer. If a student responds incorrectly, the next question is easier; if a student responds correctly, the test software progresses to a more difficult question. The MAP assessment provides accurate scores on a common scale over a broad range of student ability from kindergarten to grade 11. These features of MAP make it sensitive to student growth even if that growth is off grade level. As such, MAP can capture growth that might not be captured on other assessments designed to measure on-grade-level attainment of standards.

Six of the ObD districts either already used MAP or agreed to administer it in their ObD schools. We generally lacked within-district comparison data for MAP, so we obtained matched comparison data from NWEA, which maintains a database of tests taken by students across the country. For each ObD student and subject, NWEA provided comparison data for up to 51 students, matched on grade, gender, starting MAP score, time elapsed between starting and ending MAP score, and the school-level variables of locale (e.g., urban) and percentage of students eligible for free or reduced-price lunch.

Behavioral Data

Districts provided data on attendance and suspensions for all the grade levels enrolled in the ObD schools for a particular year. Six of the seven districts provided these data through at least the 2017–2018 academic year.

College Readiness Data

We sought a variety of college readiness indicators, including scores from the PSAT (also known as the *PSAT 10* or *PSAT/NMSQT*) and SAT, credit accumulation, the number of AP tests passed (score of 3 or higher), and graduation indicators.

Six districts provided grade 10 PSAT scores. The same six districts reported SAT scores for either grade 11 or 12, and we used the reported data. For both the PSAT and the SAT, we used the combined mathematics and verbal scores.

Five districts reported credit accumulation, which generally spanned all enrolled grades, except for one district that omitted grade 9. We analyzed all of the reported data. Two districts also provided the number of AP tests students passed. One district, in which the ObD school served grades 9–11, reported AP pass data for those three grades; the other district reported AP pass data for grade 12 only.

Three districts reported graduation status for seniors in 2017–2018 and five districts in 2018–2019. One district had not yet enrolled seniors by 2018–2019.

Sample

Figure 2.1 summarizes some of the demographic characteristics of the ObD sample, along with the demographics of the district-provided comparison sample and of the district overall, according to the National Center for Education Statistics' Common Core of Data. Most of the districts had high-minority, high-poverty students in their ObD schools, and one had particularly high English-learner populations. As described in the next section, although the characteristics of the comparison schools do not always closely align to the ObD schools, our methods make the comparison sample more similar to the ObD sample for analysis.

Henceforth, we deidentify the districts with anonymous labels (District 1, etc.) to protect their anonymity when reporting findings related to ObD implementation and outcomes. The labels and placement in subsequent figures and tables have no correspondence with the alphabetical ordering displayed in Figure 2.1.

FIGURE 2.1 **Summary of Select Demographic Characteristics, by District, 2018–2019**

		Brooklyn LAB	Cleveland	Denver	New York	Philadelphia	Prince George's	Providence
Black (%)	ObD	27	68	15	26	56	8	17
	Comparison	28	64	13	28	54	53	16
	Whole district	23	65	13	23	50	58	17
Hispanic (%)	ObD	53	18	26	46	39	90	66
	Comparison	40	16	51	40	20	28	67
	Whole district	41	16	55	41	20	33	65
ELL/LEP (%)	ObD	29	7	2	4	12	83	32
	Comparison	12	6	17	12	11	15	31
	Whole district	15	9	28	15	11	19	28
Disability (%)	ObD	16	26	16	16	19	4	9
	Comparison	18	22	13	18	16	10	14
	Whole district[a]	26	21		26	18	11	17
Poverty (%)	ObD	86		36	74	53	94	80
	Comparison	74		61	74	44	58	81
	Whole district	72		67	72	96	66	87

NOTE: ELL = English language learner; LEP = limited English proficiency.
[a] Whole district data include all students with individual education plans, including gifted.

Methods

For both the district-supplied data and the MAP data, we applied a two-stage approach to adjust for observed differences between the ObD and comparison samples. To reiterate, analyses of district data used non-ObD students in the district for comparison, while analyses of MAP scores used assessment data from students across the country. The first stage made the two samples as similar as possible by weighting (district data) or matching (MAP data) on eighth-grade assessment scores and demographic characteristics related to student outcomes, such as those displayed in Figure 2.1. We elaborate the full sets of variables in the appendix. The second stage included the same sets of variables as covariates in estimating the ObD effects. This two-stage process is often referred to as doubly robust because failure of either stage to properly adjust for sample differences can be mitigated by the other stage (Bang and Robins, 2005; Hullsiek and Louis, 2002). However, as noted in the limitations section, even the doubly robust approach does not guarantee that estimated effects are unbiased.

The analyses of district-supplied and MAP data differ in regard to the baseline used for matching and covariate adjustment. In the case of district-supplied data, eighth-grade baselines were used for all analyses. This means that the analysis captures the cumulative effect of ObD enrollment from entry in ninth grade through the date when the outcome was collected. The MAP matching methods do not function well over multiyear timespans; therefore, MAP matching and covariate adjustment are based on fall baselines from the same academic year as the spring outcomes. As a result, the MAP analyses capture single-year fall-to-spring effects of ObD enrollment. To explore longer-term cumulative effects, we also examined student performance relative to national MAP norms over several years, as discussed later.

Because the outcome measures and grade levels often differed across districts, we used the following approach to estimating district-specific and overall ObD effects. First, within each district, we estimated standardized effects within each grade level and used meta-analysis to combine the grade-level estimates into an overall estimate for the district. Then, we again used meta-analysis to combine the district-specific estimates into estimates of the overall effect of ObD on that outcome. For this second stage, we used random-effects meta-analysis to help ensure that the largest districts would not dominate the estimated overall effects.

We estimated 75 effects of ObD overall or within districts. By domain, there were 30 estimates of academic effects, 14 estimates of behavioral effects, and 31 estimates of college readiness effects. Within each domain, we performed corrections for the multiplicity of statistical tests, using the Benjamini-Hochberg (1995) method for controlling false discovery.

The ObD sample included large percentages of students of color, low-income students, and students whose achievement was relatively low. For achievement in particular, on entry into ninth grade, most cohorts of ObD students ranked below the 25th percentile on national MAP norms (Thum and Hauser, 2015). In our main analyses, we compare the performance of ObD students with that of peers with similar characteristics and baseline academic performance. We complemented this with a descriptive analysis of how ObD students performed relative to national norms, with no adjustment for how they differ from a nationally representative sample. We followed cohorts of students during their ObD enrollment in a time series of their rankings relative to norms for their grade level. To avoid trends being influenced by students entering or leaving ObD schools in the interim, students were included only if their scores were available for both the beginning and end of the time series. Nationally, normative mean MAP scores increase from about 140 at entry into kindergarten to about 220 (reading) and 230 (mathematics) at entry into ninth grade, then increase very little through 11th grade, remaining below 223 (reading) and 235 (mathematics). Although the reason for this plateau is unclear, one possibility is that the assessments do not capture more advanced topics typically covered in high school. By that hypothesis, our time series analysis would be most sensitive to whether ObD students gain on the more basic skills that are captured by MAP, which may be more likely where schools implement a personalized, mastery-based instructional approach.

All of these methods are documented in greater detail in Appendix A, along with tables showing the improvement in similarity between the ObD and comparison samples after the matching or weighting was applied, as well as more-detailed output from statistical models than is reported in this chapter.

Limitations

The study employed a broad set of outcome measures yet may not fully capture the effects of ObD. Academic measures were limited to mathematics and ELA assessments that do not reflect the breadth of subject matter that high schools cover; behavioral measures were limited to attendance and suspension rates, which may not be very sensitive to the full range of PYD skills that ObD schools sought to impart; and college readiness measures were confined to predictive high school outcomes rather than direct measures of college admission, persistence, or graduation.

Although our outcomes analyses adjusted for observed differences between the ObD and comparison samples, these methods could not eliminate all possible differences between the two. Any remaining differences could influence results, making our analyses vulnerable to selection bias. That is, where analyses find statistically significant differences between the ObD and comparison samples, those differences might have been caused by ObD or might have been caused by preexisting differences between the two samples that were not adequately controlled for with our methods.

Because the MAP time series analysis did not attempt to control for differences between the ObD and nationally representative norming samples, it is particularly vulnerable to selection bias.

We discuss these and other more technical limitations in Appendix A.

Results

Figure 2.2 summarizes the results of the outcomes analyses. The figure displays standardized effect sizes both numerically and with bar graphs, with zero indicated by a vertical line. In the interest of improving readability and interpretability of the multitude of results, certain details are unavailable in the figure: Bar graphs are truncated where effects are extremely large in magnitude, and standard errors or confidence intervals are not displayed. Standard errors are available in Appendix A. Statistical significance (after correction for multiple hypothesis tests) is indicated by pink shading for negative effects and green shading for positive effects; otherwise, the bars are gray. In every case, "good" results are displayed as positive estimates; specifically, a positive estimate for suspensions indicates a decreased suspension rate. Column 2 shows the results for ObD overall, and subsequent columns show district-specific results. The order in which the districts appear has been randomized to protect their anonymity with respect to outcomes.

For ObD overall, most estimates were nonsignificant and no larger than 0.05 in absolute magnitude. Significant negative effects were found for SAT scores (standardized effect size of –0.08) and credit accumulation (–0.27). A relatively large negative effect on state mathematics assessments (–0.17) was not significant.

District by district, larger and statistically significant estimates are more prevalent; however, we discern no clear patterns in these results. The discussion in this paragraph and the next includes estimates of 0.10 or larger even if they are not statistically significant. District 1 had positive effects on attendance, suspensions, and 2018–2019 graduation rates, and negative effects on academic outcomes and SAT scores. District 2 had negative effects on suspensions and credit accumulation. District 3 had positive effects on MAP assessments, suspensions, and credit accumulation, and negative effects on state mathematics assessments and attendance. District 4 had positive effects on AP tests passed and negative effects on state mathematics assessments, attendance, suspension, PSAT and SAT scores, and 2018–2019 graduation rates. District 5 had positive effects on MAP ELA scores, suspensions, and graduation rates (2017–2018 and 2018–2019), and negative effects on state mathematics assessments, attendance, SAT scores, and AP tests passed. District 6, with few available outcomes, had positive effects on MAP mathematics scores. Finally, District 7 had positive effects on MAP ELA scores and negative effects on state ELA scores, suspensions, credit accumulation, and 2017–2018 graduation rates.

Horizontally, across districts, in the academic outcomes domain, effects seem more likely to have been negative for state mathematics assessments and positive for MAP assessments. Effects on behavioral outcomes exhibit no clear pattern. Among college

FIGURE 2.2

Results of Statistical Models Estimating Effects of Enrollment in ObD Schools

All data from 2017–2018 except where noted	Overall	District 1	District 2	District 3	District 4	District 5	District 6	District 7
Academic outcomes								
State mathematics assessment (algebra) or PSAT8/9, grade 9	–.17	–.17	.02	–.28	–.15	–.52	.00	–.02
State ELA assessment or PSAT8/9, earliest grade tested 9–11	–.05	–.12	–.07	.02	.00	–.05	.08	–.16
Mathematics growth (MAP grades 9–11)	.05	–.18	.01	.11		.03	.11	.03
ELA growth (MAP grades 9–11)	.04	–.84	–.06	.04		.11	.01	.16
Behavioral outcomes								
Attendance	–.03	.46	.07	–.15	–.21	–.07		.00
Suspensions (reduction displayed as positive)	.03	.26	–.11	.17	–.08	.04		–.13
College readiness								
PSAT mathematics + verbal, grade 10	–.03	–.02	–.05		–.10	.04		–.03
SAT mathematics + verbal, grade 11 or 12	–.08	–.09	–.01		–.17	–.04		–.05
Credit accumulation	–.27	.10	–1.65	.06		.00		–.12
AP tests passed	.00				.24	–.19		
Graduation 2017–2018	.02	.01				.18		–.15
Graduation 2018–2019	.01	.23	.05		–.43	.20		–.09

NOTE: The horizontal length of the bar represents the magnitude of the standardized program effect estimate, with the vertical lines representing zero. For some extreme estimates, bars are truncated. Bars are green or pink where results are statistically significant after correction for multiple hypothesis tests; otherwise they are gray. Results are not shown where insufficient data were available.

About Figure 2.2

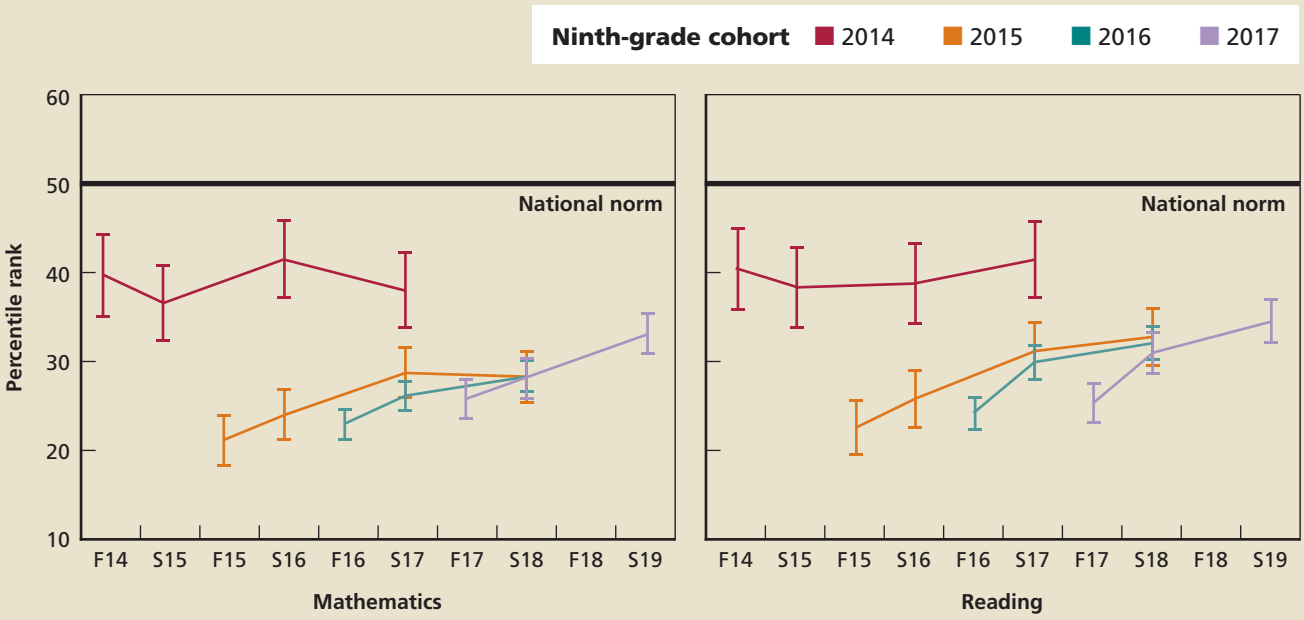
- **These analyses compare ObD student performance relative to matched peers from other schools in the same district, except for MAP analyses, where matched comparisons were drawn from a national database.**
- **Numerical results are presented in terms of standardized effect sizes and graphically portrayed as bar graphs, with the vertical lines representing zero. For some extreme estimates, bars are truncated.**
- **Gray bars indicate that the estimated effect is not statistically significant after adjustment for multiple comparisons.**
- **Green bars indicate statistically significant positive estimates, and pink bars indicate statistically significant negative estimates.**
- **Additional details and standard errors are presented in Appendix A.**

readiness outcomes, effects on PSAT and SAT scores and credit accumulation tended to be negative.

Figure 2.3 displays how cohorts of ObD students ranked versus national norms for their grade level over several years. The left panel shows mathematics, and the right panel shows reading. Each color represents students enrolling in ObD schools as ninth-graders in a certain year from 2014 through 2017. The first data point for each cohort is their national ranking in the fall of ninth grade; subsequent data points show their ranking in the springs of ninth grade and subsequent grades. Although the 2014 cohort started at a higher

FIGURE 2.3

Time Series of Ranking of Cohorts of ObD Students Versus National MAP Norms for Their Grade Level



About Figure 2.3

Figure 2.3 displays how cohorts of ObD students ranked versus national norms for their grade level over several years. The left panel shows mathematics, and the right panel shows reading. Each color represents students enrolling in ObD schools as ninth-graders in a certain year from 2014 through 2017. The first data point for each cohort is their national ranking in the fall of ninth grade; subsequent data points show their ranking in the springs of ninth grade and subsequent grades. An upward trend indicates that the cohort of ObD students experienced achievement gains relative to their national peers on skills measured by MAP. Baseline differences between cohorts are affected by entry of additional schools into the ObD initiative each year, each with their own particular student characteristics.

baseline, this is likely due to the specific students and schools that were part of the ObD initiative that year. Baseline means decreased when more schools entered into the initiative the next year. The upward trend seen for most cohorts indicates that ObD students experienced achievement gains relative to national norms on skills measured by MAP.

Summary of Outcomes Analysis

Focusing on estimated effects of the ObD initiative overall, the study found no statistically significant evidence of improved student performance as captured by 12 measures of academic, behavioral, and college readiness outcomes. As seen in column 2 of Figure 2.2, there was a significant negative effect on credit accumulation. This was driven by a large negative effect in one district, without which the overall effect would have been near zero and nonsignificant. There

was also a significant negative effect on SAT scores, reflecting negative estimates in all five districts reporting this metric. We investigated whether ObD schools encouraged more lower-performing students to take the SAT. Four of the five districts did exhibit higher rates of SAT-taking in the ObD group than the comparison group (by 1 to 11 percent), with the other exhibiting a lower rate (by 6 percent). However, there were insufficient data to confirm that this explains the negative SAT effect.

As we examined patterns in district-specific analyses, we found that estimated academic effects tended to be negative on state assessments of standards attainment and positive on achievement growth measured by MAP. This pattern holds for significant academic effects, except for one large negative effect on MAP in one district, and is echoed in generally upward-trending MAP growth relative to national norms in Figure 2.3, although that time series analysis is a weaker approach. Such a pattern could emerge from implementation of personalized and mastery-based learning if ObD schools placed relatively greater emphasis on student growth, even if off grade level, and less emphasis on attainment of grade-level standards—or if MAP was better able to measure the concepts and skills that students improved on than were state assessments. Using implementation data, we cannot determine the extent to which the ObD schools emphasized off grade level student growth, but it is likely that the ObD schools emphasized this to a greater degree than non-ObD schools. As we discussed earlier, normative MAP growth plateaus in high school, possibly because the assessment is better at measuring more-basic skills than those covered in higher-level high school courses. Sites implementing personalized or competency-based learning have exhibited similar patterns where effect estimates were available both for measures of achievement growth and for attainment of grade-level standards (e.g., Margolis, 2019; compare with Ready et al., 2019; Zimmerman and Kuhlmann, 2019).

Estimated district-specific effects on attendance and suspension were mixed, with a similar number of positive and negative estimates. However, there is suggestive evidence that attendance and suspension may be key antecedents of improved credit accumulation and graduation. Although based on a qualitative interpretation of the quantitative results, there is theoretical logic that students must be present and engaged to accumulate credits and graduate. Positive effects appeared in credit accumulation or graduation in only three districts—the same three districts that showed positive effects on at least one of these behavioral outcomes. Otherwise, effects on college readiness outcomes were mixed, with negative effects more prevalent than positive.

CHAPTER THREE

School-Level Implementation of the Key Design Principles

Summary

This chapter addresses how the ObD schools defined and implemented the key design principles related to instruction—mastery, personalization, and PYD. As conceived by CCNY, the three key design principles are interrelated. This chapter also addresses several design principles more closely related to school operations because they support high-quality instructional practices. The implementation analysis focuses on the final year of the ObD initiative. The case study that concludes this chapter describes how one school integrated several design principles much in the way that CCNY envisioned.

KEY TAKEAWAYS

Practices to Support Mastery-Based Learning

- In most schools, mastery-based learning was conceptualized as deep knowledge of content and skills.
- In practice, mastery-based learning entailed offering students multiple attempts to demonstrate mastery.

Practices to Support Personalized Learning

- In most schools, personalization entailed accommodating students' interests.
- Personalization did not typically involve other practices, such as providing students with extensive choices in content or materials.

Practices to Support PYD

- Skills were emphasized to support student academic achievement and positive behavior.
- Social and emotional competencies can help students meet mastery expectations but were not explicitly taught or assessed in most schools.

Perceived Enablers and Barriers

- Teachers were satisfied with many aspects of school operations and supports.
- Teachers perceived unclear administrator expectations and some student factors as obstacles to successful implementation.

This chapter addresses RQ 2: How did teachers and other school staff implement the key design principles, and what factors might have facilitated or hindered implementation?

We draw on the implementation data described in Chapter One to focus on the three design principles that are most closely related to school culture and instructional practices: prioritizing mastery, personalizing learning, and PYD. We also touch on several of the design principles more closely related to school operations (e.g., hiring, data use) because they support high-quality instructional practices. These results help contextualize the outcomes findings by providing detailed descriptions and examples of school-level practices, as well as enablers of and barriers to implementation. Some of the findings are relevant to more than one design principle; in those cases we discuss the finding in the section that seems most applicable. We conclude this section with a detailed description of one school's approach to integrating many of the design principles as envisioned by the initiative.

As we noted in Chapter One, the three key design principles often intersect: PYD can support effective personalization and mastery-based approaches. Even though our discussion treats them separately, we encourage readers to remember that mastery-based and personalized instructional practices, as conceived by CCNY, were intended to be mutually reinforcing, and that PYD, as the foundation of school culture, supported them. Box 1 displays CCNY's definitions of each principle. Our discussion in this chapter generally adapts the definitions used by school staff and students. Throughout this chapter, we present brief examples of selected instructional practices and supports, drawn from interview, focus group, and survey data, to provide concrete examples of key aspects of implementation.

When we discuss the interview data, we use such terms as *many* and *most* to refer to more than half of interview respondents in the applicable group (e.g., school leaders, teachers, or district staff) across schools and districts, and we use *several* or *some* to refer to fewer than half. We note instances where interview findings are applicable only in specific schools or districts. Percentages reported are based on survey results.

As we discussed in Chapter Two, the effects of ObD on student outcomes varied across districts and across outcomes within districts. The lack of a clear overall effect could reflect a variety of factors, including the limitations discussed in Chapter Two, as well as the varied ways in which educators in ObD schools implemented the design principles. We did not have the statistical power to use quantitative methods, such as regression analysis, to examine the extent to which differences in student outcomes might be associated with differences in implementation, so we visually inspected the student outcome results alongside the implementation findings. Because we did not find any meaningful relationships—which could be due in part to the small number of ObD schools and the lack of variability on some implementation measures—we have omitted those results from the discussion in this chapter and, instead, present more-general themes regarding implementation and the conditions that reportedly supported or hindered it. A more detailed discussion of this analysis can be found in the separate Technical Appendix.

Practices to Support Mastery-Based Learning

"Mastery is not only about hitting a standard, but teaching students to know whether they have mastered a standard and knowing what they need to do so, which requires individualized specific feedback."

—TEACHER

In half of the schools, mastery-based learning was defined as deep knowledge of content and skills. In practice, it entailed offering students multiple attempts to demonstrate mastery. Eight of the 16 ObD schools defined mastery as assessing whether a student has deeply understood a topic and can apply the knowledge in novel contexts. In contrast, staff and students in the other eight schools described mastery in terms of completion of tasks or assignments.

Across schools, the most common mastery-based practice was providing students multiple attempts

> *"As a school we do not have a strong understanding of what it really means to determine mastery.[Mastery is] more subjective teacher ratings of whether the teacher thinks the student is competent."*
>
> —TEACHER

to demonstrate mastery. Teachers favored this over other approaches, such as allowing students to work on different topics or skills at the same time or requiring that students demonstrate mastery before moving on to the next topic. Ninety percent of the surveyed teachers reported that students could make multiple attempts at a given task that counted toward mastery. Seventy-one percent of teachers reportedly emphasized identifying students' prior knowledge and skills when starting on a new topic or competency. In comparison, only about 40 percent of teachers reported emphasizing allowing students to work on different topics or skills at the same time or requiring students to demonstrate mastery before they moved on to a new topic.

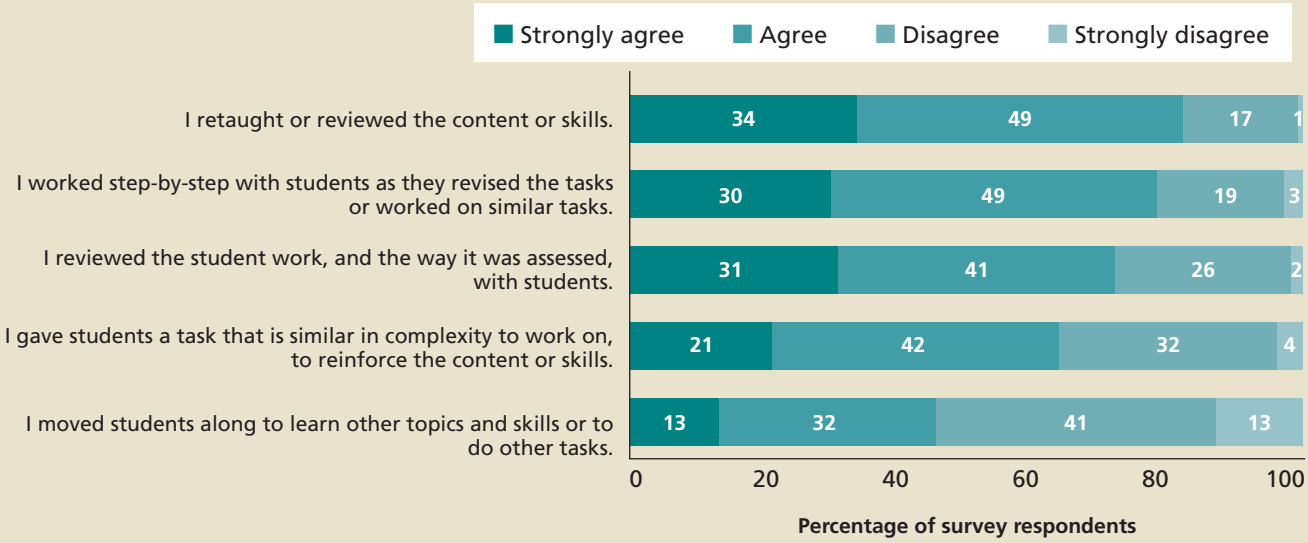
The ways in which teachers provided students with multiple attempts to demonstrate mastery varied across and within schools. Interviews of teachers, students, and leaders revealed three ways in which ObD teachers offered multiple attempts to demonstrate mastery. First, students had multiple opportunities to complete and turn in tasks to demonstrate mastery. Some schools that employed this approach established firm deadlines for turning in assignments (e.g., after three attempts), and others did not. Second, students had multiple opportunities to show improvement on assignments and could redo them if they were dissatisfied with their score or had not yet demonstrated mastery. Third, assignments and tasks were structured so that students would revisit the same competency several times during the year—a technique known as "spiraling." Spiraling provides students with multiple opportunities to demonstrate growth or attempt mastery by teaching the same skills and competencies in different contexts, using different techniques, throughout the year. At most schools, staff and students described more than one of these as central to their school's mastery approach.

Teachers reportedly offered numerous supports to help struggling students achieve mastery. Teachers in all ObD schools discussed the importance of providing supports to achieve mastery. When students did not perform well on an assignment or assessment, majorities of teachers surveyed reported that they always or often reviewed or retaught the content or skills, gave students a related task similar in complexity, and worked step-by-step with students to revise the task or work on a similar task, as shown in Figure 3.1.

Teachers' interview accounts were consistent with the survey reports. Most described offering the student another opportunity to attempt the task or revise their work, and few said that they provided a different task, or a task that was tailored to the student's learning level. In addition, many teachers we interviewed believed that supporting students to improve their performance on the same task was important for their ability to truly master it and improve over time. As one teacher said, "It's not mastery if they don't learn how to do it. So, that's it. I just keep working with them until [they achieve mastery]."

> *"The way that the program . . . is structured is that the students have many opportunities to both learn the material and then demonstrate that they understand it. It's hopefully building something that will stick."*
>
> —TEACHER

FIGURE 3.1 **Frequency of Supports Teachers Provided When Students Did Not Perform Well on a Task to Demonstrate Mastery**



NOTE: Survey question: 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.)? Responses were given on a four-point scale where "Never or hardly ever" = 1 and "Always or almost always" = 4. N = 177–179. Percentages may not sum to 100 because of rounding.

Majorities of teachers reported using student performance on assignments and tasks and conversations with students as data sources to inform instructional decisions. Personalized and mastery-based approaches call for teachers to use up-to-date data about individual students to tailor instructional strategies, content, pacing, assignments, and supports. Large majorities of ObD teachers reported using student achievement or mastery data to tailor instruction for individual students, as shown in Figure 3.2. Teachers' survey responses indicate that the student data they used was up-to-date: One-third of teachers reported receiving data at least weekly on which students needed extra assistance and which had achieved mastery, and about student performance on specific concepts or skills. Sixty to 70 percent of teachers reported receiving these data at least monthly.

> *"I gave kids a list of common misperceptions in grammar . . . things that they still consistently get wrong on their writing assignments, and . . . I've used the data that I saw, from their work, from their writing, and gathered it up and said this is a good list for them to do."*
>
> —TEACHER

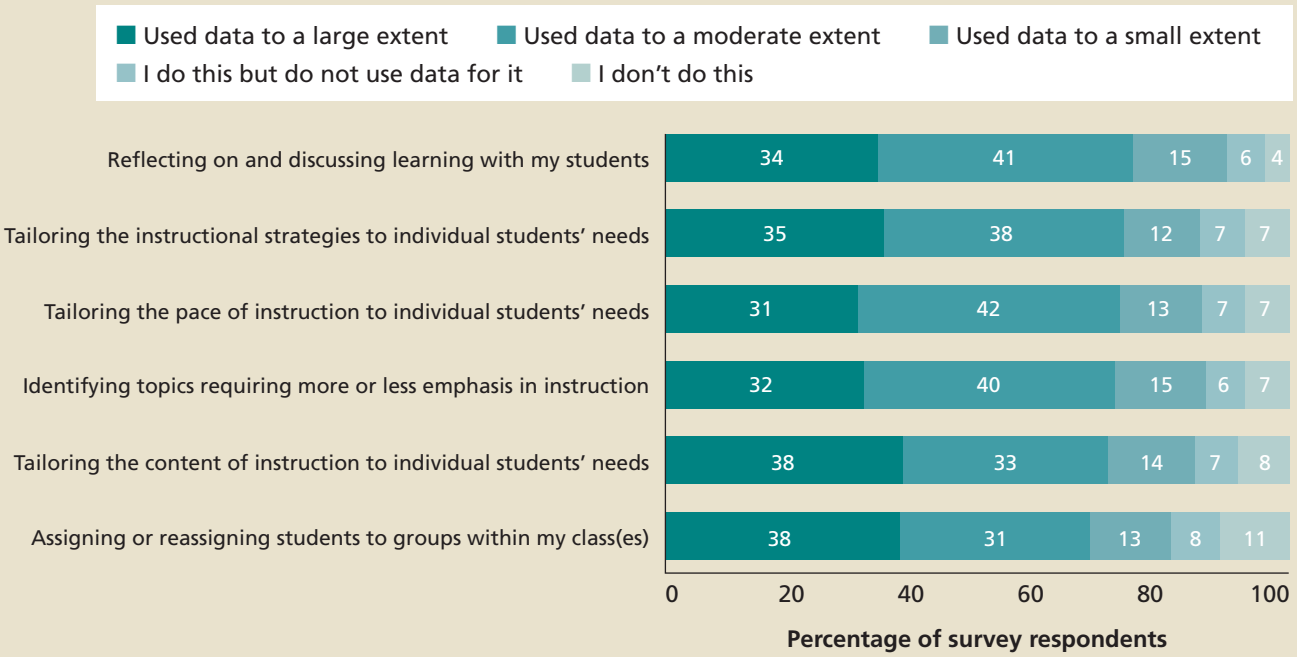
Teacher interviews revealed that their data sources were classroom assignments, tasks, and conversations with students rather than the learning management system or digital curriculum materials. Teachers reported using this information to identify which students were behind, did not understand the material, or needed more support to complete the assignment (see text box: *Inside ObD: How Do Teachers Support Students to Achieve Mastery?)*. Teachers reported relying on conversations with students to diagnose why a student did not understand the material.

Inside ObD: How Do Teachers Support Students to Achieve Mastery?

Most ObD teachers reported taking an active role in supporting students when they did not demonstrate mastery. The approaches described by teachers we interviewed fell into three categories: teacher-led support, student-led support, and a hybrid approach that combined teacher- and student-led support.

- A *teacher-led approach* (used by half of the teachers we interviewed) entailed working one-on-one or in small groups with students to reteach the content (e.g., walking through additional material or examples of completed assignments) and targeted feedback (e.g., feedback that specifically highlighted what the student needed to revise and how to revise it) to help students achieve mastery. However, few of these teachers used data from the students' first mastery attempt to inform subsequent teaching strategies.
- A *student-led approach* (used by less than one-third of teachers we interviewed) placed responsibility for incorporating teacher feedback and achieving mastery on the student. In this approach, teachers reported being less active in providing direct support (such as walking through an example of a completed assignment, as in the teacher-led approach). Instead, they pointed students to additional course materials or examples of completed assignments they could review on their own and encouraged students to revise based on feedback they had already provided.
- In a *hybrid approach* (used by about one-quarter of the teachers we interviewed), teachers used a combination of teacher-led and student-led strategies to provide support. These teachers said they provided some concrete guidance or specific resources to struggling students to help them revise their work and encouraged students to take the initiative to revise their work or make another attempt to achieve mastery.

FIGURE 3.2 **Extent to Which Teachers Reported Using Student Achievement or Mastery Data for Various Purposes**



NOTE: Survey question: 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." Responses were given on a five-point scale where "I don't do this" = 1 and "Used data to a large extent" = 5. N = 180–182. Percentages may not sum to 100 because of rounding.

Practices to Support Personalized Learning

In my classroom, I try to choose very relevant topics and present them in ways that kids care about. . . . I think I endeavor and the school as a whole endeavors to make things relevant to kids."

—TEACHER

In most schools, personalization entailed accommodating students' interests. Teachers and leaders in all schools said that personalizing learning meant that they tried to tailor their instructional practices, materials, and topics to accommodate students' interests. Seventy-six percent of surveyed teachers reported providing a variety of materials or instructional approaches to accommodate individual students' learning needs and interests. Many teachers reported making an effort to get to know their students personally and their students' interests and using that information to tailor their instruction and develop learning goals. Staff in seven schools mentioned that an advisory period, in which a teacher met with the same small group of students, was when they worked to get to know students and help them plan learning goals. Other teachers mentioned getting to know students through brief surveys and personal conversations.

Most teachers used a variety of learning materials and strategies and adapted course content to students' learning needs and interests; offering opportunities for students to make choices about their learning was less common. Large majorities of teachers emphasized most of these strategies, as shown in Figure 3.3. Teachers reported using a variety of learning materials and strategies and adapted course content to students' learning needs (see text box: *Inside ObD: How Do Teachers Personalize Assignments?*). Most teachers reported that their students were able to access learning materials inside and outside the classroom and keep track of their own progress. About half of teachers offered students opportunities to choose topics and instructional materials. Student survey responses were consistent with those of teachers—about half of students reported that they could choose instructional materials or topics.

Inside ObD: How Do Teachers Personalize Assignments?

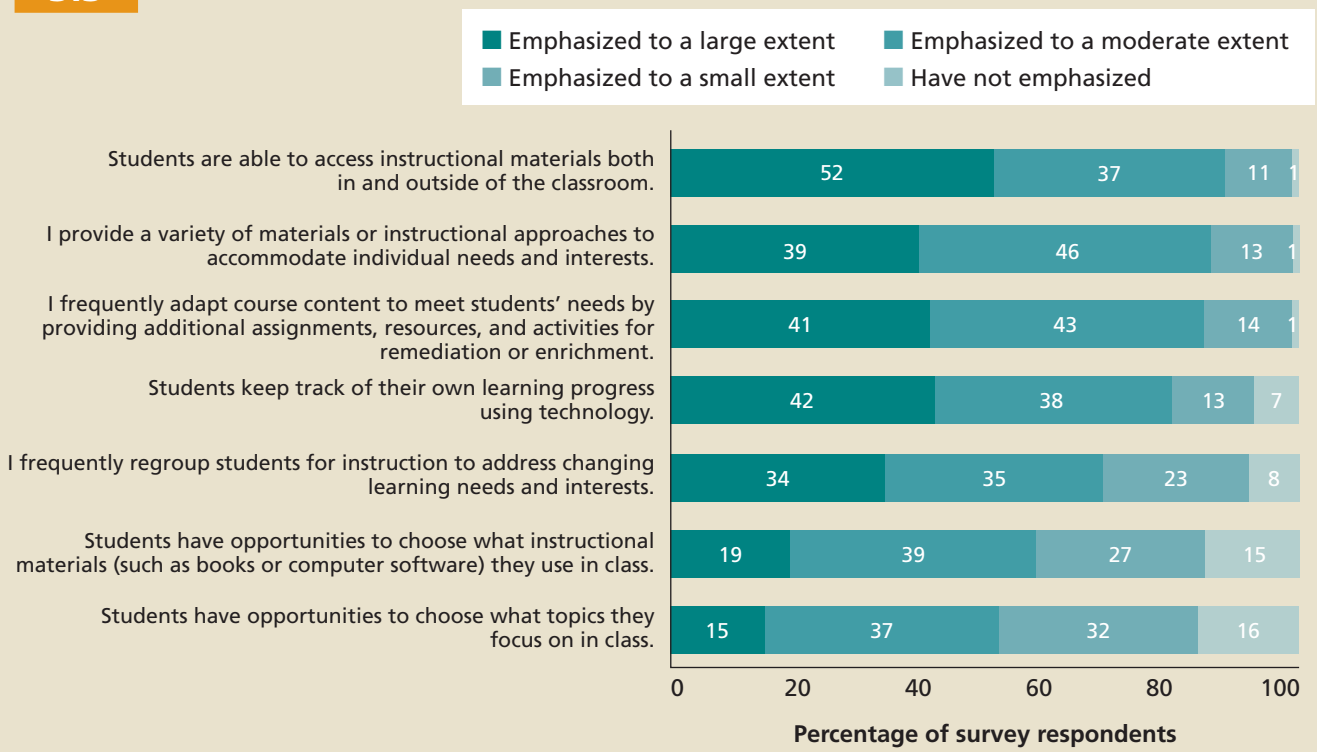
When ObD teachers did personalize assignments for students, they generally described two approaches. One involved offering a choice of topic or assignment presentation format. In these cases, students could choose the focus of the assignment—such as the topic of an essay—or the mode in which they presented their work (e.g., oral presentation, written report). Teachers who offered students a choice of topic or presentation format generally provided a limited set of three or four choices.

The other approach involved providing assignments at different levels of difficulty. Some teachers who provided assignments at varying levels of difficulty tailored assignments to student learning levels (e.g., students who struggled to master the material received a shorter version of the assignment), while others said they adjusted for difficulty when they assessed the work.

Teachers who offered students a choice of topic or presentation format generally provided a limited set of three or four choices. Most teachers tried to balance the importance of personalizing in these ways with the burden of developing and assessing many different versions of the assignment. Students in our focus groups agreed. Some teachers who did offer such a choice provided students with the assessment rubric at the beginning of the assignment so they could understand the criteria.

FIGURE 3.3

Extent to Which Teachers Reported Emphasizing Personalized Instructional Practices



NOTE: Survey question: 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. Responses were given on a four-point scale where "Have not emphasized" = 1 and "Emphasized to a large extent" = 4. N = 180–181. Percentages may not sum to 100 because of rounding.

Although the survey data indicate that several personalization strategies were common, interviews with teachers suggested that most assignments were not personalized. In general, all students received the same assignment. When personalization did occur, some teachers offered the assignment at different levels of difficulty. Others provided a limited set of choices of topic or mode of presentation (e.g., oral presentation, written report). Few teachers reported using both strategies. Tailoring the level of difficulty of an assignment and offering students the opportunity to direct their own learning by choosing the topic or mode of presentation are key aspects of personalization. But they can be time-consuming and therefore challenging for teachers to implement, particularly as schools scale up and refine their model and as teachers develop instructional materials (Pane et al., 2017; Steiner et al., 2017).

> *"There is flexibility within [the assignment], but I don't know how I could grade if they were all different. I do offer choice—for the e-magazine the written piece could be a letter to the editor, or an op-ed, or a research article, or a manifesto, or an opinion piece, or an interview. For the visual it could be memes, protest art, an infographic, a comic, a photo series."*
>
> —TEACHER

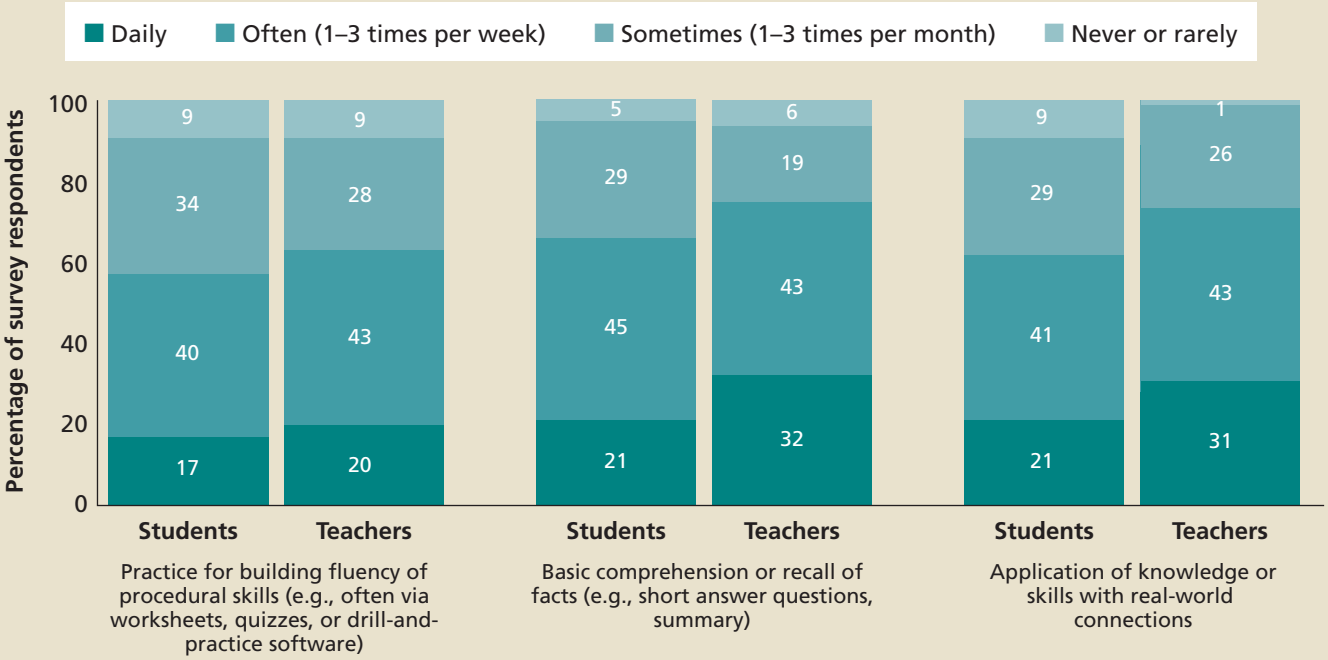
Students and teachers described tasks and assignments as varied, clear, and connected to the real world. As we described earlier, a majority of teachers used a variety of materials and assignments to personalize instruction to students' learning needs and interests. Students reported that teachers used a variety of types of tasks in their instruction—some repetitive tasks to provide students with opportunities to practice, some that focused on basic comprehension, some that connected course content to the real world, and some that required student collaboration over time. Fifty-seven to 66 percent of students reported

receiving each of these types of tasks one to three times per week. Students' perceptions of their tasks and assignments were positive. More than 80 percent of students agreed that task directions and purposes were clear, assessments were fair, and the feedback they received was helpful. Teachers' survey responses were consistent, as shown in Figure 3.4. Forty-eight to 75 percent of teachers reported that they gave each type of task at least once per week. In interviews, teachers' descriptions of actual assignments were consistent with survey responses.

Practices to Support Positive Youth Development

Most schools defined PYD as a way to support student academic achievement and positive behavior. In our interviews, we asked school staff how they defined PYD in their schools. The design principle states, "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; 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." In 12 schools, staff described PYD as a way

FIGURE 3.4 **Frequency with Which Students Reported Receiving and Teachers Reported Assigning Various Assignments and Tasks**



NOTE: Teacher survey question: How often do you give each of the following types of tasks to students to monitor or assess their learning? Responses were given on a four-point scale where "Never or hardly ever" = 1 and "Daily or almost daily" = 4. N = 180. Student survey question: 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? Responses were given on a four-point scale where "Never or hardly ever" = 1 and "Daily or almost daily" = 4. N = 2,642–2,659. Item wording differed slightly between surveys to accommodate student reading level. Percentages may not sum to 100 because of rounding.

"I'm trying really hard to get the kids to tell me more about what they want at the school. . . . We haven't developed everything [students] want yet."

—TEACHER

to foster high expectations for learning and behavior and student responsibility for meeting learning goals in their definitions of PYD. In these 12 schools, staff reported helping students build skills, such as persistence through challenging tasks, collaboration, or time management, that could help them meet their learning goals and improve their academic achievement. Staff in these 12 schools also described a focus on promoting positive behavior, connecting academic content to themes related to SEL, helping students develop a sense of agency in their education, creating a positive school culture, and encouraging teachers to build positive relationships with students to facilitate personalization. In three of the remaining schools,[4] staff described PYD exclusively as a mechanism for encouraging positive behavior. Staff in these schools described positive behavior intervention systems (e.g., offering rewards for desired behaviors) in their descriptions of PYD. Family participation and integration of community assets were not mentioned in our interviews.

PYD, mastery, and personalization practices were interconnected in most schools. The design principle of PYD includes students having "access to experiences and relationships that help them develop the skills and mindsets to succeed." As part of this conceptualization, CCNY intended that PYD, mastery, and personalization practices would be interconnected within a school. For example, PYD and mastery could be connected if students were taught skills that helped them persist through challenging tasks, which would help them revise and improve their work as they made multiple attempts to demonstrate mastery. PYD and personalization could be connected if teachers were encouraged to get to know each student individually and use this knowledge to develop lessons that focused on topics of interest to students.

We found that PYD, mastery, and personalization practices were interconnected in most schools. Students needed such skills as time-management, organization and planning, and self-regulation to meet flexible deadlines for mastery and respond to multiple opportunities for revision to demonstrate mastery. Staff in most schools described an approach to personalization that involved building positive relationships with students so they could get to know individual students' interests, context, and learning needs. More than 90 percent of teachers reported that they emphasized actively establishing relationships with their students and greeted them by name when they came to class. Staff described using this information to develop assignments and tasks that were relevant and interesting and to provide appropriate choices. In some schools, staff described getting to know students through informal conversations, surveys, and mentoring.

"I think the integration of youth development and personalization are important here. We give students a lot of opportunity to set their deadlines, to decide when they're going to get things done."

—PRINCIPAL

Social and emotional competencies can help students succeed in high school and beyond, but these skills were not explicitly taught or assessed in most schools. Although PYD practices in the ObD schools took many forms, many of the practices that teachers and principals described as aligned to CCNY's conceptualization of PYD are consistent with the increasingly prevalent emphasis on SEL, which can be defined as "the process through which children and adults understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions" (CASEL, undated). SEL can be addressed through explicit instructional programs but can also be integrated into academic instruction and fostered through the development of strong relationships between adults and youth (Aspen Institute National Commission on Social, Emotional, and Academic Development, 2019). As we described above, students needed to develop such competencies and dispositions as goal setting, collaboration, self-discipline, and communication to succeed in high

[4] We did not have enough information to describe PYD in the final school.

> *"We also do Habits of Success breakfasts or reward programs. So if a student is exhibiting one of the SEL competencies really well, we'll make it like a growth mindset ticket or something, and then at the end of the month, we invite you to a breakfast."*
>
> —TEACHER

school and meet schools' mastery expectations. These are examples of SEL competencies (CASEL, undated), and providing instruction on these skills would be consistent with ObD schools' commitment to PYD.

In about half of the ObD schools, staff said that the school had identified a set of SEL competencies to focus on (e.g., Summit Public Schools' Habits of Success [Summit Learning, 2017]), assessed students' development of specific SEL competencies, or used curricula in their advisory classes that incorporated SEL. Staff in these schools talked about the importance of

Inside ObD: How One School Integrated PYD, Mastery-Based, and Personalized Learning Practices

One school stood out for integrating PYD, mastery-based, and personalization practices in its instructional model. According to the principal, teachers, and students we interviewed, this school did several specific things:

- ***Espoused a well-rounded definition of PYD*** **that included a focus on building skills to support academic achievement and on promoting positive behavior. This school's definition of PYD emphasized building positive relationships, creating a positive school culture, and soliciting student feedback about the school model and operations.**
- ***Practiced what it preached*** **in the sense that staff and administrators emphasized social and emotional skills as a way to improve students' academic work and behavior.**
- ***Took an explicit approach to teaching students SEL skills*** **in a way that clearly stated the skill being taught and how to exhibit the skill. Teachers described lessons dedicated to social and emotional skills and provided concrete examples of those skills, along with specific feedback that was integrated with academic content.**
- ***Assessed students' social and emotional skills*** **alongside academic content in a clear, structured way through schoolwide use of a rubric accompanied by staff PD.**

> *"We try to teach our kids how to collaborate by breaking down these abstract terms into things the kids can actually do. So we break collaboration down into taking risks, and then we give them a list that's like, 'You can take risks by inviting people into the conversation. You can take risks by making eye contact. You can take risks by offering ideas when you're not even sure if they're correct.' . . . I've also shown videos of people collaborating and analyzed [the videos] and broken them down. Like, 'Oh, that person was confused and their partner could tell and asked them why they were confused, and then they worked together. Well that's a good way to collaborate.' We try to actually teach them how to do it."*
>
> —TEACHER

teaching the skills that students need to be successful in the school's mastery environment, and in life after high school more broadly—skills such as responsible decisionmaking, which supports behaviors like time management, responsibility, and organization (CASEL, undated). However, most staff in most of these schools described examples in which key SEL skills were taught implicitly. For example, most teachers described modeling skills like responsibility, collaboration, and communication without explicitly addressing them in their instruction. Other teachers said they discussed the importance of these skills and provided examples or rewarded students for exemplifying specific SEL competencies. In one school, staff described providing explicit instruction in SEL skills and assessing those skills as part of assessment of academic content (see text box: *Inside ObD: How One School Integrated PYD, Mastery-Based, and Personalized Learning Practices*). The extent to which schools were successfully teaching skills like communication, critical thinking, and collaboration has implications for how well prepared students will be to achieve their postsecondary goals. Across schools, leaders and teachers thought that these were the most important skills students need to be prepared for college and careers.

Staff in most schools worried that students would not be prepared for college because some expectations of mastery-based instruction were very different from college expectations. As we discussed earlier in this section, helping students develop the "skills and mindsets to succeed" (CCNY, 2017) is a central component of PYD. In most schools, teachers and administrators were concerned that mastery-based approaches were not preparing students for college because mastery-based instruction is so different from what students encounter in college. In college, for example, students rarely have flexible deadlines for assignments, and end-of-term exams cover all the content for the whole semester. In contrast, in most ObD schools, mastery-based instruction entailed multiple opportunities to achieve mastery, flexible deadlines for turning in assignments, and continuous revision. In addition, teachers in most schools expressed concern that the coursework was not challenging enough and expectations for student performance were too low to fully prepare students for college. Of course, some aspects of mastery-based instruction, such as the expectation that students keep track of their own progress in their courses and respond to teachers' feedback to improve their work, are aligned with college expectations and could, if implemented effectively, help students when they transition to college. Although ObD school staff did not mention these potential benefits of mastery-based instruction in our interviews, it is possible that students experienced the benefits of these practices.

Eleventh- and twelfth-graders reported greater exposure to PYD, mastery-based, and personalized instructional practices than younger students. One aspect of PYD is that students experience increasing opportunities to make decisions about their own learning. In theory, students in schools that have adopted mastery-based and personalized instructional practices should be able to make decisions about their learning. For example, they should be able to choose the assignments and subjects that interest them the most and take the time they need to successfully complete the work. Some schools that use mastery-based and personalized instructional approaches take a gradual release approach to student autonomy, in which students initially receive more support and structure in their learning, and these supports gradually decrease as students get older (Pane et al., 2017).

To explore whether the ObD schools were providing students with increasing opportunities to make decisions about their learning, we tested for differences in students' reported experiences with practices related to the three key design principles by grade level and controlled differences across schools (more information about this analysis can be found in Appendix B). Eleventh- and twelfth-graders were more likely to report experiencing practices related to all three of the key design principles. The interviews and focus groups suggest two possible explanations for this difference, although there could be others. In some schools, staff described using a gradual release approach that was designed to provide older students with more opportunities to make decisions about their own learning. Another possibility is

> *"[W]hen you go to college, you're not going to be able to work at your own pace. . . . I'm scared that they're going to go to college and be like, oh, well, I have time to work on this, and I have multiple attempts to get it right, and that's not going to be the case."*
>
> —TEACHER

that older students had more experience with PYD, mastery-based, and personalization practices and were therefore more aware of their use or more practiced at identifying them.

Perceived Enablers of and Barriers to Implementation

Teachers were satisfied with many aspects of school operations and supports. In surveys, large majorities of teachers reported satisfaction with their professional learning opportunities. More than 75 percent agreed that their professional learning opportunities met their needs, allowed them to try new things and receive feedback, addressed strategies for implementing personalized and mastery-based instructional approaches, helped them integrate SEL into academic instruction, and were aligned with what they do in the classroom. Across schools, teachers were also enthusiastic about their colleagues, with more than 80 percent reporting that teachers supported each other and were highly invested in student learning. Teachers were similarly enthusiastic about their administrators (see text box: *Inside ObD:*

Inside ObD: How Administrators Provided High-Quality Support

One school stood out for the high-quality support that administrators provided to teachers. At this Cohort I school, nearly 100 percent of teachers (compared with 65 percent to 85 percent in other schools) strongly agreed that administrators were highly supportive of teachers, were highly focused on student learning, and trusted teachers to make decisions about their own instruction. In interviews, teachers described a highly collaborative, supportive school culture that was focused on iteration and revision of instructional practices and the school model.

The school's approach to PD reflected the school's collaborative, feedback-driven culture. Instead of traditional PD "sessions," teachers participated in "inquiry groups" of four to six teachers, which met weekly to address a teacher-identified problem of practice over the course of a six-week cycle. The cycle included visiting classrooms, opening their own classrooms to observation, and presenting what they had learned to other teachers and administrators. The principal reported that this approach to PD was based on his belief that teachers learn better by working together than sitting and watching a presentation.

> *"If I incorporated multiple benchmarks into my instruction, I could assess students better and they're gonna be able to do harder tasks. I learned through visiting other classrooms and getting feedback from visits to my room that it's really important to continue revisiting vocabulary and content and to explicitly connect to the task why you're doing the learning you're doing . . . feedback happens naturally when we're doing that. You don't have to try to build conference time, you don't have to try to build revision time because if benchmarks are connected to one another, students are revising and they must have feedback. It was also a nice check for me as a teacher."*
>
> —TEACHER

How Administrators Provided High-Quality Support). Teachers also reported favorable opinions of their schools' data systems.

More than 70 percent agreed that the data systems were easy to use, produced the reports they needed, and allowed them to make good decisions. Although staff at most schools reported using more than one system (i.e., type of software, such as Google Classroom), and the specific data systems in use varied across schools, the teachers we interviewed generally described these systems as having a few common features. In particular, most teachers reported that their school's data system allowed them to record their school's mastery-based grades and provide feedback to students, and allowed students to access their assignments, feedback, and grades and keep track of their progress in the course.

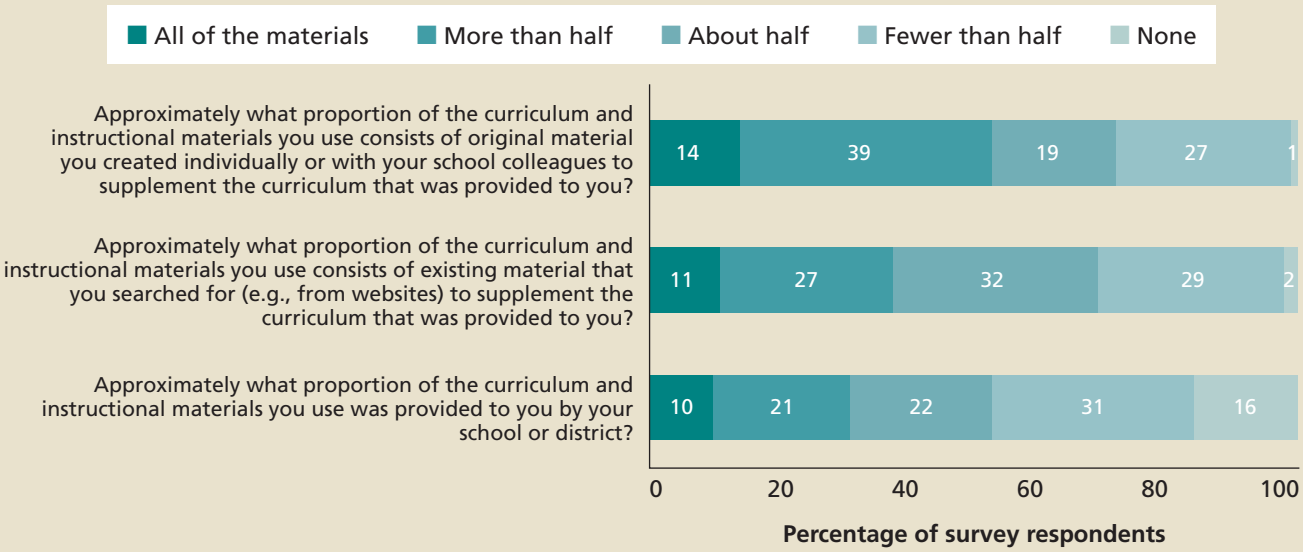
Teachers reported satisfaction with their curricula and spent several hours per week developing materials. In the survey, about three-quarters of teachers agreed that their curriculum materials were of high quality, addressed the learning needs of all their students, were easy to access, and supported mastery-based instruction. Survey data do not reveal which curriculum materials teachers had in mind when responding to the question. One possibility is that teachers were thinking of the materials that they created themselves. About half of teachers reported that they created more than half or all of their materials from scratch, about one-third said they searched for more than half or all of their materials online, and nearly half said that they received fewer than half of their curriculum materials from their districts (see Figure 3.5).

On average, teachers reported spending four and a half hours per week developing curriculum materials, but about half of teachers reported seven or more hours per week. Digital or online curriculum materials, such as Edgenuity, were reportedly only one source of instructional materials. In interviews, most teachers reported using digital curriculum materials as remediation tools or supplements to lessons they had developed. This is consistent with recent nationally representative survey results, which indicate that teachers nationally, in all grade levels, generally used digital materials as a supplement rather than as the main instructional materials (Tosh et al., 2020).

These findings raise the question of the extent to which teachers' PD focused on developing or adapting curriculum materials. In the prior report (Steiner et al., 2017), teachers' reports of their PD experiences were mixed but improved over time. Our data do not allow us to systematically describe ObD teachers'

FIGURE 3.5

Teachers' Reports of Sources of Their Curriculum Materials



NOTE: Responses were given on a five-point scale, where "None" = 1 and "All of the materials" = 5. N = 180. Percentages may not sum to 100 because of rounding.

PD experiences but indicate that in some schools, curriculum-related PD was provided by external organizations, such as Summit Learning and New Tech Network.

Unclear administrator expectations were a key obstacle to successful implementation. In half of the ObD schools, teachers described shifting or unclear expectations from school administrators as a barrier to successful implementation of mastery, personalization, and PYD. In five of these eight schools, teachers reported that loss of their founding principal was the cause. The ObD initiative supported founding principals to engage in a yearlong planning process, during which they would establish the school mission and design the school model, including the school's approach to mastery. Springpoint played a key support role during the design year and through at least the first two years of implementation (some principals received more than two years of implementation support after their schools opened; we discuss the role of Springpoint more extensively in Chapter Four). According to teachers, newly appointed school leaders did not always support sustaining the original design and, in several cases, made substantial changes to the design. Teachers in schools that experienced this challenge reported that their administrators—regardless of tenure—did not always provide clear or consistent expectations and policies for implementing mastery, personalization, and PYD.

Teachers perceived student factors to be barriers to successful implementation of mastery-based practices. Many teachers perceived that students lacked intrinsic motivation to learn, were not ready to be independent learners, and took advantage of the flexibility in the mastery-based system to do things like skip class and turn in assignments at the last minute. Additionally, many teachers perceived that the mastery approach did not allow them to enforce strict deadlines or hold students responsible for their academic behaviors. In addition, teachers from several schools reported that students struggled to adjust to the idea that their work needed to be revised until it reached a standard of mastery.

Most school leaders, however, believed that these student behaviors could change if teachers emphasized social and emotional skills in their instruction. Student survey reports also differed from teacher perceptions. A majority of students surveyed reported that they completed their assignments and submitted their work on time, with enough time to get feedback on how to revise. Eighty-six percent of students reported that they tried to do well on their schoolwork even when it was not interesting to them.

> *"[I]t's really hard to get teachers to think about what [teaching social and emotional skills] will look and sound like if they're doing this well. . . . So, the whole shift in PD this year was because of that."*
>
> —PRINCIPAL

Case Study: Integrating the Design Principles in Practice

In our analysis of the teacher survey data, one Cohort II school stood out for having the most positive responses on almost all topics. These positive survey responses were confirmed by our interviews. In this section, we highlight this school as an example of successful integration of the three key design principles and several other design principles in much the way CCNY initially envisioned and conclude with a discussion of supporting conditions that may have enabled successful implementation.

> *"I feel like PYD is one of the cornerstones of our school and makes us different and separates us from other high schools."*
>
> —TEACHER

This school integrated multiple design principles by choosing one principle—PYD—as the foundation for the school model. Staff and students at this school said PYD was the foundation of the school model and the principle around which school operations and instructional approaches were organized. According to staff and students, school culture was supportive and inclusive. Students reportedly felt that school staff cared about them, and one teacher described the school as being "like a family." PYD was evident in personalization when staff used their knowledge of individual students to provide assignments at varying levels of difficulty and personalized schedules based on social and emotional skills. PYD was linked with mastery; large majorities of surveyed teachers reported that they addressed social and emotional topics in their instruction. Most staff we interviewed described consistent use of a schoolwide rubric to assess students' social and emotional skills. Teachers reported assessing students on social and emotional skills (e.g., making good decisions frequently and in the context of assignments for demonstrating mastery), and students agreed. PYD informed the principle of empowering and supporting students for life after high school. According to students, teachers often connected social and emotional skills to preparation for college. College visits were informed by staff members' knowledge of individual students—staff took students to visit colleges with academic and social and emotional supports to ease the transition.

> *"So it's all based on positive youth development. We have built the social-emotional structure. I think it's key to our students being successful. I think that we're able to build a lot more on the academic experiences because we invest so much time in the social-emotional."*
>
> —PRINCIPAL

> *"Some teachers, you build . . . a teacher-student relationship. Sometimes they will talk to you. They will come and ask, 'Why you didn't do a good job, is something going on?'"*
>
> —STUDENT

The atmosphere at this school was collegial and supportive of staff and students. More than 90 percent of teachers reported that staff were collegial, were focused on improving student learning, and supported each other to improve student learning. A similar majority reported that school administrators trusted teachers to make their own decisions about instruction and were supportive of teachers. Large majorities of teachers reported that students were respectful of their peers and school staff and motivated to achieve. Students agreed that their teachers were focused on helping them succeed in high school and life after high school. Teachers at this school also reported positive perceptions of their PD opportunities, which focused on SEL in addition to mastery and personalization strategies.

Teachers at this school reported having the supports and resources they needed to use personalized practices. Teachers reported emphasizing PL practices and using student data to personalize instruction for students to a greater extent than teachers in other schools. Teachers at this school may also have had access to supports and resources that enabled this focus on PL, such as PD that addressed mastery-based and PL strategies.They reported greater satisfaction with their curriculum materials than teachers in other schools. They were also more likely than teachers in other schools to report that their curriculum materials were of high quality, addressed their students' learning needs, and supported mastery-based instruction. Teachers developed most of the curriculum with the support of an intermediary organization and received two periods per day to develop materials and plan with other teachers. They also perceived fewer obstacles to PL—such as lack of curriculum flexibility, inadequate data to personalize instruction, student absenteeism, and student disciplinary problems—than teachers in other ObD schools.

The school principal prioritized fit with school culture when hiring teachers. The principal said that a good fit with school culture was his priority when hiring teachers. Teachers' survey responses are consistent with this priority, citing the following factors more frequently than teachers in other ObD schools as important in their decision to take a position at the school: the opportunity to work with disadvantaged students, the school's fit with their own background and experience, interest in an environment that emphasizes PL, and opportunities to focus on youth development.

Three years of experience, a stable staff, and intermediary organization support may have contributed to successful integration of the design principles. It may take some time for ObD schools to become adept at implementing the design principles (Steiner et al., 2017). This school had three years of implementation experience by spring 2018. It benefited from stable leadership and teaching staff. Finally, its leader received extensive support from an intermediary organization (in addition to Springpoint) to develop curriculum materials, design and implement mastery-based instructional and assessment approaches, and design PD for teachers.

> *"It's [developing curriculum] something we're always sharing at our meetings, looking at curriculum we've created and how it connects to mastery. So it's not something we just do over the summer. It's something that's done all year round."*
>
> —TEACHER

CHAPTER FOUR

Local Contextual Conditions

Summary

This chapter describes local contextual factors that enabled and constrained implementation of the ObD school models. The discussion is focused on two areas: the support provided by Springpoint and other external partners and a set of enabling conditions, identified at the start of the initiative by CCNY, that could support comprehensive high school redesign. We describe the enabling conditions that remained most salient to ObD implementation, along with those that emerged according to the accounts of school and district staff.

KEY TAKEAWAYS

External Partner Support

- Springpoint played a unique role by providing school leaders with tailored support focused on the design principles.
- The time limit on Springpoint support posed a challenge to sustainability in schools and districts that experienced leadership turnover.
- Intermediary organizations and other external partners supplemented district and Springpoint support.

District Enabling Conditions

- Alignment of ObD school and district grading policies and the capacity of grading systems were key supports for mastery-based instruction in some schools.
- In some districts, principals perceived district financial instability and inflexible policies as threats to implementation.
- Autonomy from district curriculum and PD requirements were enablers of implementation but were not often accompanied by district supports related to curriculum or PD.

This chapter addresses RQ 3: What system-level conditions supported or hindered implementation?

As we describe in the introduction, CCNY designed the selection criteria for the ObD initiative to identify districts that exhibited a series of contextual conditions that could enable comprehensive high school redesign. As part of the program design, CCNY ensured that grantees would have expert external support through Springpoint and (in some districts) intermediary organizations and other external partners. However, school districts engaged in comprehensive reforms often experience tension between the need to provide schools with the autonomy and flexibility necessary for innovation and constraints (e.g., accountability requirements) present in the district context. In this chapter, we discuss the inputs of the ObD initiative, focusing first on support from Springpoint and other external partners and then on district contextual conditions.

External Partner Support

Springpoint provided extensive supports for school design and implementation. All ObD schools received this support during the design year and the first two years of operation, per the design of the initiative. Some schools chose to receive additional support from Springpoint during the third or fourth years of implementation, and some schools received support—such as access to PD for teachers and curriculum materials—from intermediary organizations and other external partners. In this section, we present findings related to the support provided by these external partners, with an emphasis on Springpoint because it played a key role in advising school leaders on design and implementation.

Springpoint played a unique role in the initiative by providing school leaders with tailored support focused on the design principles. Springpoint was established at the beginning of the initiative as a dedicated support to ObD school leaders, a feature not common to most comprehensive school-reform efforts. Springpoint staff reported tailoring support to the phase of school development (i.e., design year, first or second year of implementation), student population, and school leader needs. They said the design year support focused on helping school leaders integrate the design principles into their school models and plan for the launch of the school. For example, principals reported working with Springpoint to design teacher PD plans that focused on personalized and mastery-based instruction, PYD, and mastery grading approaches. Some principals worked with Springpoint to design staff recruitment and hiring strategies.

During the first and second years of a school's operation, Springpoint focused on working with school leaders to address implementation challenges. Springpoint staff visited each ObD school to conduct observations and provide targeted coaching to school leaders. These visits often included district representatives to ensure that school leaders did not receive conflicting feedback. In addition, Springpoint collected and disseminated resources (e.g., research articles, case studies, curriculum materials) related to CCNY's design principles, according to each school leader's needs. Springpoint also connected school leaders with other external partners to fill gaps in support from district sources; we discuss this further later.

Springpoint was a key source of support to school leaders throughout the initiative. In each year of the study, a large majority of ObD school leaders reported that the support they received from Springpoint was very helpful, even after the formal relationship had ended. Nearly all school leaders across years and cohorts reported that Springpoint provided valuable support. They particularly appreciated the coaching and feedback Springpoint provided during school visits and the opportunity to connect with and learn from leaders at other ObD schools. Some school leaders appreciated how Springpoint helped them look ahead and plan for upcoming implementation needs and challenges, and they shared Springpoint's recommendations with teachers through PD. Several school leaders in Cohort I and II schools reported that they remained engaged in with Springpoint through hosting study tours[5] and coaching Cohort III and IV school leaders after their formal relationship had ended. Many school leaders also valued Springpoint's

[5] Study tours were "visits to innovative schools that were successfully implementing models aligned with key design principles. Study tours included classroom observations; conversations with teachers, students, and school leaders; and sharing of materials (e.g., mastery rubrics, student handbooks, lesson plans) so that staff from the visiting schools could understand the host school's design and, if they wished, incorporate attractive features into their own school models. After the first year of the initiative, study tours included other ObD schools to facilitate connections among the ObD schools so they could learn from each other" (Steiner et al., 2017, p. 3.).

> *"The most common way we differentiate support to schools is by developmental stage. . . . Another big driver of differentiation is student population. . . . A third differentiation factor is school team capacity. If I have a strong leader with good instructional background I don't need to spend as much time coaching in this area."*
>
> —SPRINGPOINT STAFFER

> *"Springpoint, what they say is on point. Wednesday, tomorrow . . . I'll be bringing teachers [Springpoint's] feedback. I'll be showing the teachers, these are our strengths, these are our recommendations."*
>
> —PRINCIPAL

role in connecting them to external sources of supports, such as consultants or service providers who could provide support in areas of need.

Springpoint refined its approach to supporting ObD school leaders over the course of the initiative. In keeping with the design principles, Springpoint's approach to supporting ObD schools emphasized continuous improvement in two ways—by refining its own strategies for supporting the ObD schools, and by encouraging school leaders to use a continuous improvement approach as they implemented their school models. For each school leader, Springpoint staff used data from school visits (e.g., observations, conversations with other stakeholders within the school and district) and monthly coaching calls with school leaders to inform a customized program of support around a standardized set of topics.

Over time, Springpoint's approach shifted from focusing on areas the leaders identified for themselves to a more standardized set of topics and strategies based on challenges that were common across the ObD schools. For example, in years three and four of the initiative, Springpoint staff devised a data-driven approach for targeting areas for improvement that the leaders might not recognize on their own. For example, one Springpoint leader described helping a principal spend more time providing instructional feedback to teachers by observing the principal during the school day and tracking use of time. This exercise reportedly helped the principal identify tasks to delegate and perform related tasks in sequence (e.g., perform office work all at one time during the day)—and use the remaining time to work with teachers.

The time limit on Springpoint support posed a challenge to sustainability in schools that experienced turnover of principals or key district staff. CCNY designed the initiative to provide schools with three years of Springpoint support—the design year and the first two years of implementation. Thus, Springpoint focused on building the capacity of the founding principal and teachers and select district staff. After the second year of the school's implementation, Springpoint often did not have the resources to provide the same level of support to new principals or district staff. And in some cases, principals and district staff opted not to receive support beyond the first two years of implementation (although some school leaders opted to access this support longer). Several principals and key district administrative staff who entered their positions in the third year of implementation or later reported that they did not have the benefit of receiving targeted coaching from Springpoint because the relationship with Springpoint ended before they replaced their predecessors. Principals in several districts expressed frustration that, due to turnover, the leaders in their district had a limited understanding of the ObD design principles—particularly mastery and continuous improvement—and lacked Springpoint support to help them become more familiar with them. According to these principals, the combination of these conditions often limited the relevance and usefulness of district supports.

Intermediary organizations and other external partners supplemented district and Springpoint support. Three ObD schools received support from intermediary organizations, and several others partnered with other external organizations for training or other services.[6] One school leader described the coach provided by an intermediary organization as a consistent, supportive partner who helped

[6] Intermediary organizations mediate between two educational organizations. In the context of ObD, intermediary organizations mediated between the ObD schools and the school districts. These organizations consisted of networks of innovative schools that provided PD and support that aligned with the design principles. External organizations included curriculum providers, PD providers, and other support organizations and vendors.

> *"[The coach from my intermediary organization] has been the most consistent support. She's the person who gets the ObD principles. . . . There's never a time where I have to stop and give four months of context about why we're doing something. She's in the moment with us."*
>
> —PRINCIPAL

develop key features of the school's design and kept staff focused on the design principles in the midst of changing district priorities. Many of the external organizations were recommended by Springpoint to supplement Springpoint and district support. For example, several ObD schools adopted the Summit Learning online learning management system and partnered with Summit to participate in training on how to use the system to facilitate mastery-based learning (Summit Learning, undated). In the last two years of the initiative, two ObD schools enlisted external organizations for assistance with developing curriculum materials and providing teachers with PD in mastery-based assessment.

District Contextual Conditions

CCNY drew on school-reform implementation research to compile a set of district-level conditions that were likely to enable ObD implementation. These were district leader support for ObD, financial stability, a commitment to a portfolio strategy, a commitment to innovation, technology quantity and quality, and an emphasis on ambitious college- and career-ready standards (CCNY, 2014). In this section, we describe the enabling conditions that remained most salient to ObD implementation, along with those that emerged according to the accounts of school and district staff.

District support for innovation promoted collaboration and cross-school learning opportunities. Principals described varying levels of district support for ObD schools and for developing innovative practices as part of the school models. Principals who described stronger district-level support had strong district-wide networks of innovative schools they could turn to for support. They also described clear intentions on the part of district staff to learn from the ObD schools. For instance, principals and central office staff in two districts mentioned several networks of innovative schools and district-wide mechanisms (e.g., communities of practice, network-specific meetings) that allowed the leaders of these innovative schools to learn from one another and share insights with district leaders.

Alignment of ObD school and district grading policies and the capacity of grading systems emerged as key factors in implementing mastery-based learning. Across districts, ObD teachers and principals described district data systems as important facilitators of (or, in some cases, barriers to) implementing mastery-based learning. As we noted in Chapter Three, school and district grading policies were often misaligned, which hindered successful implementation of mastery at the school level. Specifically, principals and teachers in several ObD schools expressed frustration that their school's mastery-based student data system did not interface with the district's data system. In these schools, teachers reported that converting mastery-based grades to traditional letter grades created extra work for them and confusion for students and families—and was contradictory to the mastery-based approach they were trying to implement. The challenge created by misalignment between mastery-based grading and external systems (e.g., state-level reporting and college admissions) is not unique to ObD (Pane et al., 2017; Steele et al., 2014). However, in one district, district leaders described an effort to find or create a data reporting system that would integrate mastery-based and non–mastery-based grading to support ObD schools' use of such approaches.

In some districts, principals said they perceived district financial instability and limited district-level support as a threat to ObD schools' ability to plan and implement the design principles. ObD principals from two districts reported that they believed the sustainability of their schools was threatened due to district financial instability and limited support from district leadership. According to these principals, the uncertainty about future district funding was a barrier to continuous improvement planning. One principal described that the only way to ensure future funding was to demonstrate the efficacy of the school model by improving student scores on high-stakes tests, which might not be sensitive to the value of the ObD model. This principal said, "I think the challenge is that we've been asked to do something (through ObD) that doesn't currently produce super valuable test results." Another principal expressed the belief that advocacy was required to secure district support for the school and that advocacy for the model might become increasingly difficult once the grant ended.

Autonomy from district curriculum and PD requirements was an enabler of implementation but was not often accompanied by district support. All of the ObD schools had the autonomy to adopt or develop curricula that were aligned to their school model and supported mastery-based instruction and assessment and personalized instructional strategies. As discussed in Chapter Three, staff at most schools used a blend of materials—some they developed themselves; some were provided by external partners (e.g., Summit or New Tech Network), and others provided by the district; and some were off-the-shelf (e.g., Edgenuity, Khan Academy). The ObD schools were given similar flexibility for PD requirements and were able to seek tailored support from external partners (e.g., ReDesign, NewTechNetwork) or to develop their own PD. This autonomy and flexibility were key enablers of implementation because they allowed the ObD schools to select materials and PD that were tailored to their school models.

Although teachers generally reported being satisfied with their curriculum materials and PD opportunities (as we discussed in Chapter Three), respondents to our survey reported spending about four and a half hours per week developing curriculum materials. A majority (71 percent) reported that the excessive amount of time they spent developing curriculum materials was an obstacle to implementing PL in their school. Teachers and principals in most ObD schools reported that they rarely received district support for selecting and/or designing curriculum and PD. Most district staff characterized the decision to provide the ObD schools with autonomy to develop or select their own curriculum materials and PD as a support, but many teachers and principals said they desired district support to assess the quality of prepackaged curriculum materials and help them design PD focused on mastery. Staff in the three ObD schools that worked with intermediary organizations reported that their intermediary partners provided this support, which they found helpful in the absence of district support.

> *"I would say the district is leaving selecting and developing curriculum materials to the school leaders. . . . [T]here's no curriculum being imposed, but at the same time, I don't have any evidence of collaboration taking place to help either school select what their curriculum will be. In both schools, teachers create curriculum."*
>
> —DISTRICT LEADER

Limited flexibility in district policies presented a barrier to ObD implementation. Most principals reported that they perceived many of their districts' human capital policies to be constraints. For example, principals in several districts expressed concern that they would be unable to find teachers who were a good fit for their school models because the district required them to prioritize hiring teachers from within the district. In addition, many principals said they perceived other district policies, such as setting inflexible school schedules months before the beginning of the school year and central approval of school grading systems, as barriers to innovation because they limited principals' ability to make changes to the school model. Although some districts granted the ObD schools a temporary reprieve from such policies in the early years of the initiative, in general, principals said they perceived this to be insufficient. For example, in one district, ObD schools were exempt from the requirement to prioritize hiring teachers from within the district in the first year of the initiative. The principals found this helpful in the first year, but reported that the requirement constrained their ability to hire their preferred staff in later years.

CHAPTER FIVE

ObD in a National Context

Summary

This chapter compares ObD teachers' survey results with those from a nationally representative sample of high school teachers. The results suggest that the typical student experience in an ObD school generally looks different from that in high schools nationally in terms of supports for mastery-based learning, personalization of learning, and PYD. Although these summary findings mask some variability in practices and supports within both the ObD and the ATP groups, on average, we observed fairly consistent and sometimes sizable differences. Although we cannot make any claims regarding the reasons for these differences, it is possible that they stem, at least in part, from the resources and guidance provided to ObD teachers as part of the initiative.

KEY TAKEAWAYS

Compared with the national sample, ObD teachers reported

- emphasizing mastery-based and personalized instructional practices and PYD in their instruction to a greater extent
- more extensive use of student data to inform mastery-based instruction
- using a variety of approaches to support struggling students more frequently
- more-positive opinions about data access and data system quality
- spending more time developing curriculum materials.

This chapter addresses RQ 4: How did ObD teachers' practices and perceptions of implementation enablers and challenges compare with those of a nationally representative sample of high school teachers?

The findings presented in Chapter Three indicate that teachers in ObD schools engaged in a variety of practices related to mastery, personalization, and PYD. What is not clear from those findings is the extent to which these practices differ from what was happening in high schools across the United States at the same time. In this chapter, we provide data from a nationally representative sample of high school teachers to explore whether and how practices in the ObD schools differed from those of the typical high school.

Using RAND's ATP,[7] we administered the ObD teacher survey to a nationally representative sample of 1,008 high school teachers in spring 2018. This was year 4 of the initiative, the focal year of this report. We sampled teachers of core academic subjects (mathematics, ELA, social studies, and science) who taught at least one high school grade (9–12). We gathered these data at roughly the same time as we collected the survey data from ObD teachers. The national survey omitted a few questions that were specific to ObD but was otherwise identical. We compared the responses of the national sample with those of the ObD teachers to understand the ways in which the reported practices and contextual conditions might differ between the two groups of teachers. These comparisons are purely descriptive; any differences might be attributable to factors other than the ObD initiative. We examined these comparisons primarily to understand the national context and to identify ways in which ObD teachers' experiences might be different from what is typical in high schools across the United States. The ATP provided a unique opportunity to gather data to conduct that analysis.

In this chapter, we summarize key findings from these comparisons. We conducted tests of significance on differences between the item means rather than on the categorical results shown throughout this chapter to better capture differences throughout the full distribution of responses. We first present data on reported practices related to the three key design

[7] Information about the ATP is available on RAND's Education and Labor webpage (RAND Education and Labor, 2020).

principles discussed in previous chapters. We then discuss the few areas in which we observed differences in perceptions of contextual conditions. Additional details about the sample and analytic approach, as well as the full results of questions presented in this report, are available in the separate Technical Appendixes. A separate report (Steiner, Doss, and Hamilton, 2020) presents more-comprehensive analyses of results from the ATP sample.

Practices to Support Mastery-Based Learning, Personalization of Learning, and Positive Youth Development

Our comparison of reported practices of ObD teachers with those of the national sample revealed numerous differences between the two groups, particularly in their approaches to the three key design principles. The differences varied in magnitude but generally showed ObD teachers adopting practices to support mastery, personalization, and PYD to a greater extent than the national sample. We discuss these findings next.

Mastery-Based Learning

ObD teachers consistently reported a greater emphasis on mastery-based instructional practices than high school teachers nationally. Teachers in ObD schools were more likely to report emphasizing many mastery-based practices to a moderate or large extent, as shown in Table 5.1. We observed differences of 20 percentage points or more for three key mastery practices—requiring students to demonstrate mastery before moving on, allowing different students to work on different topics or skills at the same time, and allowing students to work through material at different rates. Although all of these practices were relatively common in the ATP sample, these differences suggest that ObD teachers were using many mastery-based practices to a greater extent than was typical nationally.

We also asked teachers to indicate how various mastery-based assessment practices and ways students engaged with instructional tasks resembled their own practices. The percentages of teachers who reported that these practices resembled their own were similar (within 10 percentage points) for most items, as shown

Table 5.1. Percentage of Teachers Who Reported Emphasizing Mastery-Based Practices in Their Instruction to a Moderate or Large Extent

Mastery-Based Practice	Emphasized to a Moderate or Large Extent	
	ATP (%)	ObD (%)
When students are working on an assignment or activity, they know what the goals of the assignment or activity are.	86	90
Students have opportunities to review or practice new material until they fully understand it.**	68	84
I give students the chance to work through instructional material at a faster or slower pace than other students in this class.**	60	82
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.**	66	77
Different students work on different topics or skills at the same time.**	43	76
I require students to demonstrate mastery of a topic before they can move onto a new topic.**	42	62

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

ATP results are weighted to be nationally representative.

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, Emphasized to a small extent, Emphasized to a moderate extent, Emphasized to a large extent. ObD N = 180–181; ATP N = 998–999.

in Table 5.2. However, ObD teachers reported that three of these practices resembled their own at a rate more than 20 percentage points higher than teachers in the ATP sample: providing students with mastery tasks at the beginning of the unit, providing mastery tasks that differed in difficulty depending on the student's ability level, and allowing students to make multiple attempts at a given task that counts toward mastery to a greater extent than teachers nationally.

ObD teachers reported more extensive use of student data to inform mastery-based instruction than those in the national sample. Data use is a crucial component of mastery-based instruction, and ObD teachers reported more extensive use of student achievement or mastery data to inform their instruction compared with high school teachers nationally (Table 5.3). Differences for each of the purposes ranged from 11 to 30 percentage points, indicating that ObD teachers were much more likely than other teachers to report using data to inform their instruction to a moderate or large extent.

Teachers in both samples reported a variety of approaches to supporting struggling students, with ObD teachers indicating more frequent use of some of these approaches. A final question about mastery asked teachers how often they engaged in several different approaches to supporting students who did not achieve mastery (Table 5.4). ObD and ATP teachers reported using several of these approaches at similar rates, but ObD teachers were more likely to engage in some of them, including reteaching or reviewing the content or skills, placing students in groups to receive support, and working step-by-step with students as they revised the task or worked on similar tasks.

Table 5.2. Percentage of Teachers Who Reported Mastery-Based Instructional Practices That Resemble Their Own to a Moderate or Large Extent

Mastery-Based Instructional Practice	Resemble Own Practices to a Moderate or Great Extent	
	ATP (%)	ObD (%)
Students have multiple opportunities, throughout a unit or throughout the year, to demonstrate mastery of certain content and skills.**	90	97
Students can make multiple attempts at a given task that counts toward mastery.**	64	92
Tasks that are assessed for mastery require students to apply knowledge and skills they have learned to a new problem or context.*	87	89
Tasks that are assessed for mastery closely resemble tasks that students have already seen or experienced.	85	87
Students are considered to have achieved mastery only when they have demonstrated the knowledge or skills consistently.	83	84
Mastery is assessed as what students can accomplish mostly independently.*	85	82
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.**	56	76
Students attempt a task that is assessed for mastery when I believe they have a good chance at success on it.	80	74
Tasks that are assessed for mastery of a given competency differ in difficulty, depending on the student's ability level.**	50	74
When starting on a new topic or competency, I first identify students' prior knowledge and skills with a diagnostic assessment or task.*	63	71
Student work counts toward mastery only when it reaches an adequate performance level.	69	70
Students are considered to have achieved mastery when they have completed the expected number of tasks.	52	55

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

*** indicates statistically significant difference in item means at alpha < 0.01.*

ATP results are weighted to be nationally representative.

Question text: Please rate the extent to which each of the following descriptions resemble your instructional practices. Response options: Not at all, To a slight extent, To a moderate extent, To a great extent. ObD N = 163–166; ATP N = 810–811.

Table 5.3. Percentage of Teachers Who Reported Using Student Mastery or Achievement Data for Instructional Purposes to a Moderate or Large Extent

Instructional Purpose	Used Data to a Moderate or Large Extent	
	ATP (%)	ObD (%)
Reflecting on and discussing learning with my students**	49	76
Tailoring the instructional strategies to individual students' needs**	47	73
Tailoring the pace of instruction to individual students' needs**	43	73
Identifying topics requiring more or less emphasis in instruction**	51	72
Tailoring the content of instruction to individual students' needs**	43	71
Assigning or reassigning students to groups within my class(es)**	42	69
Developing recommendations for tutoring or other educational support services for particular students**	43	57
Recommending students for extended learning opportunities**	26	49
Providing college/career advice or guidance**	31	45
Allowing students to skip units or lessons if they've demonstrated mastery of the content in some other way**	18	40
Allowing students to skip courses or grades if they've demonstrated mastery of the content in some other way**	12	23

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

ATP results are weighted to be nationally representative.

Question text: This school year (2017–18), 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.

Table 5.4. Percentage of Teachers Who Reported Using Approaches to Support Students Who Struggled to Achieve Mastery

Approach	Used Approaches Often or Always/Almost Always	
	ATP (%)	ObD (%)
I retaught or reviewed the content or skills.*	71	82
I worked step-by-step with students as they revised the task or worked on similar tasks.**	65	79
I provided students with samples or models of finished work for the task they were assigned.**	64	75
I reviewed the student work, and the way it was assessed, with students.	71	72
I pointed students to materials or resources to practice or review on their own.**	59	69
I placed students in groups to receive support with learning the content or skills.**	53	66
I gave students a task that is similar in complexity to work on, to reinforce the content or skills.	62	64
I gave students a simpler task that covers the same content or skill.**	38	51
I moved students along to learn other topics and skills or to do other tasks.	44	46

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

***indicates statistically significant difference in item means at alpha < 0.01.*

ATP results are weighted to be nationally representative.

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 5.5. Percentage of Teachers Who Reported Emphasizing Personalized Instructional Practices to a Moderate or Large Extent

Personalized Instructional Practice	Emphasized Practice to a Moderate or Large Extent	
	ATP (%)	ObD (%)
Students are able to access instructional materials both in and outside of the classroom.**	81	89
I provide a variety of materials or instructional approaches to accommodate individual needs and interests.**	76	86
I frequently adapt course content to meet students' needs by providing additional assignments, resources, and activities for remediation or enrichment.**	63	84
Students keep track of their own learning progress using technology (for example, by using an online gradebook or portfolio).**	62	80
I frequently regroup students for instruction to address changing learning needs and interests.**	52	69
Students have opportunities to choose what instructional materials (such as books or computer software) they use in class.**	36	58
Students have opportunities to choose what topics they focus on in class.**	28	52

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

***indicates statistically significant difference in item means at alpha < 0.01.*

ATP results are weighted to be nationally representative.

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; Emphasized to a small extent; Emphasized to a moderate extent; Emphasized to a large extent. ObD N = 180–181; ATP N = 998–999.

Personalization of Learning

ObD teachers' reported more extensive use of practices to personalize instruction than teachers in the ATP sample. Teachers in both the ObD and national samples reported extensive use of several practices for personalized instruction, though ObD teachers' reported use was higher. As Table 5.5 shows, with one exception, we found differences of 10 percentage points or more in teachers' reports of emphasizing each practice listed in Table 5.5 to a moderate or large extent. Differences in the extent to which teachers offered students choice in topics or materials were especially noteworthy. Even though ObD teachers reported emphasizing student choice less than they did other practices related to personalization, they did support choice to a significantly greater degree than their counterparts in other high schools. The exception was enabling students' access to instructional materials both inside and outside the classroom; more than 80 percent of teachers in both groups indicated doing this to a moderate or large extent.

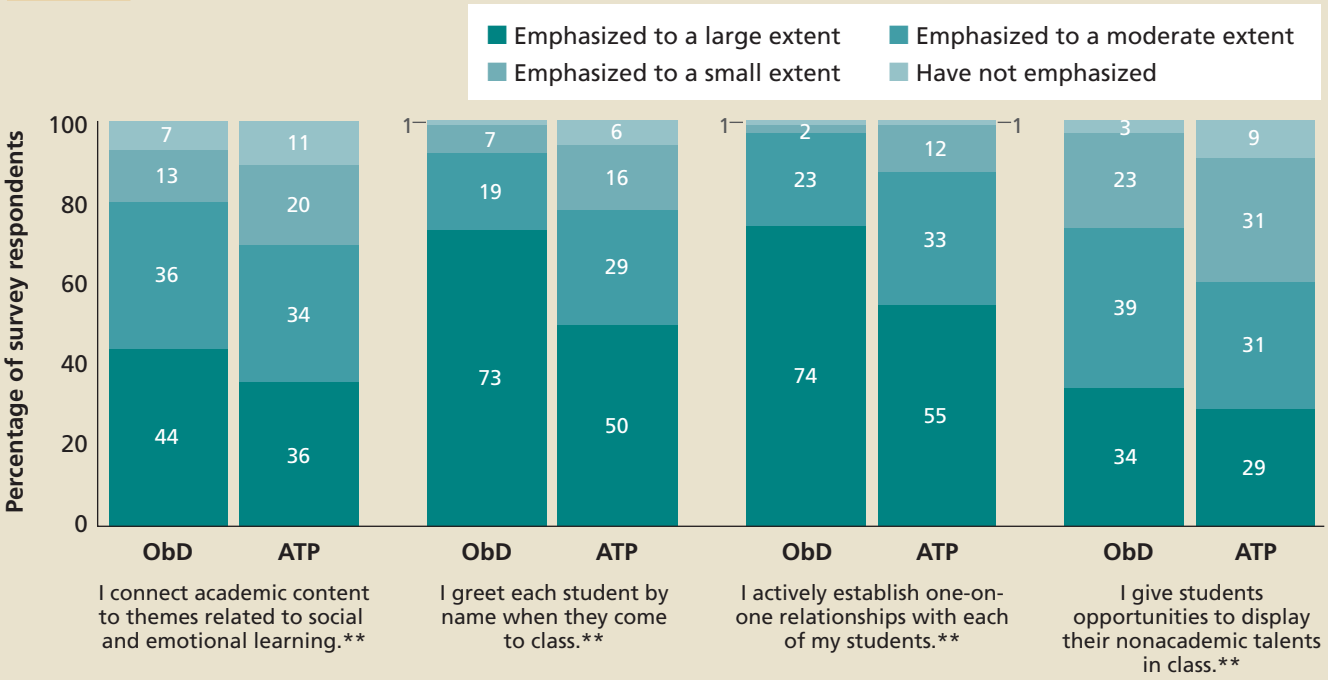
Positive Youth Development

ObD teachers reported emphasizing PYD to a greater extent than teachers in the national sample. The final set of practices that we examine in this chapter relates to support for PYD. As we discuss in Chapter Three, many PYD practices are consistent with an emphasis on SEL.[8] We observed patterns similar to those for mastery-based and personalized practices: ObD teachers reported greater emphasis on practices to support PYD than did teachers in the national sample. As shown in Figure 5.1, majorities of teachers in both groups reported engaging in a variety of practices to support PYD, such as emphasizing establishing one-on-one relationships and connecting academic content to themes related to SEL, to a large extent, but ObD teachers were more likely to emphasize these practices to a large or moderate extent.

An important component of incorporating PYD/SEL into academic instruction is addressing such competencies as collaboration, emotion management, and responsible decisionmaking in instruction. Table 5.6

[8] SEL can be defined as "the process through which children and adults understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions" (CASEL, undated).

FIGURE 5.1 **Percentage of Teachers Who Reported Emphasizing PYD in Their Instruction to a Moderate or Large Extent**



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

ATP results are weighted to be nationally representative.

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; Emphasized to a small extent; Emphasized to a moderate extent; Emphasized to a large extent. ObD N = 180–181; ATP N = 998.

shows the percentages of ObD teachers and high school teachers nationally who reported that they emphasized these practices in their instruction. We report the percentage of teachers who responded "to a large extent" only because similar proportions of teachers in the two groups responded "to a moderate extent" on these items. Consistent with the more extensive reported emphasis on PYD practices in Figure 5.1, we found that ObD teachers reported emphasizing almost all of these competencies to a large extent significantly more often than teachers nationally. The two exceptions were "handling stress" and "developing a sense of identity," the two items likely to be emphasized to a large extent in both groups. The implementation findings provide some insight about the ways in which ObD teachers emphasized PYD/SEL practices in their instruction. In most schools, such competencies as collaborating with other students and establishing positive relationships were not taught explicitly. That is, most teachers modeled these skills and expected students to learn by example; relatively few teachers explicitly taught students how to enact the skill or provided feedback on or assessed their performance.

Together, the comparisons presented in this chapter indicate clear and consistent differences between ObD teachers and their national counterparts on the use of practices aligned with the key ObD design principles. As noted earlier, we cannot determine the reasons for these differences, but they do suggest that students in ObD schools are, on average, experiencing a somewhat different approach to learning than is typical for other

Table 5.6. Percentage of Teachers Who Reported Addressing PYD/SEL Competencies in Their Instruction to a Large Extent

PYD/SEL Competency	Addressed to a Large Extent	
	ATP (%)	ObD (%)
Making responsible decisions*	59	66
Persisting through challenging tasks**	55	63
Establishing and maintaining positive relationships*	53	60
Collaborating with other students**	51	59
Feeling and showing empathy for others**	44	56
Learning mindsets**	38	56
Communicating their thoughts and emotions**	37	49
Developing a sense of identity	35	44
Understanding and managing emotions**	32	43
Handling stress	32	38

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

***indicates statistically significant difference in item means at alpha < 0.01.*

ATP results are weighted to be nationally representative.

Question text: Please indicate to what extent you address each of the following topics in your instruction. Response options: Not addressed; To a small extent; To a moderate extent; To a large extent. ObD N = 178–180; ATP N = 998.

U.S. high school students—an approach characterized by stronger supports for students to demonstrate mastery and advance through content based on that demonstration, to pursue instructional opportunities that are aligned with their learning needs and interests, and to develop the social and emotional competencies that they will need as they move on to college or careers. As we discussed in Chapter Three, these approaches were not ubiquitous in ObD schools, and their intensity and frequency varied across those schools. But the survey findings suggest that ObD teachers have adopted practices aligned with the key design principles to a greater degree than have other high school teachers, on average.

Perceived Barriers and Supports

ObD teachers' perceptions regarding barriers to, and supports for, implementation of the design principles were similar to those of the national sample on several topics. For instance, roughly equal majorities of ObD and ATP teachers agreed that they had access to curriculum materials that were of high quality, addressed the learning needs of all their students, contributed to their efforts to promote college and career readiness, and were accessible to students outside of the school day and school building. In this section, we focus on two areas where we observed fairly large differences—data access and the source of curriculum materials.

Opinions about data access and data system quality were more positive among ObD teachers than among high school teachers nationally. ObD teachers reported more-frequent receipt of data on the performance of individual students, including data about students who needed extra assistance and students who had achieved mastery, compared with their national counterparts (Table 5.7). These are the kinds of data that can help teachers tailor their instruction to individual students' needs, so it is noteworthy that ObD teachers reported greater access to this information. These findings are consistent with differences in teachers' reported use of data shown in Table 5.3.

Table 5.7. Percentage of Teachers Who Reported Receiving Student Data at Least Weekly

Student Data	Received Student Data at Least Weekly	
	ATP (%)	ObD (%)
Information about student performance on specific concepts or skills***	17	36
Identification of specific students who need extra assistance***	17	35
Identification of specific students who have achieved mastery***	11	35
Youth development outcomes***	9	27
Identification of specific students who are at risk of dropping out or not progressing to the next grade***	9	22

*NOTES: ***indicates statistically significant difference in item means at alpha < 0.001.*

ATP results are weighted to be nationally representative.

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 a year, Approximately monthly, A few times per month, Approximately weekly, A few times per week, At least daily. ObD N = 182–183; ATP N = 1,002–1,004.

Table 5.8. Percentage of Teachers Who Agreed or Strongly Agreed with Statements Regarding Access to Data to Inform Instruction

Statement	Agree or Strongly Agree	
	ATP (%)	ObD (%)
I have the necessary skills and experience to use data to guide my instruction.**	82	90
Students have access to information from data systems that track their progress on particular tasks, skills, or for the course overall.**	57	78
I have access to high-quality data that help me adapt the pace, content, or strategies of instruction to meet students' needs.**	54	73
Our school's data system provides real-time data that are actionable.**	54	73
Our school's data system and assessments enable me to make good decisions about mastery-based progression for individual students.**	53	72
Our school's data system and assessments provide adequate information about students' progress toward specific learning objectives.**	53	71
I can use the school's data system to easily produce the views or reports I need.*	57	64
Students regularly review data on their own progress using the school's data system.**	44	61

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

ATP results are weighted to be nationally representative.

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.

We also found that, compared with the national sample, ObD teachers expressed more-favorable opinions about access to data to inform instruction (Table 5.8). ObD teachers were more likely than teachers in the national sample to agree that their schools' data systems produced the data they needed, were easy to use, provided real-time data that were actionable, enabled them to make good decisions about mastery-based progression for individual students, and provided adequate information about students' progress toward specific learning objectives—all of which are key enablers of mastery-based progression. Differences between ObD and ATP teachers' responses ranged from 8 to 21 percentage points. Although our data do not speak to the specific features of school data systems to which teachers nationally had access, data from a separate, recent national survey suggest that the data systems of most teachers allow them to see data they enter themselves (e.g., grades, attendance), as well as standardized test results. In general, the data systems used by teachers nationally do not provide access to other student information, such as discipline records or course enrollment, or other more-complex data that might require processing or analysis before use (Berglund and Tosh, 2020).

Compared with the national sample, ObD teachers reported spending more time developing curriculum materials. Another difference between ObD teachers and the national sample is related to the source of curriculum materials. Teachers in the national sample reported spending an average of three to four hours per week developing their own materials, compared with the five to six hours that ObD teachers reported. A possible reason for this is that teachers in ObD schools lacked access to curriculum materials that supported their implementation of the key design principles. We find some support for this hypothesis in a comparison of perceptions regarding curriculum materials. Although these perceptions were largely similar between the two groups, as noted earlier, only 27 percent of ObD teachers strongly agreed that that their curriculum materials supported mastery-based instruction, and 25 percent described lack of curriculum flexibility as a major barrier. These findings suggest that implementation of personalized and mastery-based practices might benefit from additional curriculum flexibility and support.

CHAPTER SIX

Implications

Implications

- **Although teachers in ObD schools reported more extensive use of personalized, mastery-based, and PYD instructional approaches, these did not result in discernable improvement on available student outcome measures.**
- **Mastery-based instruction requires data that are not typically available in traditional high school classrooms.**
- **District leaders, external support providers, principals, and teachers should consider working together to select and/or develop high-quality curricula suitable to mastery-based and PL environments.**
- **Such resources as adaptable curriculum materials and PD could help teachers address the needs of students who struggle to achieve mastery.**
- **PYD and related social and emotional competencies may help students succeed in mastery-based systems.**
- **Schools and districts should develop strategies to mitigate the negative effects of principal turnover on reform efforts.**
- **Support providers for schools that are engaged in complex reforms should bring a continuous improvement lens.**
- **Complex reforms require a dedicated, aligned system of supports that includes district staff, principals, and teachers.**
- **As in most studies of broad high school initiatives, measurement limitations may have hampered our ability to capture the full breadth of potential ObD effects.**

Many of the strategies undertaken by the ObD schools are popular, and schools will continue to adopt them. Schools and districts interested in adopting the comprehensive school design process advocated by ObD should be mindful of the fact that the design principles—particularly the key design principles—are difficult to implement well. In implementing these strategies, the ObD schools faced challenges—such as principal turnover and selecting and developing appropriate curriculum materials—that are commonly faced by other schools. To be successful, ObD school staff and students may need different resources and supports than other high schools.

This study did not find statistical evidence that ObD's principles-based design process and supports resulted in a set of school models that produced improved student outcomes. It is possible that the available measures and the limited study duration did not adequately capture how ObD may have improved student prospects for life success. For example, we were not able to track the majority of ObD students beyond high school graduation. Ideally, future studies would follow students long enough to directly measure at least some college outcomes, such as admission, persistence, and graduation. Although it may not be possible for all studies to track student outcomes over

a long period of time, we encourage measurement of these long-term outcomes where possible.

In this section, we discuss lessons drawn from ObD implementation that may help other schools that use mastery-based, personalized, or PYD approaches to manage similar challenges and implement similar reforms in a high-quality way. The implications from this study can also be informative for researchers and others who are interested in understanding how high schools enact innovative practices and how those practices influence students' learning.

Although teachers in ObD schools reported more extensive use of personalized, mastery-based, and PYD instructional approaches, these did not result in discernable improvement on available student outcome measures. In their survey responses, ObD teachers reported emphasizing these three key design principles in their instruction to a greater extent than teachers nationally. ObD teachers also reported more extensive use of student data to inform their instruction and more-positive opinions about data access and school data system quality. Although these summary findings mask some variability in practices and supports within both the ObD and the ATP groups, on average, the differences were consistent and sometimes large. However, we did not find any evidence that these changes in instructional practice were linked to changes in student outcomes. We were not able to quantitatively investigate the relationship between implementation and outcomes, but our interview data suggest that one possible explanation for the lack of improvement in student outcomes is that the ObD schools were still working to implement the key design principles in a deep and integrated way. Staff in most schools reported implementing only one or two of the instructional practices that characterized each of the three key design principles. In addition, teachers in most schools reported that they were not explicitly teaching or assessing the social and emotional skills that would help students succeed in personalized and mastery-based environments. The key design principles are complex, and it is not surprising that staff in most schools focused on a few aspects of the design principles, but it is possible that this relatively limited implementation was sufficient to be different from practices reported by teachers nationally—but not sufficient to have a measurable effect on student outcomes.

Mastery-based instruction requires data that are not typically available in traditional high school classrooms. Mastery is a marked departure from how most high schools typically assess and promote students. Teachers are expected to prepare students for culminating, high-stakes exams and face pressure to advance through the content even if some students are struggling to master previous material. Ideally, information about which skills and concepts each student did—or did not—master would be available with well-designed mastery-based assessments. Teachers in the ObD schools used data from conversations with students and performance on classroom tasks, as well as student achievement data, to inform instructional decisions. ObD teachers used student achievement data to inform their instructional decisions to a greater extent than teachers nationally. These findings suggest that mastery-based instruction can be facilitated by multiple sources of high-quality data about which concepts and skills students have mastered and school data systems that provide easy access to this information. Principals in schools that aim to implement mastery-based instructional practices should work closely with teachers to ensure that they can access the data they need easily and frequently. Principals and district staff should also ensure that assignments and assessments provide detailed information about which concepts and skills students have and have not mastered.

District leaders, external support providers, principals, and teachers should consider working together to select and/or develop high-quality curricula suitable to mastery-based and PL environments. At the beginning of the initiative, CCNY recognized that the options for high-quality curricula that were well-suited to mastery-based and PL environments were limited. CCNY therefore intentionally selected districts that could provide the ObD schools with flexibility to select or develop their own curricula. A high level of autonomy was important; the ObD schools needed to be able to use curricula that were compatible with mastery, amenable to personalization, and aligned with the competencies they wanted to teach.

Our findings indicate that the lack of high-quality, appropriate curricula was a barrier to implementation of the key design principles. Although most ObD teachers enjoyed the opportunity to use their creativity and professional skills to develop curricula (Steiner

et al., 2017) and (perhaps as a result) were generally satisfied with them, we also heard that selecting and developing materials was a challenge. ObD teachers spent several hours a week—more than teachers in other U.S. high schools—selecting or developing most of their materials and used a combination of several online programs and teacher-developed materials.

The time ObD teachers spent selecting and developing curriculum materials prevented them from focusing on other things, such as tailoring those materials to the learning needs of individual students or developing multiple versions of an assignment to offer students choice (Steiner et al., 2017), a concern that has been raised in other studies (Pane et al., 2017). Teachers worried that the materials they developed were not challenging enough to prepare students for college and careers, a concern that is consistent with other research (Gross and DeArmond, 2018).

In the spirit of providing teachers with the autonomy to make choices about curricula that were right for their school models and students, most districts did not provide the ObD schools with curriculum materials. Most teachers also reported that they would have liked more support from their districts in vetting off-the-shelf curricula to reduce the amount of time spent researching these materials. Taken together, these findings suggest that while autonomy in curriculum design is important, ObD teachers wanted some resources, suggestions, and support as a place to start.

The balance of providing curriculum resources and autonomy for curriculum development may be different in each school context, and it may not be feasible for district staff to provide customized support to every unique school model. District and school leaders could therefore consider ways to work with teachers and external support providers to find a balance of recommended and shared resources while also providing space for teachers to develop original materials. For example, districts could develop a repository of mastery-aligned digital materials that teachers could browse or facilitate ways for teachers to share the assignments and materials they develop broadly within the district. External support providers, such as Springpoint or intermediary organizations, could also play a role in developing such a repository. In addition, teachers may need training in how to develop curricula for personalized and mastery-based environments. Districts could partner with external support providers to offer such training to help school staff share resources and efficiently develop challenging materials that are suitable for their schools and address the learning needs of their students.

Such resources as adaptable curriculum materials and PD could help teachers address the needs of students who struggle to achieve mastery. Adapting curriculum materials to meet the diverse learning needs of students is a challenge many teachers face. Like most teachers, teachers in mastery-based schools need to be able to use information about students' mastery of concepts and skills to help those who struggle to achieve mastery. This skill is crucial in schools—like the ObD schools—where most students perform well below grade level. ObD teachers recognized that different students need different supports to achieve mastery, and most reported adjusting the supports they provided depending on the student. ObD teachers used such strategies as reviewing or reteaching the content or skills, giving students a task similar in complexity, and working step-by-step with students to revise or work on a similar task. Few ObD teachers provided a different task, or a task that was tailored to the student's learning level, largely because creating such tasks is time-consuming. These findings suggest that teachers in mastery-based schools need a number of resources to help them support students who did not achieve mastery.

While tailored support for each school model may not be feasible for some districts, district staff should consider how to provide mastery-based schools with resources and materials that may benefit a broader group of schools. For example, district staff can leverage their own expertise, or that of expert external partners, to provide teachers with curriculum materials that include multiple versions of assignments or that can be easily adapted to the learning needs of different students. District staff and school principals can provide teachers with PD opportunities—including support from expert external organizations, such as Springpoint—that help them use multiple types of student data to address the needs of students who did not achieve mastery.

PYD and related social and emotional competencies may help students succeed in mastery-based systems. Initially, most schools did not emphasize PYD and related social and emotional competencies (Steiner et al., 2017) in a way that was integrated with academic instruction. As the initiative continued, evidence regarding the important role that social and emotional competencies play in academic and postsecondary success became more widely available (Aspen Institute National Commission on Social, Emotional, and Academic Development, 2019). Although the phrase "positive youth development"

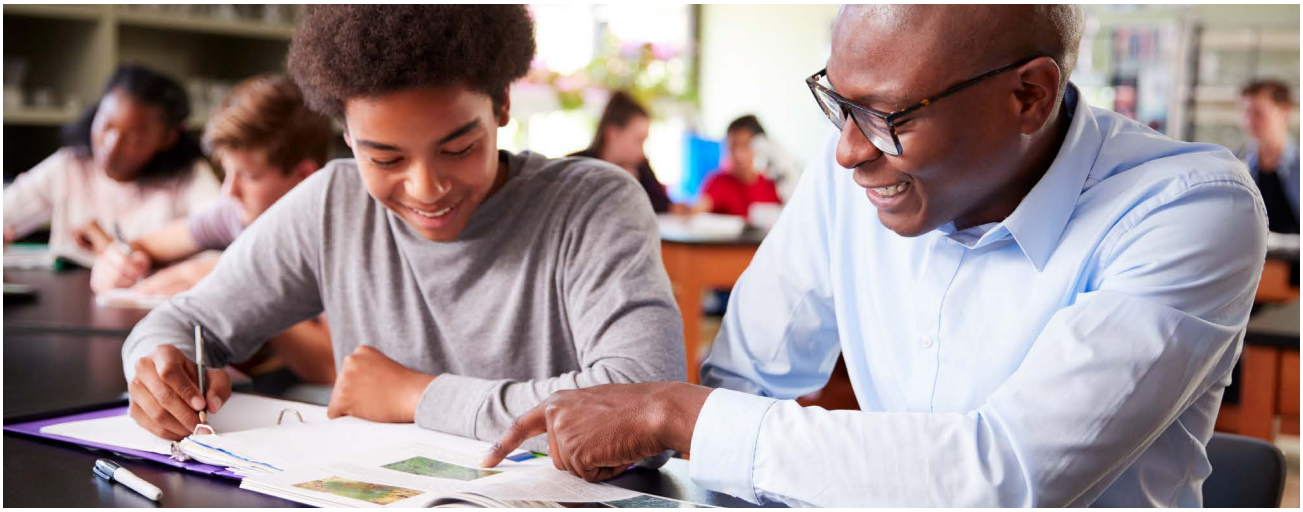
covers more than just students' social and emotional competencies, in the final year of the initiative, many ObD teachers emphasized social and emotional competencies in their instruction and did so to a greater extent than high school teachers nationally. The extent to which schools focus on skills like communication, critical thinking, and collaboration can have implications for how well prepared students will be to achieve their academic and postsecondary goals (Allensworth et al., 2018; Durlak et al., 2011; Taylor et al., 2017). Mastery-based schools should consider including competencies for SEL skills in mastery frameworks, providing teachers with research-based curricula and assessments aligned to those competencies, and providing targeted training to help teachers implement them. Training in teaching SEL in connection with academic content may be particularly important, given that most teachers do not receive such training in their pre-service programs (Hamilton, Doss, and Steiner, 2019). Our findings about how student experiences with mastery-based and personalized instructional practices varied over time suggest that staff in some schools may have adopted a gradual release approach and allowed students more autonomy as they advanced through grade levels. Future schools could consider a similar approach or could consider orientation programs or other ways to help students adjust to the expectations of a mastery-based environment.

Schools and districts should develop strategies to mitigate the negative effects of principal turnover on reform efforts. Changes in principal leadership are a challenge that many schools face, and they can be particularly disruptive in schools that are undertaking complex reforms (Schwartz et al., 2020). Many of the ObD schools experienced a change of principal during the course of the study, and for some, it was perceived as a barrier to successful implementation of the design principles. Change in principal leadership is likely inevitable and will probably be disruptive to some extent, but it does not have to result in redesigning the school model. Districts, funders, and external support providers could plan for changes in principal leadership over the course of a reform. For example, district staff should work with external support providers and sitting principals to develop operations manuals or other materials that would help new principals get up to speed and develop other support geared specifically to new principals. Funders, if providing external support, could consider ways to continue the support for new principals. When changes in principal leadership do occur, principals can develop clear systems and channels for communicating changes to the school design or policies.

Support providers for schools that are engaged in complex reforms should bring a continuous improvement lens. CCNY anticipated that the ObD schools' models and support needs would evolve over time. After the first two years of the initiative, CCNY observed that most schools focused on a subset of design principles—particularly mission and culture, PYD, personalization, and mastery. CCNY advised schools to prioritize the three design principles of mastery, personalization, and PYD and designated them as power principles (Steiner et al., 2017). Encouraging schools to focus on a subset of the design principles facilitated implementation by not overwhelming the schools and offered a refined strategy for later cohorts of schools. The support Springpoint provided also evolved over time to accommodate this shift and respond to schools' changing needs as the initiative unfolded.

Support providers and developers of schools engaged in complex reforms should communicate their expectation that schools will experiment and devise ways to collect and disseminate lessons learned. Support providers should also be ready to adapt and should consider what resources they will need to provide adequate supports as school needs change. The fact that the three key design principles did not have commonly accepted definitions in the field highlights the importance of clear communication while schools experiment with implementing the design principles in the way that is best suited to their context. In view of this, districts and external support providers should consider ways to help schools engaged in complex reforms develop a continuous improvement process and strategies to communicate changes clearly and consistently.

Complex reforms require a dedicated, aligned system of supports that includes district staff, principals, and teachers. CCNY established Springpoint as a dedicated support for ObD principals, a feature not common to most large-scale school-reform efforts. Springpoint staff focused on helping school leaders implement the design principles and tailored their support as school leaders' needs evolved. Nearly all ObD school leaders reported that they valued the expertise of Springpoint staff and that Springpoint was an important supplement to the supports they received from their districts and other



organizations. For example, many school leaders appreciated how Springpoint feedback helped them look ahead and plan for upcoming implementation needs and challenges. The school leaders who received support from intermediary organizations reportedly appreciated those supports as well. Broadly, these findings are consistent with those in the prior report (Steiner et al., 2017), in that Springpoint continued to play a key role in supporting school leaders and, along with intermediary organizations and other external partners, filled some gaps in district support.

However, the constraints on Springpoint support, such as the inevitable time limit and focus primarily on principals, suggest potential areas for improvement. First, Springpoint was focused primarily on supporting principals but also provided support to district staff and helped principals support teachers, in some ObD sites. Many district staff and school principals wished that Springpoint could expand its focus to systematically include district staff and teachers. Second, although there is an inherent time limit on externally funded resources and the ObD schools were intended to be self-sustaining, the need for PD may increase rather than decrease over time, as school staff become more familiar with new instructional practices and more aware of implementation nuances and challenges (Kaufman et al., 2016). These findings suggest that an aligned system of supports that can persist and adapt as PD needs change over time and that includes district leaders, intermediary organizations, school leaders, and teachers could benefit schools engaged in complex reforms. District staff and funders could work together to ensure that staff engaged in complex reforms at all levels receive continued expert implementation support.

As in most studies of broad high school initiatives, measurement limitations may have hampered our ability to capture the full breadth of potential ObD effects. Although we examined a broad set of outcomes, the academic measures available across districts were limited to mathematics and ELA assessments that do not reflect the breadth of subject matter that high schools cover across all grade levels (e.g., most states assess mathematics in the ninth grade). Behavioral measures were limited to attendance and suspension rates, which may not be very sensitive to the full range of PYD skills that ObD schools sought to impart; college readiness measures were confined to predictive high school outcomes rather than direct measures of college admission, persistence, and graduation. To understand the extent to which ObD or other initiatives for high school students are preparing students for postsecondary education or careers, researchers would need to follow students into their postsecondary lives to document post–high school performance. When this type of long-term study is infeasible, researchers should explore opportunities to incorporate a wider variety of measures, including grade-point average and engagement in learning (National Academies of Sciences, Engineering, and Medicine, 2019). These measures were not available for us in the present research owing to lack of consistent metrics across participating districts. This limitation points to the need for more common, standardized indicators of postsecondary readiness and success to enable educators and researchers to generate evidence-based guidance for high schools.

Acknowledgments

The authors are grateful to the ObD students, teachers, and administrators who voluntarily participated in project data collection; district, CCNY, and Springpoint staff who facilitated connections with the schools and participated in interviews; and NWEA staff who facilitated assessment administration and provided assessment data and analytic support.

We appreciate the efforts of the RAND American Educator Panels research team, particularly David Grant, Christopher Young, Claude Setodji, Matthew Strawn, and Gerald Hunter. We thank the National Education Association and the Bill & Melinda Gates Foundation, which were the original sponsors of the ATP. Finally, we are extremely grateful to the U.S. public school teachers who agreed to participate in the ATP and took this particular survey. Their time and willingness to share their experiences are invaluable for this effort and for helping us to understand more about how to better support their hard work in schools.

We are grateful to the following RAND staff who contributed to the research: Hannah Acheson-Field, Olamigoke Akinniranye, Matt Baird, Andy Bogart, Michelle Bongard, Karen Christiansen, Amanda Edelman, Suzette Gambone, Mark Harris, Brittany Joseph, Courtney Kase, Katelynn Kelly, Serafina Lanna, Stephanie Lonsinger, Scott Naftel, Marian Oshiro, Alyssa Ramos, Rachel Ross, Melanie Rote, Mollie Rudnick, Anna Saavedra, Clare Stevens, Quinton Stroud, Lindsey Thompson, Tiffany Tsai, Elaine Wang, Katie Whipkey, and Paul Yoo.

This document benefited substantively from feedback from Fatih Unlu, Julia Kaufman, and Jennifer Steele, as well as Saskia Levy Thompson and Jennifer Timm and their colleagues at CCNY. Samantha Bennett provided expert editing. Monette Velasco gracefully managed the publications process. Any flaws that remain are solely the authors' responsibility.

References

Allensworth, Elaine M., Camille A. Farrington, Molly F. Gordon, David W. Johnson, Kylie Klein, Bronwyn McDaniel, and Jenny Nagaoka, *Supporting Social, Emotional, and Academic Development: Research Implications for Educators*, Chicago, Ill.: UChicago Consortium on School Research, 2018. As of June 30, 2020: https://consortium.uchicago.edu/sites/default/files/2019-01/Supporting%20Social%20Emotional-Oct2018-Consortium.pdf

Aspen Institute National Commission on Social, Emotional, and Academic Development, *From a Nation at Risk to a Nation at Hope: Recommendations from the National Commission on Social, Emotional, and Academic Development*, Washington, D.C.: The Aspen Institute, 2019. As of May 12, 2020: http://nationathope.org/report-from-the-nation

Balfanz, Robert, "Can the American High School Become an Avenue of Advancement for All?" *Future of Children*, Vol. 19, No. 1, 2009, pp. 17–36.

Bang, Heejung, and James M. Robins, "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, Vol. 61, No. 4, 2005, pp. 962–973.

Barnum, Matt, "Maine Went All In on 'Proficiency-Based Learning'—Then Rolled It Back. What Does That Mean for the Rest of the Country?" *Chalkbeat*, October 18, 2018. As of June 16, 2020: https://www.chalkbeat.org/2018/10/18/21105950/maine-went-all-in-on-proficiency-based-learning-then-rolled-it-back-what-does-that-mean-for-the-rest

Benjamini, Yoav, and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 289–300.

Berglund, Tiffany, and Katie Tosh, *Educator Access to and Use of Data Systems*, Santa Monica, Calif.: RAND Corporation, RR-2575/8-1-BMGF, 2020. As of October 1, 2020: https://www.rand.org/pubs/research_reports/RR2575z8-1.html

Bloom, Howard S., Saskia Levy Thompson, and Rebecca Unterman, *Transforming the High School Experience: How New York City's New Small Schools Are Boosting Student Achievement and Graduation Rates, Executive Summary*, New York: MDRC, 2010.

Bloom, Howard S., and Rebecca Unterman, *Sustained Positive Effects on Graduation Rates Produced by New York City's Small Public High Schools of Choice*, New York: MDRC, January 2012.

Carnegie Corporation of New York, verbal communication to the RAND project team about the enabling conditions ObD grantees were assessed for prior to funding, May 2014.

———, email communication to the RAND project team about the purpose of the ObD initiative and definitions of key terms, January 31, 2017.

CASEL—*See* Collaborative for Academic, Social, and Emotional Learning.

CCNY—*See* Carnegie Corporation of New York.

Center on Reinventing Public Education, "Portfolio Strategy," webpage, 2018. As of June 16, 2020: https://www.crpe.org/research/portfolio-strategy

Collaborative for Academic, Social, and Emotional Learning, "What Is SEL?" webpage, undated. As of February 23, 2020:
https://casel.org/what-is-sel/

Committee for Children and Collaborative for Academic, Social, and Emotional Learning, *Why Social and Emotional Learning and Employability Skills Should Be Prioritized in Education*, Seattle, Wash.: Committee for Children, 2016. As of June 3, 2020:
https://www.cfchildren.org/wp-content/uploads/policy-advocacy/sel-employability-brief.pdf

Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, Vol. 132, No. 4, 2017, pp. 1593–1640.

Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger, "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions," *Child Development*, Vol. 82, No. 1, January/February 2011, pp. 405–432.

Grant, Sean, Laura S. Hamilton, Stephani L. Wrabel, Celia J. Gomez, Anamarie A. Whitaker, Jennifer T. Leschitz, Fatih Unlu, Emilio R. Chavez-Herrerias, Garrett Baker, Mark Barrett, Mark Harris, and Alyssa Ramos, *Social and Emotional Learning Interventions Under the Every Student Succeeds Act: Evidence Review*, Santa Monica, Calif.: RAND Corporation, RR-2133-WF, 2017. As of January 23, 2019:
https://www.rand.org/pubs/research_reports/RR2133.html

Gross, Betheny, and Michael DeArmond, *Personalized Learning at a Crossroads: Early Lessons from the Next Generation Systems Initiative and the Regional Funds for Breakthrough Schools Initiative: Executive Summary*, Seattle, Wash.: Center on Reinventing Public Education, June 2018. As of September 22, 2019:
https://www.crpe.org/sites/default/files/crpe-personalized-learning-at-crossroads-executive-summary.pdf

Gross, Betheny, Sivan Tuchman, and Susan Patrick, *A National Landscape Scan of Personalized Learning in K–12 Education in the United States*, Vienna, Va.: iNACOL, June 2018. As of September 22, 2019:
https://www.inacol.org/wp-content/uploads/2018/06/iNACOL_ANationalLandscapeScanOfPersonalizedLearning.pdf

Hamilton, Laura S., Christopher Joseph Doss, and Elizabeth D. Steiner, *Teacher and Principal Perspectives on Social and Emotional Learning in America's Schools: Findings from the American Educator Panels*, Santa Monica, Calif.: RAND Corporation, RR-2991-BMGF, 2019. As of June 16, 2020:
https://www.rand.org/pubs/research_reports/RR2991.html

Hamilton, Laura S., David Grant, Julia H. Kaufman, Melissa Diliberti, Heather L. Schwartz, Gerald P. Hunter, Claude Messan Setodji, and Christopher J. Young, *COVID-19 and the State of K–12 Schools: Results and Technical Documentation from the Spring 2020 American Educator Panels COVID-19 Surveys*, Santa Monica, Calif.: RAND Corporation, RR-A168-1, 2020. As of June 3, 2020:
https://www.rand.org/pubs/research_reports/RRA168-1.html

Hamilton, Leah, "Barr Announces $4.5 Million in New Grants for 'Doing High School Differently,'" *Barr Foundation Blog*, June 27, 2018. As of June 16, 2020:
https://www.barrfoundation.org/blog/announcing-new-partners-doing-high-school-differently

Hamilton, Leah, and Anne Mackinnon, *Opportunity by Design: New High School Models for Student Success*, New York: Carnegie Corporation of New York, 2013. As of June 16, 2020:
https://www.carnegie.org/media/filer_public/83/72/8372b753-7f6e-4213-bd05-2663587610d6/ccny_challenge_2013_opportunity.pdf

Hullsiek, Katherine Huppler, and Thomas A. Louis, "Propensity Score Modeling Strategies for the Causal Analysis of Observational Data," *Biostatistics*, Vol. 3, No. 2, 2002, pp. 179–193.

Jackson, C. Kirabo, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, and Sebastián Kiguel, *School Effects on Socio-Emotional Development, School-Based Arrests, and Educational Attainment*, Washington, D.C.: National Bureau of Economic Research, NBER Working Paper No. 26759, February 2020.

Jerald, Craig, Neil Campbell, and Erin Roth, *High Schools of the Future: How States Can Accelerate High School Redesign*, Washington, D.C.: Center for American Progress, December 4, 2017. As of June 16, 2020: https://cdn.americanprogress.org/content/uploads/2017/12/01061134/110617_HSRedesign-brief.pdf

Kaufman, Julia H., Laura S. Hamilton, Brian M. Stecher, Scott Naftel, Michael Robbins, Lindsey E. Thompson, Chandra Garber, Susannah Faxon-Mills, and V. Darleen Opfer, *What Supports Do Teachers Need to Help Students Meet Common Core State Standards for English Language Arts and Literacy? Findings from the American Teacher and American School Leader Panels*, Santa Monica, Calif.: RAND Corporation, RR-1374-1, 2016. As of October 1, 2020: https://www.rand.org/pubs/research_reports/RR1374-1.html

Lake, Robin, Paul T. Hill, and Tricia Maas, *Next Generation School Districts: What Capacities Do Districts Need to Create and Sustain Schools That Are Ready to Deliver on Common Core?* Seattle, Wash.: Center on Reinventing Public Education, April 2015.

Le, Cecilia, and Jill Frankfort, *Accelerating College Readiness: Lessons from North Carolina's Innovator Early Colleges*, Boston, Mass.: Jobs for the Future, March 2011.

Mahoney, Joseph L., Joseph A. Durlak, and Roger P. Weissberg, "An Update on Social and Emotional Learning Outcome Research," *Phi Delta Kappan*, Vol. 100, No. 4, 2018, pp. 18–23.

Margolis, Jesse, *Three-Year MAP Growth at Schools Using Teach to One: Math*, MarGrady Research, February 2019. As of October 1, 2020: http://margrady.com/tto/

Marsh, Julie A., Susan McKibben, Heather J. Hough, Michelle Hall, Taylor N. Allbright, Ananya M. Matewos, and Caetano Siqueira, *Enacting Social-Emotional Learning: Practices and Supports Employed in CORE Districts and Schools*, Stanford, Calif.: Policy Analysis for California Education, April 19, 2018.

Nagaoka, Jenny, Camille A. Farrington, Stacy B. Ehrlich, and Ryan D. Heath, *Foundations for Young Adult Success: A Developmental Framework*, Chicago, Ill.: UChicago Consortium on Chicago School Research, June 2015.

National Academies of Sciences, Engineering, and Medicine, *Monitoring Educational Equity*, Washington, D.C.: The National Academies Press, 2019. As of September 21, 2020: https://doi.org/10.17226/25389

National Assessment of Educational Progress, "The Nation's Report Card," webpage, undated. As of June 3, 2020: https://www.nationsreportcard.gov/

Pane, John F., Elizabeth D. Steiner, Matthew D. Baird, Laura S. Hamilton, and Joseph D. Pane, *Informing Progress: Insights on Personalized Learning Implementation and Effects*, Santa Monica, Calif.: RAND Corporation, RR-2042-BMGF, 2017. As of June 16, 2019: http://www.rand.org/pubs/research_reports/RR2042.html

RAND Education and Labor, "American Educator Panels," webpage, 2020. As of June 16, 2020: https://www.rand.org/education-and-labor/projects/aep.html

Ready, Douglas D., Katharine Conn, Shani S. Bretas, and Iris Daruwala, *Final Impact Results from the i3 Implementation of* Teach to One: Math, New York: Consortium for Policy Research in Education, January 2019. As of October 1, 2020: https://www.tc.columbia.edu/media/centers/cpre/Final-Impact-Results-i3-TtO-190224.pdf

Schwartz, Heather L., Laura S. Hamilton, Susannah Faxon-Mills, Celia J. Gomez, Alice Huguet, Lisa H. Jaycox, Jennifer T. Leschitz, Andrea Prado Tuma, Katie Tosh, Anamarie A. Whitaker, and Stephani L. Wrabel, *Early Lessons from Schools and Out-of-School Time Programs Implementing Social and Emotional Learning*, Santa Monica, Calif.: RAND Corporation, RR-A379-1, 2020. As of November 3, 2020: https://www.rand.org/pubs/research_reports/RRA379-1.html

Steele, Jennifer L., Matthew W. Lewis, Lucrecia Santibañez, Susannah Faxon-Mills, Mollie Rudnick, Brian M. Stecher, and Laura S. Hamilton, *Competency-Based Education in Three Pilot Programs: Examining Implementation and Outcomes*, Santa Monica, Calif.: RAND Corporation, RR-732-BMGF, 2014. As of October 1, 2020: http://www.rand.org/pubs/research_reports/RR732.html

Steiner, Elizabeth D., Christopher Joseph Doss, and Laura S. Hamilton, *High School Teachers' Perceptions and Use of Personalized Learning: Findings from the American Teacher Panel*, Santa Monica, Calif.: RAND Corporation, RR-A322-1, 2020. As of October 1, 2020: https://www.rand.org/pubs/research_reports/RRA322-1.html

Steiner, Elizabeth D., Laura S. Hamilton, Laura Stelitano, and Mollie Rudnick, *Designing Innovative High Schools: Implementation of the Opportunity by Design Initiative After Two Years*, Santa Monica, Calif.: RAND Corporation, RR-2005-CCNY, 2017. As of June 16, 2020: https://www.rand.org/pubs/research_reports/RR2005.html

Sturgis, Chris, "Boston Day and Evening Academy: Where Competency Education Is Good Teaching Practice," Aurora Institute, June 21, 2012. As of May 12, 2020: https://aurora-institute.org/cw_post/boston-day-and-evening-academy-where-competency-education-is-good-teaching-practice

Summit Learning, homepage, undated. As of June 16, 2020: https://www.summitlearning.org/

———, "The Science of Summit: School Models That Drive Student Success," webpage, August 7, 2017. As of June 16, 2020: https://blog.summitlearning.org/2017/08/science-of-summit-framework-research/

Taylor, Rebecca D., Eva Oberle, Joseph A. Durlak, and Roger P. Weissberg, "Promoting Positive Youth Development Through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-Up Effects," *Child Development*, Vol. 88, No. 4, 2017, pp. 1156–1171.

Thum, Yeow Meng, and Carl H. Hauser, *NWEA 2015 MAP Norms for Student and School Achievement Status and Growth*, Portland, Ore.: Northwest Evaluation Association, July 1, 2015.

Tosh, Katie, Sy Doan, Ashley Woo, and Daniella Henry, *Digital Instructional Materials: What Are Teachers Using and What Barriers Exist?* Santa Monica, Calif.: RAND Corporation, RR-2575/17-BMGF/SFF/OFF, 2020. As of April 13, 2020: https://www.rand.org/pubs/research_reports/RR2575z17.html

U.S. Department of Education, "Next Generation High Schools: Redesigning the American High School Experience," webpage, undated. As of June 16, 2020: https://www.ed.gov/highschool

XQ, "Let's Rethink High School Together," website, undated. As of June 3, 2020: https://xqsuperschool.org/

Zimmerman, Geoff, and Jillian Kuhlmann, *Personalized, Competency-Based Learning: Analysis and Reflections on Student Outcome Data in RSU2*, Cincinnati, Ohio: KnowledgeWorks, 2019. As of October 1, 2020: https://knowledgeworks.org/wp-content/uploads/2019/11/rsu2-data-analysis.pdf

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

For more information on this publication, visit www.rand.org/t/RRA322-3

Published by the RAND Corporation, Santa Monica, Calif.

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

Abbreviations

AP	Advanced Placement
ATP	American Teacher Panel
CCNY	Carnegie Corporation of New York
CEM	Coarsened Exact Matching
CMO	charter management organization
ELA	English language arts
ELL	English language learner
FRL	free and reduced-price lunch
LEP	limited English proficiency
ObD	Opportunity by Design
PD	professional development
RIT	Rasch Unit
VCG	virtual comparison group
WWC	What Works Clearinghouse

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

Variable	Cleveland	Denver	New York	Philadelphia	Prince George's	Providence
Race/ethnicity	✓	✓	✓	✓	✓	✓
Male	✓	✓	✓	✓	✓	✓
English language learner (ELL)/Limited English proficiency			✓	✓		
Free or reduced-price lunch Eligible (FRL)			✓	✓		✓
Days attended (percentage)	✓	✓	✓	✓		✓
Suspension				✓		✓
ELA (baseline)	✓	✓	✓	✓	✓	✓
Math (baseline)	✓	✓	✓	✓	✓	✓
Science (baseline)				✓		

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):

$$y_{igsd} = \beta_0 + \beta_1 T_{gsd} + X_{igsd}{}'\lambda + e_{igsd} \quad (1)$$

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 σ^2.

$\hat{\beta}_1$, the estimate of β_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:

$$\hat{\beta}_{1g} = \theta + \varepsilon_{ig} \quad (2)$$

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

$$\hat{\theta} = \frac{\sum_{i=1}^{g} \frac{1}{v_i} \hat{\beta}_{1i}}{\sum_{i=1}^{g} \frac{1}{v_i}} \quad (3)$$

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:

$$\hat{\theta}_d = \mu + \zeta_d + \varepsilon_{id} \tag{4}$$

where μ is the mean of all district effects, ζ_d is a between-district random effect with mean 0 and variance τ, and ε_{id} is within-district error, with mean 0 and variance ϑ^2. An estimate of the weighted mean impact (across *d* districts) is given by

$$\hat{\mu} = \frac{\sum_{i=1}^{d} \frac{1}{v_i} \hat{\theta}_i}{\sum_{i=1}^{d} \frac{1}{v_i}} \tag{5}$$

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 τ is relative to ϑ^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, τ 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) scale[3] 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.

$$w_i = \begin{cases} 1 & if\ T_i = 1 \\ \sum_{\{j|i \in VCG_j\}} \frac{1}{|VCG_j|} & if\ T_i = 0 \end{cases} \quad (6)$$

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

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

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

	Grade 9	
Characteristic	**ObD**	**Other**
Male (%)	53.3	47.7
Race/ethnicity (%)		
American Indian or Alaskan Native	0.8	0.1
Asian	12.3	14.5
Black	23.8	27.2
Hispanic	57.4	40.9*
Multiple races	0.0	0.6
White	2.5	12.4*
FRL (%)	76.3	70.3
ELL (%)	24.7	13.3*
Attendance (%)	87.5	90.3
Math	271.9	293.8*
ELA	291.5	309.9*
Total (*N*)	**122**	**97,509**

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

Table A.4.
Cleveland's Descriptive Statistics Before Propensity Score Weighting

	Grade 9		Grade 10		Grade 11		Grade 12	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	52.0	52.4	51.3	50.0	45.1	49.4	44.9	50.5
Race/ethnicity (%)								
Asian	1.9	1.1	1.5	1.1	1.3	2.0	0.0	2.1
Black	55.9	64.9	67.4	65.5	88.9	63.6*	99.3	62.6*
Hispanic	22.8	16.8	22.3	17.0	2.6	13.6*	0.7	11.9*
Other	4.1	3.3	0.4	3.3*	2.6	2.5	0.0	2.5
White	15.3	13.9	8.4	13.1	4.6	18.3*	0.0	20.8
Attendance (%)	89.3	89.7	91.6	91.0	90.2	92.6*	92.3	92.8
Math	665.1	666.3	654.3	673.1*	700.2	710.6*	390.3	404.92*
ELA	649.7	655.8	641.8	661.8*	713.1	723.8*	394.2	411.63*
Total (*N*)	**675**	**2,650**	**273**	**2,616**	**153**	**2,338**	**147**	**2,238**

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.5.
Denver's Descriptive Statistics Before Propensity Score Weighting

	Grade 9		Grade 10		Grade 11	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	73.8	53.4*	71.6	51.7*	52.7	50.9
Race/ethnicity (%)						
American Indian or Alaskan Native	0.0	0.6	0.0	0.6	0.0	0.8
Asian	6.0	2.7	7.5	2.5	3.3	2.8
Black	17.9	12.1	10.4	11.9	18.7	12.5
Hispanic	26.2	51.2*	16.4	48.3*	22.0	50.1*
Multiple races	7.1	4.4	6.0	3.4	7.7	4.0
Native Hawaiian	0.0	0.3	1.5	0.2	0.0	0.2
White	41.7	17.5*	49.3	17.2*	44.0	16.6*
Attendance (%)	86.0	79.3*	85.6	82.9	83.6	83.5
Math	729.4	723.3	744.0	728.9*	732.8	731.6
ELA	749.9	739.7*	755.6	743.7*	751.1	739.0*
Total (*N*)	**84**	**8,276**	**67**	**6,888**	**91**	**6,109**

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

	Grade 9		Grade 10		Grade 11		Grade 12	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	68.5	47.7*	67.3	45.3*	69.0	46.2*	59.5	48.6*
Race/ethnicity (%)								
American Indian or Alaskan Native	1.4	1.1	1.0	0.9	1.4	0.9	0.0	0.8
Asian	8.2	14.5	11.9	15.1	7.0	16.9*	13.9	18.6
Black	26.0	27.2	28.7	28.7	29.6	27.4	21.5	26.9
Hispanic	47.9	40.9	42.6	41.0	50.7	39.1	44.3	37.8
Multiple races	0.7	0.6	3.0	0.6	0.0	0.6	0.0	0.5
White	10.3	12.4	9.9	12.2	8.5	13.8	12.7	14.4
FRL (%)	66.9	70.3	63.6	71.8	79.4	71.4	76.9	75.2
ELL (%)	8.5	13.3	0.0	12.6	6.3	11.6	3.1	13.2*
Attendance (%)	91.5	90.3	89.7	88.6	98.5	95.4*	93.9	93.3
Math	286.9	293.8	302.3	298.5	308.0	300.4	304.1	301.6
ELA	306.7	309.9	311.3	308.2	301.7	303.0	304.7	301.7
Total (*N*)	**146**	**97,509**	**101**	**96,739**	**71**	**78,189**	**79**	**69,179**

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

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

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

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

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

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

	Grade 9	
Characteristic	**ObD**	**Other**
Male (%)	53.3	52.1
Race/ethnicity (%)		
American Indian or Alaskan Native	0.8	0.8
Asian	12.3	12.8
Black	23.8	23.8
Hispanic	57.4	55.7
Multiple races	0.0	0.1
White	2.5	3.0
FRL (%)	76.3	76.3
ELL (%)	24.7	24.5
Attendance (%)	87.5	88.0
Math	271.9	273.0
ELA	291.5	292.6
Total (*N*)	**122**	**97,509**

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

	Grade 9		Grade 10		Grade 11		Grade 12	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	52.0	51.9	51.3	51.6	45.1	47.9	44.9	47.2
Race/ethnicity (%)								
Asian	1.9	1.6	1.5	1.1	1.3	1.3	0.0	0.1
Black	55.9	57.7	67.4	67.8	88.9	87.0	99.3	97.8
Hispanic	22.8	21.6	22.3	20.0	2.6	3.7	0.7	0.9
Other	4.1	3.9	0.4	1.3	2.6	2.5	0.0	0.1
White	15.3	15.1	8.4	9.8	4.6	5.6	0.0	1.1
Attendance (%)	89.3	89.5	91.6	91.6	90.2	90.4	92.3	92.3
Math	665.1	664.5	654.3	656.0	700.2	700.3	390.3	390.7
ELA	649.7	650.1	641.8	643.8	713.1	713.5	394.2	395.3
Total (*N*)	**675**	**2,650**	**273**	**2,616**	**153**	**2,338**	**147**	**2,238**

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

	Grade 9		Grade 10		Grade 11	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	73.8	70.5	71.6	67.1	52.7	51.1
Race/Ethnicity (%)						
American Indian or Alaskan Native	0.0	0.3	0.0	0.2	0.0	0.2
Asian	6.0	6.0	7.5	7.6	3.3	3.4
Black	17.9	17.5	10.4	9.1	18.7	17.7
Hispanic	26.2	26.3	16.4	19.3	22.0	23.4
Multiple races	7.1	6.3	6.0	6.2	7.7	7.8
Native Hawaiian	0.0	0.1	1.5	1.1	0.0	0.1
White	41.7	41.7	49.3	47.4	44.0	42.5
Attendance (%)	86.0	85.9	85.6	85.4	83.6	83.9
Math	729.4	728.0	744.0	742.3	732.8	732.7
ELA	749.9	748.5	755.6	754.2	751.1	750.7
Total (*N*)	**84**	**8,276**	**67**	**6,888**	**91**	**6,109**

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

	Grade 9		Grade 10		Grade 11		Grade 12	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	68.5	66.8	67.3	66.4	69.0	67.3	59.5	58.4
Race/ethnicity (%)								
American Indian or Alaskan Native	1.4	1.6	1.0	1.2	1.4	1.0	0.0	0.2
Asian	8.2	8.5	11.9	11.7	7.0	7.3	13.9	14.6
Black	26.0	25.7	28.7	29.5	29.6	28.9	21.5	19.8
Hispanic	47.9	47.9	42.6	41.9	50.7	51.8	44.3	44.1
Multiple races	0.7	0.8	3.0	1.9	0.0	0.1	0.0	0.1
White	10.3	9.9	9.9	10.8	8.5	8.2	12.7	13.3
FRL (%)	66.9	67.2	63.6	65.6	79.4	78.5	76.9	75.8
ELL (%)	8.5	8.8	0.0	1.2	6.3	6.5	3.1	3.2
Attendance (%)	91.5	92.1	89.7	90.2	98.5	98.5	93.9	94.1
Math	286.9	288.6	302.3	303.8	308.0	308.0	304.1	307.2
ELA	306.7	307.3	311.3	311.2	301.7	301.8	304.7	306.3
Total (*N*)	**146**	**97,509**	**101**	**96,739**	**71**	**78,189**	**79**	**69,179**

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

	Grade 9		Grade 10		Grade 11		Grade 12	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	50.2	52.3	46.0	45.8	54.4	51.9	50.0	50.0
Race/ethnicity (%)								
American Indian or Alaskan Native	1.0	0.5	0.7	0.7	0.0	0.0	0.0	0.0
Asian	1.0	2.6	0.7	1.7	0.0	1.1	1.5	2.0
Black	46.3	46.2	55.8	56.4	59.6	59.7	66.4	65.3
Hispanic	47.8	43.8	35.5	34.6	36.8	34.6	29.9	28.7
Multiple races	2.0	2.2	2.2	2.2	2.9	3.0	0.7	0.9
Native Hawaiian or Pacific Islander	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unknown	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0
White	2.0	4.7	4.3	4.3	0.7	1.5	1.5	3.1
FRL (%)	83.2	76.7	69.8	70.6	66.7	66.4	78.2	77.3
LEP (%)	12.7	15.6	12.5	13.7	11.8	13.1	13.8	12.0
Attendance (%)	87.6	88.0	89.3	89.5	93.3	93.3	94.2	94.4
Suspension (%)	21.4	22.1	26.0	25.8	23.5	21.0	24.1	21.2
Math	820.0	833.6	817.0	818.7	852.4	857.9	1,201.2	1,211.5
ELA	910.5	913.6	902.2	902.2	909.4	916.6	1,228.0	1,246.3
Science	1,097.6	1,100.4	1,019.4	1,021.7	1,068.4	1,074.5	1,067.3	1,073.9
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 A.15.
Prince George's County's Descriptive Statistics After Propensity Score Weighting

	Grade 9		Grade 10		Grade 11	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	54.5	55.4	55.3	55.0	58.4	58.0
Race/Ethnicity (%)						
American Indian or Alaskan Native	0.0	0.0	0.0	0.0	0.0	0.0
Asian	1.1	0.9	1.1	1.1	1.5	1.5
Black	5.7	5.8	7.6	8.1	13.9	14.9
Hispanic	91.6	91.5	89.5	89.0	83.2	82.4
Multiple races	0.0	0.1	0.0	0.1	0.0	0.2
Native Hawaiian or Pacific Islander	0.0	0.0	0.0	0.0	0.0	0.0
White	1.6	1.5	1.7	1.5	1.5	1.1
Math	683.1	682.6	682.2	682.3	679.0	677.7
ELA	679.4	680.7	683.5	685.5	691.7	693.4
Total (*N*)	**367**	**10,356**	**19**	**10,384**	**137**	**8,185**

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

	Grade 9		Grade 10		Grade 11	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	61.7	60.2	49.4	52.5	41.1	42.3
Race/Ethnicity (%)						
American Indian or Alaskan Native	2.0	1.7	2.3	1.9	0.9	0.9
Asian	2.0	2.8	1.7	2.4	2.7	3.0
Black	13.4	14.0	18.4	17.9	18.8	18.3
Hispanic	68.5	67.1	66.1	66.8	64.3	66.0
Multiple races	3.4	3.4	5.2	4.6	5.4	3.7
Native Hawaiian or Pacific Islander	0.0	0.1	0.0	0.1	0.0	0.1
White	10.7	10.9	6.3	6.4	8.0	8.0
FRL (%)	58.4	56.0	44.8	44.5	35.7	37.6
Attendance (%)	82.9	83.1	89.5	89.5	91.1	91.5
Suspension (%)	5.6	5.0	16.7	18.9	10.0	9.9
Math	692.6	690.4	701.8	701.2	707.4	708.09
ELA	701.3	700.1	710.2	710.5	713.0	715.51
Total (*N*)	**149**	**2,147**	**174**	**2,072**	**112**	**1,778**

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

	Cleveland		Denver		New York City		Philadelphia		Providence	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	42.1	47.8	52.4	52.7	71.8	52.2*	50.5	48.7	44.7	52.4*
Race/ethnicity (%)										
American Indian or Alaskan Native			0.0	1.0	1.4	1.1	0.0	0.1	0.0	0.9
Asian	1.7	2.0	3.6	3.2	7.0	16.0*	0.0	12.1	0.0	0.1
Black	90.1	62.9*	17.9	15.7	33.8	28.6	54.6	52.4	2.1	4.4
Hispanic	1.7	13.9*	20.2	57.8*	45.1	39.8	41.2	17.4*	23.4	15.1
Multiple races			7.1	4.2	0.0	0.8	3.1	5.1	63.8	70.0
Native Hawaiian or Pacific Islander			0.0	0.3			0.0	0.0	0.0	0.1
Unknown	1.7	2.6					0.0	0.0		
White	5.0	18.6*	51.2	17.8*	9.9	13.1	1.0	12.9*	6.4	6.4
FRL (%)					76.9	72.3	61.2	59.0	40.4	60.3*
LEP (%)					4.6	12.9	17.9	7.4*	92.5	92.3
Attendance (%)	91.1	93.2*	85.1	82.8	96.4	94.8	93.8	94.1	10.6	13.6
Suspension (%)							20.9	8.6*	40.4	60.3*
Math	702.1	711.9*	738.0	727.9*	306.5	300.7	855.0	928.8*	711.5	704.8*
ELA	714.4	726.1*	746.6	728.5*	299.6	303.2	908.9	992.0*	719.2	721.5
Science							1,070.8	1,218.7*		
Total (*N*)	**121**	**2,100**	**84**	**8,088**	**71**	**73,168**	**97**	**7,581**	**47**	**677**

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.

Table A.18.
Descriptive Statistics for All Districts After Propensity Score Weighting for 2018–2019 12th-Graders

	Cleveland		Denver		New York City		Philadelphia		Providence	
Characteristic	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**	**ObD**	**Other**
Male (%)	42.1	45.7	52.4	52.2	71.8	69.8	50.5	49.0	44.7	44.8
Race/ethnicity (%)										
American Indian or Alaskan Native			0.0	0.4	1.4	0.9	0.0	0.0	0.0	3.5
Asian	1.7	1.7	3.6	3.6	7.0	7.1	0.0	2.7	0.0	0.0
Black	90.1	87.3	17.9	17.3	33.8	33.3	54.6	53.9	2.1	2.9
Hispanic	1.7	3.7	20.2	22.8	45.1	46.1	41.2	37.0	23.4	17.5
Multiple races			7.1	6.7	0.0	0.1	3.1	3.4	63.8	67.9
Native Hawaiian or Pacific Islander			0.0	0.1			0.0	0.0	0.0	0.1
Unknown	1.7	1.6					0.0	0.0		
White	5.0	5.8	51.2	49.0	9.9	9.8	1.0	3.0	6.4	6.8
FRL (%)					76.9	75.1	61.2	65.6	40.4	45.1
LEP (%)					4.6	5.2	17.9	18.8		
Attendance (%)	91.1	91.3	85.1	85.0	96.4	96.5	93.8	93.3	92.5	92.7
Suspension (%)							20.9	17.2	10.6	10.4
Math	702.1	702.4	738.0	737.8	306.5	306.7	855.0	864.2	711.5	710.4
ELA	714.4	715.1	746.6	746.3	299.6	299.9	908.9	920.1	719.2	719.3
Science							1,070.8	1,082.7		
Total (*N*)	**121**	**2,100**	**84**	**8,088**	**71**	**73,168**	**97**	**7,581**	**47**	**677**

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

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

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

			Unweighted VCG			Weighted VCG		
Subject	Variable	ObD Mean	Mean	Difference	Standardized Difference	Mean	Difference	Standardized Difference
Reading	Start RIT	204.34	206.69	–2.35	–0.12	204.38	–0.04	0.00
	FRL	93.90	90.30	3.60	0.52	89.91	3.99	0.58
	Elapsed	168.84	163.17	5.67	0.10	169.39	–0.55	–0.01
Math	Start RIT	212.34	214.97	–2.63	–0.14	212.40	–0.06	0.00
	FRL	94.40	90.19	4.22	0.58	90.30	4.10	0.57
	Elapsed	171.14	172.70	–1.55	–0.03	172.18	–1.03	–0.02

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

	Cleveland	Denver	New York City	Philadelphia	Prince George's County	Providence
Academic Outcomes						
Math	EOC Algebra (9)	PSAT 8/9 (9)	Regents CC ELA (9)*	Keystone Algebra (9)	Algebra 1 (9)	PSAT 8/9 (9)
ELA	EOC ELA (9)	PSAT 8/9 (9)	Regents CC Algebra (9)*	Keystone Literature (10)	ELA (11)	PSAT 8/9 (9)
Behavioral Outcomes						
Attendance	✓	✓ (9,10,11)	✓*	✓	N/A	✓ (9,10,11)
Suspension	✓	✓ (9,10,11)	✓*	✓	N/A	✓ (9,10,11)
College Readiness Outcomes						
PSAT	✓ (10)	✓ (10)	✓ (10)	✓ (10)	N/A	✓ (10)
SAT	✓ (12)	✓ (11)	✓ (11)	✓ (12)	N/A	✓ (11)
Credits	✓	N/A	✓*	✓	N/A	✓ (10,11)
AP tests passed	N/A	✓ (10,11)	✓ (12)	N/A	N/A	N/A
Graduation	✓ (12)	N/A	✓ (12)	✓ (12)	N/A	N/A

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

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

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

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

Math					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.00	–0.18	1.00	–2.37	4.86
Denver	0.00	–0.11	1.00	–3.21	2.99
New York City	0.00	0.26	1.00	–4.06	2.25
Philadelphia	0.00	–0.16	1.00	–3.25	5.64
Prince George's	0.00	–0.11	1.00	–2.49	4.36
Providence	0.00	–0.02	1.00	–2.55	3.14
Brooklyn Lab	0.00	0.26	1.00	–4.06	2.25
ELA					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.00	–0.06	1.00	–3.10	3.38
Denver	0.00	–0.20	1.00	–3.13	2.95
New York City	0.00	0.26	1.00	–2.57	1.51
Philadelphia	0.00	0.03	1.00	–3.09	3.45
Prince George's	0.00	0.09	1.00	–1.93	3.11
Providence	0.00	–0.14	1.00	–2.41	3.20
Brooklyn Lab	0.00	0.26	1.00	–2.56	1.51
Attendance					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.92	0.95	0.11	0.012	1.00
Denver	0.74	0.80	0.23	0.00	1.00
New York City	0.86	0.94	0.21	0.00	1.00
Philadelphia	0.89	0.95	0.16	0.00	1.00
Prince George's	N/A	N/A	N/A	N/A	N/A
Providence	0.83	0.90	0.19	0.00	1.00
Brooklyn Lab	0.86	0.95	0.22	0.00	1.00
Suspension					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.24	0.00	0.43	0.00	1.00
Denver	0.07	0.00	0.25	0.00	1.00
New York City	0.05	0.00	0.22	0.00	1.00
Philadelphia	0.10	0.00	0.30	0.00	1.00
Prince George's	N/A	N/A	N/A	N/A	N/A
Providence	0.08	0.00	0.27	0.00	1.00
Brooklyn Lab	0.06	0.00	0.24	0.00	1.00

Table A.23—Continued

PSAT					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.00	–0.26	1.00	–3.60	4.15
Denver	0.00	–0.14	1.00	–3.07	3.21
New York City	0.00	–0.15	1.00	–3.13	3.26
Philadelphia	0.00	–0.17	1.00	–2.80	3.74
Prince George's	N/A	N/A	N/A	N/A	N/A
Providence	0.00	–0.25	1.00	–2.80	4.16
Brooklyn Lab	N/A	N/A	N/A	N/A	N/A
SAT					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.00	–0.20	1.00	–3.04	4.36
Denver	0.00	–0.15	1.00	–2.89	3.05
New York City	0.00	–0.15	1.00	–2.81	2.89
Philadelphia	0.00	–0.16	1.00	–2.64	3.51
Prince George's	N/A	N/A	N/A	N/A	N/A
Providence	0.00	–0.23	1.00	–2.63	4.05
Brooklyn Lab	N/A	N/A	N/A	N/A	N/A
Credits					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	3.27	4.00	2.38	0.00	48.00
Denver	N/A	N/A	N/A	N/A	N/A
New York City	11.49	13.00	4.45	0.00	89.38
Philadelphia	3.65	4.00	1.25	0.00	15.00
Prince George's	N/A	N/A	N/A	N/A	N/A
Providence	11.39	11.00	6.79	0.00	30.50
Brooklyn Lab	11.32	13.00	4.67	0.00	89.38
AP Test					
District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland					
Denver	1.63	1.00	0.98	1.00	6.00
New York City	1.95	2.00	1.19	1.00	10.00
Philadelphia	N/A	N/A	N/A	N/A	N/A
Prince George's	N/A	N/A	N/A	N/A	N/A
Providence	N/A	N/A	N/A	N/A	N/A
Brooklyn Lab	N/A	N/A	N/A	N/A	N/A

Table A.23—Continued

	Graduation				
	Mean	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.87	1.00	0.34	0.00	1.00
Denver					
New York City	0.77	1.00	0.42	0.00	1.00
Philadelphia	0.79	1.00	0.41	0.00	1.00
Prince George's	N/A	N/A	N/A	N/A	N/A
Providence	N/A	N/A	N/A	N/A	N/A
Brooklyn Lab	N/A	N/A	N/A	N/A	N/A

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

District	**ObD % (*N*)**	**Others % (*N*)**
Cleveland	54 (65)	45 (954)
Denver	25 (21)	15 (1,238)
New York	69 (49)	47 (34,347)
Philadelphia	60 (58)	47 (3,565)
Providence	100 (47)	100 (677)

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

District	**Mean**	**Median**	**SD**	**Minimum**	**Maximum**
Cleveland	0.89	1.00	0.32	0.00	1.00
Denver	0.59	1.00	0.49	0.00	1.00
New York City	0.71	1.00	0.45	0.00	1.00
Philadelphia	0.89	1.00	0.31	0.00	1.00
Providence	0.83	1.00	0.37	0.00	1.00

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

		District						
Outcome	**Overall**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Math	–0.170*	–0.173**	0.016	–0.275***	–0.146***	–0.524***	–0.001	–0.018
	0.081	0.062	0.082	0.019	0.025	0.016	0.068	0.102
ELA	–0.050	–0.119**	–0.073	0.021	0.005	–0.049	0.083	–0.155**
	0.029	0.044	0.110	0.045	0.028	0.046	0.193	0.056
Attendance	–0.027	0.460***	0.069	–0.150***	–0.209***	–0.075***		–0.001
	0.038	0.088	0.040	0.013	0.031	0.005		0.048
Suspension	–0.029	–0.258***	0.107*	–0.171***	0.085**	–0.039***		0.132**
	0.050	0.031	0.049	0.019	0.032	0.009		0.042
PSAT	–0.030	–0.024	–0.049		–0.100**	0.038*		–0.029
	0.029	0.035	0.041		0.031	0.018		0.045
SAT	–0.076**	–0.090**	–0.013		–0.175***	–0.043**		–0.052
	0.025	0.033	0.097		0.047	0.013		0.049
Credits	–0.272**	0.099*	–1.655***	0.060**		0.004		–0.121
	0.104	0.045	0.116	0.022		0.011		0.104
AP course	0.003				0.244	–0.193***		
	0.217				0.144	0.029		
Graduation	0.018	0.011				0.182***		–0.153**
	0.106	0.054				0.022		0.054

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

District	Estimate	Standard Error
Overall	0.007	0.106
1	0.229*	0.090
2	0.054	0.147
4	–0.428*	0.120
5	0.198***	0.012
7	–0.086	0.082

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

	Mathematics		Reading	
District	Estimate	Standard Error	Estimate	Standard Error
Overall	0.053	0.047	0.041	0.069
1	–0.185	0.088	–0.844***	0.067
2	0.011	0.080	–0.056	0.174
3	0.112***	0.012	0.036**	0.011
5	0.032	0.048	0.107	0.075
6	0.105**	0.030	0.010	0.024
7	0.031	0.028	0.156***	0.028

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.[5] 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

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

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,f] We did not conduct classroom observations or collect artifacts in spring 2018.

Annual Spring School Visits

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

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

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

[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

		2014–2015				2015–2016				2016–2017				2017–2018			
		Fall		Spring		Fall		Spring		Fall		Spring		Fall		Spring	
Cohort	School	*N*	RR (%)	*N*	RR (%)	*N*	RR (%)	*N*	RR (%)	*N*	RR (%)	*N*	RR (%)	*N*	RR (%)	*N*	RR (%)
I	J	91	83	76	71	150	82	150	87	183	91	152	94	212	91	112	51
	P	109	92	97	78	138	95	141	87	149	88	138	93	186	76	97	44
	H	54	53	78	98	137	98	152	94	198	85	123	59	222	71	184	62
	I	105	87	87	81	186	95	128	70	208	84	212	90	222	89	209	90
	C	78	79	71	86	139	94	118	86	137	64	140	65	197	73	173	72
II	K	—	—	—	—	88	90	82	88	133	81	151	95	198	88	171	80
	O	—	—	—	—	92	96	89	100	163	94	159	95	192	89	168	81
	G	—	—	—	—	91	100	85	99	169	92	170	99	243	96	198	78
	D	—	—	—	—	67	96	63	100	129	90	111	80	168	92	173	94
	N	—	—	—	—	70	95	62	91	106	69	112	72	135	80	133	81
III	F	—	—	—	—	—	—	—	—	95	94	71	93	187	94	163	93
	A	—	—	—	—	—	—	—	—	121	95	109	83	148	91	149	92
IV	E	—	—	—	—	—	—	—	—	—	—	—	—	115	93	119	97
	B	—	—	—	—	—	—	—	—	—	—	—	—	82	92	80	94
	L	—	—	—	—	—	—	—	—	—	—	—	—	98	87	99	85
	M	—	—	—	—	—	—	—	—	—	—	—	—	90	87	83	81
Total		**437**	**79**	**409**	**81**	**1,161**	**93**	**1,070**	**88**	**1,791**	**85**	**1,648**	**84**	**2,695**	**86**	**1,918**	**85**

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

Type of Task	Percentage of Students Who Responded			
	Never or Hardly Ever	Sometimes	Often	Daily or Almost Daily
Memory-based or repetitive tasks, like solving a page of equations, filling in the blanks, or taking a short quiz	9	34	40	17
Basic comprehension tasks, like answering short questions or writing a summary based on what you learned	5	29	45	21
Connecting or applying what you learned to real-world problems	10	29	41	21
Projects where you worked with other students over more than one class period	7	32	41	21

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

Statement	Percentage of Students Who Responded			
	Strongly Disagree	Disagree	Agree	Strongly Agree
The assignments the teachers give have a clear purpose; I know how they help me to learn what I'm supposed to learn.	5	12	65	18
The instructions the teachers give about what I need to do for each assignment are clear.	4	14	64	18
The way the teachers assess my work is fair.	4	14	63	18
Based on the feedback teachers give me on my work, I know how to improve in future assignments.	5	10	61	25

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

	Percentage of Students Who Responded			
Statement	**Strongly Disagree**	**Disagree**	**Agree**	**Strongly Agree**
I pay attention and try not to get distracted when I'm doing schoolwork.	4	13	62	21
I make time to do schoolwork outside of class time.	6	20	57	17
I try to do well on my schoolwork even when it isn't interesting to me.	4	10	60	26

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

	Percentage of Students Who Responded			
Statement	**Never**	**Once in a While**	**Usually**	**Always**
Complete all the classwork and activities you are assigned	4	14	56	26
Submit your work by the due date	3	21	54	21
Submit your work with enough time to get feedback and make changes before it is due	5	26	48	21
Revise or make changes to your work when given the opportunity to improve on it	3	16	50	30

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

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

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

	Percentage of Students Who Responded				
Classroom Experience	Never	Rarely	Sometimes	Most of the Time	Always
I have opportunities to choose which instructional materials (such as books or computer software) I use in class.	7	11	34	31	18
I have opportunities to choose which topics I focus on in class.	7	13	36	30	14
I am required to show that I understand a topic before I move onto a new topic.	4	8	33	36	19
I work on different topics or skills than what my classmates are working on at the same time.	6	11	36	32	16
I am given the chance to work through instructional material at a faster or slower pace than other students in this class.	5	8	34	35	19
When I am working on an assignment or activity, I know what the goals of the assignment or activity are.	3	6	32	38	21
I keep track of my learning progress using technology (for example, by using an online gradebook or portfolio).	4	7	30	35	24
I have opportunities to review or practice new material until I really understand it.	3	7	30	37	23

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

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

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.

Table D.8.
Extent to Which Students' Teachers Addressed Social and Emotional Learning Topics in the Classroom

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

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

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

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.

Table D.10.
Scales Included in Spring 2017–2018 Regression Analyses

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

	Percentage of ObD Teachers Who Responded			
Working Condition	**Strongly Disagree**	**Disagree**	**Agree**	**Strongly Agree**
Teachers at my school engage in regular, productive conversations with one another about how to improve instruction.	4	11	46	39
Teachers at my school are highly focused on improving student learning.	1	6	54	39
Teachers at my school have the skills needed to foster meaningful student learning.	2	8	59	31
Teachers at my school support each other in their efforts to improve student learning.	2	5	60	34
Teachers at my school really believe every child can learn and be college and career ready.	2	12	56	29
Teachers at my school are fully prepared to teach based on the Common Core State Standards, Next Generation Science Standards or other relevant standards.	1	11	66	21
I feel respected as a colleague by staff members at my school.	2	8	47	43
If I had concerns about my school, I would feel comfortable raising them with administrators at the school.	12	14	36	38
Administrators at my school are highly *supportive of teachers*.	15	11	41	33
Administrators at my school are highly *focused on student learning*.	13	12	43	33
Administrators at my school trust teachers to make decisions about their own instruction.	12	11	35	41
Students in this school respect *one another*.	8	29	51	12
Students in this school respect *the school staff*.	9	27	50	14
Students in this school are motivated to achieve.	9	28	55	9

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 E.2.
Conditions That Obstructed Teachers' Efforts to Personalize Student Learning

	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
Condition	**N/A**	**Not an Obstacle**	**Minor Obstacle**	**Major Obstacle**	**N/A**	**Not an Obstacle**	**Minor Obstacle**	**Major Obstacle**
Too much diversity in achievement levels among my students	0	35	39	26	13	34	37	17
Lack of flexibility in the curriculum I am required to teach (i.e., need to teach specific material in a specific time frame)	0	38	37	25	31	30	26	13
Pressure to cover specific material as a result of state or district standards or testing requirements	0	35	33	33	13	26	37	23
Excessive amounts of time I need to spend developing content that meets individual students' needs	0	29	43	28	14	30	35	21
Inadequate opportunities to participate in professional development related to personalizing learning	0	37	44	19	31	33	26	11
Inadequate data to help me personalize instruction for students	0	46	44	11	37	31	26	6
Lack of high-quality content or materials	0	42	39	19	37	26	29	8
High levels of student *absenteeism* **	0	11	37	52	13	22	36	29
High levels of student *disciplinary problems* **	0	20	36	44	23	30	30	17
Scheduling constraints	0	25	50	25	12	28	41	20

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

	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
Type of Student Performance Information	**Never**	**Once a Year**	**A Few Times per Year**	**Approx. Monthly**	**Never**	**Once a Year**	**A Few Times per Year**	**Approx. Monthly**
Scores on district or state assessments in *mathematics***	13	16	53	7	18	44	28	5
Scores on district or state assessments in *language arts***	9	15	53	9	16	45	30	4
Scores on district or state assessments in *science*	31	24	31	8	24	47	21	3
Scores on district or state assessments in subjects other than mathematics, language arts, or science**	31	18	34	5	39	30	21	4
Information about student performance on specific concepts or skills**	13	4	19	15	23	19	22	10
Assessment data that are built into curriculum software**	31	7	22	7	43	13	18	9
Identification of specific students who *need extra assistance***	6	7	18	19	8	15	37	13
Identification of specific students who *have achieved mastery***	19	4	16	12	31	21	20	9
Youth development outcomes (for example, student behavior, attitudes, or motivation)**	18	14	19	10	40	13	19	10
Identification of specific students who are at risk of dropping out or not progressing to the next grade**	13	12	24	17	22	16	35	12

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.

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

Type of Student Performance Information	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
	A Few Times per Month (%)	Approx. Weekly (%)	A Few Times per Week (%)	At Least Daily (%)	A Few Times per Month (%)	Approx. Weekly (%)	A Few Times per Week (%)	At Least Daily (%)
Scores on district or state assessments in *mathematics***	6	3	2	1	3	2	1	1
Scores on district or state assessments in *language arts***	8	3	3	1	3	2	0	1
Scores on district or state assessments in *science*	4	1	2	0	3	2	1	1
Scores on district or state assessments in subjects other than mathematics, language arts, or science**	6	4	2	1	3	2	1	1
Information about student performance on specific concepts or skills**	13	14	11	11	9	8	5	4
Assessment data that is built into curriculum software**	8	10	6	10	6	7	3	2
Identification of specific students who *need extra assistance***	16	14	9	12	11	9	5	3
Identification of specific students who *have achieved mastery***	14	17	8	10	8	6	3	2
Youth development outcomes (for example, student behavior, attitudes, or motivation)**	13	13	7	7	8	6	2	1
Identification of specific students who are at risk of dropping out or not progressing to the next grade**	12	10	7	6	7	6	2	1

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.

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

Purpose of Data Use	Percentage of ObD Teachers Who Responded					Percentage of Teachers Nationally Who Responded				
	I Don't Do This	I Do This Without Data	Small Extent	Moderate Extent	Large Extent	I Don't Do This	I Do This Without Data	Small Extent	Moderate Extent	Large Extent
Tailoring the *pace* of instruction to individual students' needs**	7	7	13	42	31	13	20	24	26	17
Tailoring the *content* of instruction to individual students' needs**	8	7	14	33	38	12	21	25	28	15
Tailoring the *instructional strategies* to individual students' needs**	7	7	13	38	35	8	21	24	30	17
Developing recommendations for tutoring or other educational support services for particular students**	12	9	21	35	22	21	13	23	29	14
Assigning or reassigning students to groups within my class(es)**	11	8	13	31	38	18	18	22	27	15
Allowing students to skip *units or lessons* if they've demonstrated mastery of the content in some other way**	36	7	17	20	20	64	6	13	12	6
Allowing students to skip *courses or grades* if they've demonstrated mastery of the content in some other way**	65	3	9	15	8	76	5	7	7	5
Recommending students for extended learning opportunities (for example, extended-day programs, Saturday classes, or an extended school year)**	23	10	18	26	23	47	10	18	17	9
Identifying topics requiring more or less emphasis in instruction**	7	6	16	40	32	8	17	24	32	19
Identifying areas where I need to strengthen my content knowledge or teaching skills**	4	13	15	41	27	6	22	24	31	18
Reflecting on and discussing teaching and learning with other teachers**	6	7	18	39	30	7	21	28	29	15
Reflecting on and discussing *learning with my students***	4	6	15	41	34	7	21	23	31	18

Purpose of Data Use	Percentage of ObD Teachers Who Responded					Percentage of Teachers Nationally Who Responded				
	I Don't Do This	I Do This Without Data	Small Extent	Moderate Extent	Large Extent	I Don't Do This	I Do This Without Data	Small Extent	Moderate Extent	Large Extent
Providing college/career advice or guidance**	12	18	26	29	15	18	28	24	21	10
Revising our school's strategic goals or educational plan**	39	11	10	25	15	50	10	16	15	9

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.

Table E.5.
Teachers' Assessment of Their Schools' Data System

Statement	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
	Strongly Disagree	Disagree	Agree	Strongly Agree	Strongly Disagree	Disagree	Agree	Strongly Agree
I have access to high-quality data that help me adapt the pace, content, or strategies of instruction to meet students' needs.**	6	20	53	21	13	34	43	11
I can use the school's data system to easily produce the views or reports I need.*	8	27	44	21	13	29	44	13
Our school's data system provides real-time data that are actionable.**	8	18	47	26	15	31	42	12
I have the necessary skills and experience to use data to guide my instruction.**	2	8	48	42	5	13	57	25
Our school's data system and assessments enable me to make good decisions about mastery-based progression for individual students (i.e., using data to determine whether students have mastered a set of competencies and should move on to new material).**	8	21	47	25	15	33	42	11
Our school's data system and assessments provide adequate information about students' progress toward specific learning objectives.**	7	22	52	19	13	34	44	9
Students have access to information from data systems that track their progress on particular tasks, skills, or for the course overall.**	5	17	47	31	16	26	45	12
Students regularly review data on their own progress using the school's data system.**	13	25	38	24	24	32	36	8

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

	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
Practice	**None**	**Small Extent**	**Moderate Extent**	**Large Extent**	**None**	**Small Extent**	**Moderate Extent**	**Large Extent**
Students have opportunities to choose what instructional materials (such as books or computer software) they use in class.**	15	27	39	19	29	35	25	12
Students have opportunities to choose what topics they focus on in class.**	16	32	37	15	38	34	20	8
I frequently adapt course content to meet students' needs by providing additional assignments, resources, and activities for remediation or enrichment.**	2	14	43	41	6	31	36	27
I clearly present the goal or objective for each assignment.	2	11	37	51	2	12	31	55
I have adopted strategies that allow students to keep track of their own learning progress.**	3	18	46	33	16	30	36	19
I require students to demonstrate mastery of a topic before they can move onto a new topic.**	9	28	46	16	25	34	29	13
Different students work on different topics or skills at the same time.**	8	16	44	32	25	32	27	15
I give students the chance to work through instructional material at a faster or slower pace than other students in this class.**	3	15	42	40	14	26	37	22
When students are working on an assignment or activity, they know what the goals of the assignment or activity are.	2	9	45	45	2	13	36	50
Students keep track of their own learning progress using technology (for example, by using an online gradebook or portfolio).**	7	13	38	42	15	23	27	35
Students have opportunities to review or practice new material until they fully understand it.**	3	12	45	39	5	27	40	29
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.**	5	18	46	31	10	24	44	23

Table E.6—Continued

Practice	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
	None (%)	Small Extent (%)	Moderate Extent (%)	Large Extent (%)	None (%)	Small Extent (%)	Moderate Extent (%)	Large Extent (%)
I frequently regroup students for instruction to address changing learning needs and interests.**	8	23	35	34	17	31	32	20
Students are able to access instructional materials both in and outside of the classroom.**	1	11	37	52	5	14	34	47
I provide a variety of materials or instructional approaches to accommodate individual needs and interests.**	1	13	46	39	4	20	45	31
I connect what students are learning with experiences they have throughout the rest of the school day or outside of school.*	3	16	43	38	6	23	37	34

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

Practice	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
	None	Small Extent	Moderate Extent	Large Extent	None	Small Extent	Moderate Extent	Large Extent
I connect academic content to themes related to social and emotional learning (e.g., by discussing resilience in the context of a character in a book or emphasizing perseverance or time management in solving math problems).**	7	13	36	44	11	20	34	36
I greet each student by name when they come to class.**	1	7	19	74	6	16	29	50
I actively establish one-on-one relationships with each of my students.**	1	2	23	74	1	12	33	55
I give students opportunities to display their nonacademic talents in class.**	3	23	39	34	9	31	31	29

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

Topic	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
	None	Small Extent	Moderate Extent	Large Extent	None	Small Extent	Moderate Extent	Large Extent
Collaborating with other students**	0	4	36	59	1	12	37	51
Understanding and managing emotions**	3	16	38	43	8	25	36	32
Handling stress	3	20	39	38	5	23	40	32
Persisting through challenging tasks**	1	3	33	63	1	9	36	55
Making responsible decisions*	0	6	28	66	1	8	31	59
Establishing and maintaining positive relationships*	2	6	32	60	3	9	35	53
Feeling and showing empathy for others**	1	10	32	56	3	16	37	44
Learning mindsets (e.g., growth mindset, sense of purpose and belonging)**	2	12	30	56	8	16	38	38
Communicating their thoughts and emotions**	2	11	38	49	4	19	40	37
Developing a sense of identity	6	16	35	44	6	20	39	35

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 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

I have adequate access to curriculum materials (whether technology- or non–technology-based) that . . .	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
	Strongly Disagree	Disagree	Agree	Strongly Agree	Strongly Disagree	Disagree	Agree	Strongly Agree
are of high quality.	5	14	52	28	3	16	59	22
address the learning needs of all of my students.	7	23	46	24	4	18	59	19
contribute to my efforts to promote college and career readiness.	4	16	54	26	4	20	56	21
are accessible to students outside of the school building and outside of the regular school day.	5	23	42	29	5	18	51	26
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).*	6	21	46	27	6	25	51	18

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 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

Response Option	Percentage of ObD Teachers
None; I don't receive any materials from my school or district	16
Fewer than half	31
About half	22
More than half	21
All of the materials	10

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

Response Option	Percentage of ObD Teachers
None; I don't search for supplemental materials	2
Fewer than half	29
About half	32
More than half	27
All of the materials	11

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

Response Option	Percentage of ObD Teachers
None; I don't create original material	1
Fewer than half	27
About half	19
More than half	39
All of the materials	14

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**

Response Option	Percentage of ObD Teachers	Percentage of Teachers Nationally
Less than 1 hour	0	3
1–2 hours	6	17
3–4 hours	28	33
5–6 hours	24	25
7–8 hours	19	10
9–10 hours	12	5
More than 10 hours	12	6

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

	Percentage of ObD Teachers Who Responded			
Task	Never or Hardly Ever	Sometimes	Often	Daily or Almost Daily
Practice for building fluency of procedural skills (e.g., often via worksheets, quizzes, or drill-and-practice software)*	9	28	43	20
Basic comprehension or recall of facts (e.g., short answer questions, summary)	6	19	43	32
Application of knowledge or skills with real-world connections	1	26	43	31
Culminating projects and performance tasks that provide opportunities for deeper learning, application of skills and content, and multiple formats for demonstrating student mastery	1	52	27	21

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

	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
Practice	**Never or Hardly Ever**	**Sometimes**	**Often**	**Always or Almost Always**	**Never or Hardly Ever**	**Sometimes**	**Often**	**Always or Almost Always**
I moved students along to learn other topics and skills or to do other tasks.	13	41	32	13	12	44	33	11
I gave students a task that is similar in complexity to work on, to reinforce the content or skills.	4	32	43	21	6	32	44	18
I gave students a simpler task that covers the same content or skill.**	10	40	35	16	19	43	27	11
I retaught or reviewed the content or skills.*	1	17	49	34	2	27	42	30
I reviewed the student work, and the way it was assessed, with students.	2	26	41	31	4	25	44	27
I worked step-by-step with students as they revised the task or worked on similar tasks.**	3	19	49	30	6	30	41	24
I provided students with samples or models of finished work for the task they were assigned.**	5	21	39	36	9	27	37	27
I placed students in groups to receive support with learning the content or skills.**	7	27	40	26	16	31	36	17
I pointed students to materials or resources to practice or review on their own.**	6	25	40	29	9	32	40	20

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

Description	Percentage of ObD Teachers Who Responded				Percentage of Teachers Nationally Who Responded			
	Never or Hardly Ever	Sometimes	Often	Always or Almost Always	Never or Hardly Ever	Sometimes	Often	Always or Almost Always
When starting on a new topic or competency, I first identify students' prior knowledge and skills with a diagnostic assessment or task.*	6	23	38	33	8	29	37	26
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).	4	22	47	27	3	17	52	29
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.**	3	21	37	39	18	26	36	20
Tasks that are assessed for mastery closely resemble tasks that students have already seen or experienced.	2	11	53	34	1	14	46	39
Tasks that are assessed for mastery require students to apply knowledge and skills they have learned to a new problem or context.*	0	11	46	43	2	11	52	35
Tasks that are assessed for mastery of a given competency differ in difficulty, depending on the student's ability level.**	4	22	41	33	19	31	34	16
Mastery is assessed as what students can accomplish mostly independently (i.e., without extensive peer support or scaffolds).*	5	13	52	30	3	12	46	39
Students have multiple opportunities, throughout a unit or throughout the year, to demonstrate mastery of certain content and skills.**	0	3	32	66	0	10	46	45
Students can make multiple attempts at a given task that counts toward mastery.**	1	7	27	65	5	22	39	35
Students are considered to have achieved mastery when they have completed the expected number of tasks.	14	31	40	16	20	28	39	13
Student work counts toward mastery only when it reaches an adequate performance level.	7	23	38	32	8	23	48	21
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).	2	14	45	39	3	15	54	29

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

Bang, Heejung, and James M. Robins, "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, Vol. 61, No. 4, 2005, pp. 962–973.

Benjamini, Yoav, and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 289–300.

Borenstein, Michael, Larry Hedges, and Hannah Rothstein, *Meta-Analysis: Fixed Effect vs. Random Effects*, Englewood, N.J.: Biostat, 2007. As of July 9, 2020: https://www.meta-analysis.com/downloads/M-a_f_e_v_r_e_sv.pdf

Center for Evaluation, Policy, and Research, *High School Survey of Student Engagement*, Bloomington, Ind.: Indiana University Bloomington, 2013.

Hullsiek, Katherine Huppler, and Thomas A. Louis, "Propensity Score Modeling Strategies for the Causal Analysis of Observational Data," *Biostatistics*, Vol. 3, No. 2, 2002, pp. 179–193.

Iacus, Stefano M., and Gary King, *How Coarsening Simplifies Matching-Based Causal Inference Theory*, Milan, Italy: University of Milan, Department of Economics, Business and Statistics, 2012.

Kalaian, Sema A., "Meta-Analysis Methods for Synthesizing Treatment Effects in Multisite Studies: Hierarchical Linear Modeling (HLM) Perspective," *Practical Assessment, Research, and Evaluation*, Vol. 8, No. 1, Article 15, 2002.

Ma, Lingling, and John Cronin, *Evaluating the Effect of Random Selection on Virtual Comparison Group Creation*, Portland, Ore.: Northwest Evaluation Association, 2009. As of October 22, 2020: https://files.eric.ed.gov/fulltext/ED521965.pdf

Nordmann, Alain J., Benjamin Kasenda, and Matthias Briel, "Meta-Analyses: What They Can and Cannot Do," *Swiss Medical Weekly*, Vol. 142, 2012.

Pane, John F., Elizabeth D. Steiner, Matthew D. Baird, and Laura S. Hamilton, *Continued Progress: Promising Evidence on Personalized Learning*, Santa Monica, Calif.: RAND Corporation, RR-1365-BMGF, 2015. As of June 16, 2020: http://www.rand.org/pubs/research_reports/RR1365.html

Pustejovsky, James E., and Elizabeth Tipton, "Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models," *Journal of Business & Economic Statistics*, Vol. 36, No. 4, 2018, pp. 672–683.

Ridgeway, Greg, Daniel F. McCaffrey, Andrew R. Morral, Lane F. Burgette, and Beth Ann Griffin, *Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial for the R TWANG Package*, Santa Monica, Calif.: RAND Corporation, TL-136/1-NIDA, 2014. As of June 12, 2020: https://www.rand.org/pubs/tools/TL136z1.html

Robbins, Michael W., and David Grant, *RAND American Educator Panels Technical Description*, Santa Monica, Calif.: RAND Corporation, RR-3104-BMGF, 2020. As of May 21, 2020: https://www.rand.org/pubs/research_reports/RR3104.html

Steiner, Elizabeth D., Laura S. Hamilton, Laura Stelitano, and Mollie Rudnick, *Designing Innovative High Schools: Implementation of the Opportunity by Design Initiative After Two Years*, Santa Monica, Calif.: RAND Corporation, RR-2005-CCNY, 2017. As of May 21, 2020: https://www.rand.org/pubs/research_reports/RR2005.html

Stuart, Elizabeth A., "Estimating Causal Effects Using School-Level Data Sets," *Educational Researcher*, Vol. 36, No. 4, 2007, pp. 187–198.

University of Chicago Consortium on School Research, *High School Teacher Survey*, Chicago, Ill., 1999.

———, *2015 CPS My Voice, My School Student Survey, 6th–12th Grade Version*, Chicago, Ill., 2015.

What Works Clearinghouse, *WWC Procedures and Standards Handbook, Version 4.0*, Washington, D.C.: National Center for Education Evaluation, 2018. As of July 9, 2020: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

WWC—*See* What Works Clearinghouse.