JULY 2017

Informing Progress

John F. Pane
Elizabeth D. Steiner
Matthew D. Baird
Laura S. Hamilton
Joseph D. Pane

Insights on Personalized Learning Implementation and Effects

RAND CORPORATION

Funded by BILL & MELINDA GATES foundation

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.

BILL & MELINDA GATES *foundation*

Guided by the belief that every life has equal value, the Bill & Melinda Gates Foundation works to help all people lead healthy, productive lives. In developing countries, it focuses on improving people's health and giving them the chance to lift themselves out of hunger and extreme poverty. In the United States, it seeks to ensure that all people—especially those with the fewest resources—have access to the opportunities they need to succeed in school and life. Based in Seattle, Washington, the foundation is led by CEO Susan Desmond-Hellmann and Co-chair William H. Gates Sr., under the direction of Bill and Melinda Gates and Warren Buffett.

RR-2042-BMGF

This report is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

PHOTO CREDITS | Cover: PeopleImages/DigitalVision/Getty Images; page 5: JBryson/iStock/Getty Images Plus; page 7: Ableimages/DigitalVision/Getty Images; page 8: Tyler Olson/Fotolia; page 19: asiseeit/E+/Getty Images; page 31: kali9/E+/Getty Images; page 32: WavebreakMediaMicro/Fotolia; page 34: PeopleImages/DigitalVision/Getty Images; page 38: STEEX/E+/Getty Images; page 39: Rob/Fotolia; page 42: FatCamera/E+/Getty Images

Table of Contents

Introduction

The Bill & Melinda Gates Foundation engaged the RAND Corporation to carry out a study of foundation-funded schools that are implementing personalized learning (PL). This is the third in a series of reports focused on PL school design characteristics, teacher and student perceptions, and student achievement.

The basic concept of PL—instruction that is focused on meeting students' individual learning needs while incorporating their interests and preferences—has been a longstanding practice in U.S. K–12 education. Examples include individualized education plans for students with special needs, the use of data to make instructional decisions for individuals or small groups, the use of support teachers and tutors, individual or group projects, and diverse elective course offerings.

More recently, options for personalization have increased as personal computing devices have become more affordable and available in schools and developers have created software products that can support individual student learning. Much of this work was inspired by Bloom's (1984) article showing that human tutors providing individualized instruction to students can produce large achievement gains relative to whole-class instruction. In the context of a review of hundreds of studies of human and computer-based tutoring, VanLehn (2011) made the important observation that mastery learning principles used by the tutors in Bloom's (1984) article may account for a large part of their positive effect. The studies in VanLehn's meta-analysis were not all conducted in K–12 schools, and many did not produce statistically significant results. Nonetheless, he found that systems that emulate the interactions of a human tutor tended to produce positive achievement results. Some of these systems have undergone rigorous evaluation in K–12 schools with positive results (Brodersen and Melluso, 2017).

__Personalized learning__ prioritizes a clear understanding of the needs and goals of each individual student and the tailoring of instruction to address those needs and goals. These needs and goals, and progress toward meeting them, are highly visible and easily accessible to teachers as well as students and their families, are frequently discussed among these parties, and are updated accordingly.

For the most part, all of these personalization efforts have been implemented within schools and classrooms that otherwise retain a traditional model of large-group instruction to groups of roughly 20–30 similar-age students. However, in recent years, it has become more common for schools to embrace schoolwide models of PL that depart more radically from typical practice. These schools seek to allow what and how a student learns on a daily basis to be less constrained by the needs of other students or by external requirements for grade-level content coverage; to be driven largely by the individual student's needs, interests, and context; and to be informed by ongoing conversations with the student and the adults in his or her life. In service of these objectives, the staffs at these schools are implementing a range of interconnected strategies in novel ways.

In this report, we:

- explore what PL looks like in a small sample of schools that have been focused on implementing PL approaches schoolwide
- consider how the approaches to personalization in these schools compare to a national sample that represents more-typical practice in the United States
- briefly discuss obstacles to PL implementation
- discuss how PL implementation differs between charter schools and traditional district schools in our sample, and what factors seem to support or hinder implementation
- describe how achievement growth for students in these schools differs from growth for similar students in other schools
- discuss implications for policymakers, implementers, and funders.

We collected data from the schools that received funding from the Next Generation Learning Challenges (NGLC) initiative in the Wave IIIa and Wave IV launch grants.[1] This initiative was intended to support the development of schools that took a highly personalized approach to learning. Our study began in fall 2012 and concluded in spring 2015.

The schools participating in the NGLC program were not expected to implement a specific PL intervention. Although there were some general requirements, such as allowing students to learn at varying rates, technology-enabled learning, and incorporating flexibility in the learning environment, each school had the flexibility to implement a PL model that would work best with its context, students, and goals. Of course, educational interventions are often enacted differently in practice than how they are described in theory. Therefore, the implementation component of our study seeks to describe how these schools were implementing PL, understand some of the challenges and facilitators, and consider these alongside achievement findings to discern patterns that may be informative.

To learn about implementation, we interviewed school administrators, surveyed teachers and students, and collected instructional logs (brief surveys administered daily to teachers for several weeks during the school year, focusing on instruction in that day's lesson). We also visited some of the schools to interview teachers and students and observe classrooms. The surveys were also administered to a national sample for comparison. Additional information about all of these data collection methods can be found in Appendix A of this report. Student outcome analyses in this study rely on data from the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) mathematics and reading assessments, administered in the fall and spring of each school year in a subset of the NGLC schools.

This study has numerous limitations, and so the findings should be interpreted cautiously. For example,

Key Findings

- **Schools in the NGLC sample were pursuing a wide variety of practices to focus on the learning needs of each individual student in a supportive and flexible way.**
- **Schools were implementing specific PL approaches to varying degrees, with none of the schools looking as radically different from traditional schools as theory might predict.**
- **There is suggestive evidence that greater implementation of PL practices may be related to more positive effects on achievement; however, this finding requires confirmation through further research.**

[1] The NGLC initiative is managed by EDUCAUSE, a nonprofit association dedicated to advancing the use of information technology in higher education, in association with other organizational partners, including the League for Innovation in the Community College, the International Association for K–12 Online Learning, and the Council of Chief State School Officers. NGLC receives primary funding from the Bill & Melinda Gates Foundation, with additional support from the William and Flora Hewlett Foundation, the Eli and Edythe Broad Foundation, and the Michael and Susan Dell Foundation. The initiative supports school districts, charter management organizations, and partner organizations that embrace PL as a means to dramatically increase college readiness rates, particularly among low-income students and students of color. To be considered for funding, these schools applied for a competitive grant. In their applications, schools were required to describe with specificity how their models would support PL. While all of these schools have a high degree of integrated technology as part of their school designs, they vary considerably in the methods and degrees to which they use technology to support PL.

implementation data are limited by their self-report nature and small sample sizes, which can make it difficult to detect differences between groups. Comparisons to national surveys are also limited by an untestable assumption that they represent more-traditional practices, where PL is not being implemented as intensively as in the NGLC schools. The achievement analyses use a research design that does not enable strong causal conclusions. Results can be influenced by, among other risks, selection bias or the implementation of PL in the comparison group. In addition, because it may take a few years for new PL schools to optimize implementation, the results here may not reflect how well the NGLC schools will perform in the future. Readers are encouraged to review the more-detailed discussion of limitations in Appendix B.[2]

[2] Portions of this report are adapted from Pane et al. (2015).

The Sample for Implementation Analyses

The 40 NGLC schools in the implementation sample were predominantly located in urban areas (two were rural) and served large proportions of minority students from low-income families. Many of the schools started out serving a limited range of grades, with plans to expand annually until they reach their full enrollment and grade range. Key sample characteristics include the following, based on 2014–15 school year data provided by school administrators:

- 43 percent of the schools had been implementing PL for one year, 38 percent for two years, and 20 percent for three years
- more than three-quarters of the sample were charter schools
- elementary and K–8 schools averaged about 230 students per school, and middle and high schools averaged about 270 students
- the median schoolwide proportion of students eligible for free or reduced-price lunch (FRL) was 80 percent
- the median schoolwide proportion of students of color was 96 percent.

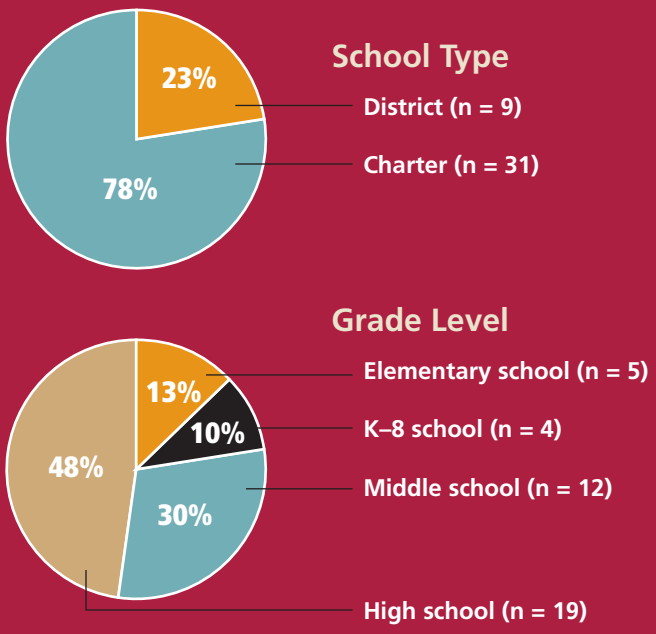
Composition of schools in the implementation analysis

About 10,600 students

40 schools participating in NGLC

6,145 students surveyed

241 teachers surveyed



Note: Percentages may not add to 100 percent due to rounding.

The Sample for Achievement Analyses

Of the 40 NGLC schools in the implementation analysis, only the 32 that administered the MAP assessment are included in the achievement analysis. All of them were relatively new at implementing PL, having started in the 2012–13 academic year or later. Moreover, most of these schools were new schools at the time they began implementing PL. As such, this report discusses schools relatively early in the implementation process. The results reported here are not directly comparable to achievement analyses presented previously in Pane et al. (2015), which focused on a larger sample of somewhat older PL schools that had been operating for at least two years. Only 16 NGLC schools were included in those prior analyses, along with 46 additional schools that were not part of the NGLC program (among the non-NGLC schools, 18 had launched as new schools implementing PL in the same time frame as the NGLC schools; most of the 28 others launched in the prior decade, though we do not know exactly when they started to focus on PL). Table 1 compares the samples used for achievement analyses in the two reports.

Table 1: Comparison of achievement analysis samples in Pane et al. (2015) and this report

	Pane et al. (2015)	This report
Participating initiatives	NGLC and other programs	NGLC only
School experience implementing PL	At least two years	At least one year
Number of PL schools in sample	62	32
Percentage of charter schools in PL school sample	92%	75%
Main achievement analysis	2-year span: 2013–15	1-year span: 2014–15
Approximate number of PL students in main achievement analysis	11,000	5,500
Approximate percentage of PL student sample in		
… grades 9–12	8%	31%
… grades 6–8	23%	48%
… grades K–5	69%	21%

What Is Personalized Learning?

Although there is not yet a widely shared definition of PL, we distilled this working definition from discussions with leading practitioners in the field:

> **Personalized learning prioritizes a clear understanding of the needs and goals of each individual student and the tailoring of instruction to address those needs and goals. These needs and goals, and progress toward meeting them, are highly visible and easily accessible to teachers as well as students and their families, are frequently discussed among these parties, and are updated accordingly.**

This aspiration contrasts with more-traditional instructional approaches, where efforts to meet individual students' needs may take less priority than having students work toward grade-level standards, progress on pace with their grade-level peers, or prepare for grade-level tests at the end of the year. In its ideal form, PL allows for greater variety in what students are working on at any moment, while still setting ambitious goals for each student's progress. The hypothesis, consistent with the research cited above, is that personalized instructional approaches and strategies will improve student outcomes in the short term (e.g., stronger rates of growth in achievement) and in the long term (e.g., successful completion of a postsecondary degree or successful transition into a career).

The best strategies for creating an educational environment that is highly personalized have yet to be identified through research. The NGLC schools in this study were taking a variety of approaches, some of which were extensions of traditional practices, often enhanced by strategic use of technology for instruction and other purposes, and some of which were more-significant departures from common approaches. Each school integrated a set of approaches to create their unique school model. At this early stage in the development of PL, such a diversity of models can be useful to help us learn which strategies and approaches, or combinations thereof, appear to be most important for PL's success. To organize our discussion we group the approaches used by the NGLC schools into four interdependent strategies.

What is the role of technology in PL?

In a variety of ways, technology holds promise to enable personalization to an extent that was not possible at large scale in an earlier era. Technology's greatest role may be to manage the complexity of the personalization process. By occasionally providing instruction or supporting independent learning, technology can also enable educators to take a more personalized approach in their own teaching efforts and other activities they undertake to support student learning and development.

Learner Profiles

A learner profile is a record of each student's individual strengths, needs, motivations, progress, and goals based on data from all available sources. Learner profiles are available not only to teachers, but also to students and their families, and are frequently reviewed, discussed, and updated to inform the student's educational plan.

Personal Learning Paths

Informed by the learner profile, personal learning paths allow for flexibility in the specific paths students take through content to enact their educational plan, while still holding them to high expectations. Within parameters set by teachers, students can make choices about the content or structure of

learning, and the school offers a variety of instructional approaches and curriculum materials, including support for meaningful learning experiences outside of school. Time is available during the school day for one-on-one academic support tailored to students' learning needs, whether for remediation, help with grade-level content, or enrichment.

Competency-Based Progression

Competency-based progression enables personalized paths to run their natural course by removing external constraints on what material each student works on, when, and for how long. Each student's progress toward clearly defined goals is continually assessed, and assessment occurs "on demand" when a student is ready to demonstrate competency. Assessment may take a variety of forms, such as projects or presentations, as well as more-traditional tests or quizzes. A student advances at his or her own pace and earns course credit (if applicable) as soon as he or she demonstrates an adequate level of competency.

Flexible Learning Environments

Flexible learning environments imply that the school adapts the use of resources such as staff, space, and time to best support personalization. For example, elements of the learning space—size, classroom organization, and furniture—are designed to support implementation of PL. The structure of learning time and student grouping strategies are flexible, responsive to student needs, and driven by data where appropriate. Technology is a key aspect of the school model and is available to all students; often schools provide a device to each student.

As we discuss above, the schools approached PL in a variety of ways, and did not necessarily plan to implement every strategy. Rather, they were all working toward the general goal of improving student achievement through PL, and were free to be creative and to adopt approaches compatible with local context and the population of students they served.

In the next section, we use these four strategies as an organizing framework for describing PL implementation in the NGLC schools. For each strategy, we first present a vignette drawn from a school we visited, as an example of relatively strong implementation of the strategy. We then compare NGLC schools to a sample representing schools across the United States with respect to implementing that strategy, discuss some of the challenges NGLC schools reported, and briefly discuss obstacles to PL implementation that cut across the four strategies. Subsequent sections examine whether PL implementation differed between district-operated and charter schools in the NGLC sample. We then turn to an analysis of achievement effects, and discuss implications for policymakers, implementers, and funders.

The Four PL Strategies Are Interrelated

Learner profiles maintain a rich and up-to-date record of student strengths, needs, goals, and progress; that information is used to define personal learning paths, which are appropriate and meaningful choices of material for each student to work on, with the necessary adult supports; competency-based progression enables these personalized paths to run their natural course by removing external constraints on what material each student works on, when, and for how long; and flexible learning environments enable schools to allocate resources in new ways to best support these processes.

What Does Personalized Learning Look Like, and How Does It Differ from Practices in Schools Nationally?

Implementation Analysis Methods

The findings we present in this section comprise a synthesis of the implementation data. The methods we used are described in greater detail in Appendix A, along with a discussion of limitations. To briefly summarize our methods, we used a holistic approach to decide what information to present, focusing on meaningful evidence of differences (or similarities) between practices in the NGLC schools and in other schools nationally. Where we were able to perform tests of statistical significance, we used those results to guide our decisions about what material to present. In some cases, we describe differences that were not statistically significant but that were large in magnitude and qualitatively meaningful in that they shed some light on substantive questions about implementation. We relied heavily on teacher and student survey data because those sources are available for more of the sample and are thus most representative of teachers' and students' attitudes and perceptions.

However, we also made use of interviews with principals and teachers, and focus groups with students; although these sources are less representative than the surveys, they provide a greater depth of information on key aspects of implementation that help to clarify or illuminate patterns we found in the survey data. We triangulated these sources with teacher logs and classroom observation data where applicable.

When we discuss the interview data, we use terms such as "many" and "most" to refer to more than half of interview respondents in the applicable group (e.g., school leaders, teachers, or students) across schools, and we use "several" or "some" to refer to less than half of respondents. Percentages reported here are based on survey results. The vignettes are drawn from the site visit data, and the discussion of implementation challenges is drawn from survey, interview, and focus group data.

Learner Profiles

In this section, we discuss:

- strategies used by students and teachers to track and discuss learning goals
- teachers' use of student achievement data to personalize instruction
- usefulness of school data systems and teachers' access to data
- challenges of using learner profiles.

*A **learner profile** is a record of each student's individual strengths, needs, motivations, progress, and goals based on data from all available sources. Learner profiles are available not only to teachers, but also to students and their families, and are frequently reviewed, discussed, and updated to inform the student's educational plan.*

KEY TAKEAWAYS ON LEARNER PROFILES

NGLC schools showed higher levels of implementation than the national sample in some ways.

- More NGLC teachers reported frequent receipt of high-quality student data and extensive use of the data to personalize instruction.
- More NGLC students reported using technology to track their learning progress.

Schools in the two samples showed similar levels of implementation in other ways.

- Students in the two samples reported similar levels of discussion with teachers regarding their learning progress or learning goals.
- Teachers in the two samples reported similar rates of keeping up-to-date documentation of student strengths, weaknesses, and goals.

There were several challenges to implementation.

- Many NGLC schools struggled to use nonachievement data (e.g., behavior, attendance, socio-emotional skills) to inform instructional decisions and goal setting, in part due to challenges measuring those skills and integrating the results with other student data.
- Data from digital curriculum programs were not always well integrated with other school data systems.

Summary

These findings suggest some important ways in which the NGLC schools exhibited greater access to and use of student data to inform personalized instructional approaches. Both national and NGLC teachers reported receiving and using student data frequently and we did not find differences in the use and characteristics of formal learner profiles. However, NGLC teachers reported receiving many types of student data (e.g., data on students who have achieved mastery or need extra assistance) more frequently, and using them to adjust instruction in ways consistent with PL practices to a greater extent than teachers in the national sample. These differences in data access and use could be related to differences in the schools' data systems, which in NGLC schools seemed more likely to contain student data which facilitated personalized instructional practices. According to principals, barriers to more-extensive use of student data included difficulties measuring nonachievement constructs (e.g., behavior or socio-emotional skills) and integrating such data, along with other data generated by curriculum products, into the school's data system. This made it harder to combine these inputs with achievement data for instructional decisions and goal setting.

Learner profiles: How do NGLC and national practices compare?

More NGLC teachers reported frequent receipt of high-quality student data and extensive use of the data to personalize instruction. NGLC teachers reported that their schools' data systems provided high-quality data useful for informing instruction (e.g., real-time, actionable data, and information about students of varying achievement levels, including students who are far above or below grade level). NGLC teachers were also more likely to report that they had access to high-quality assessment data that helped them adapt the pace or content of instruction to meet students' needs. NGLC teachers, on average, reported receiving student

VIGNETTE: What do learner profiles look like?

School A is an urban charter high school that had been implementing PL for two years at the time of our visit. Students are able to check their grades using the school's learning management system, PowerSchool, as well as their nonachievement data (e.g., behavior, attendance, socio-emotional skills) via SchoolRunner. These programs are accessible to students at home as well as in school and are also accessible to parents. Students reported that these two sources of information about their performance were updated frequently, were useful, and were easily accessible, for example, from their smartphones. Teachers said that they updated PowerSchool frequently—at least daily or weekly. A commitment to making student progress visible and accessible was also evident in posters and charts on classroom walls, which were used to track students' progress toward mastering the college-ready ACT standards. Students and teachers reported that they drew on these multiple sources of data to drive conversations about student progress and set goals.

Students reported that they discussed their grades and behavior with teachers, tracked progress, and set goals during Advisory, a daily time when students met one-on-one with teachers or caught up on classwork. Teachers reported different methods for ensuring that they met with all students during Advisory to discuss their grades and progress. One met weekly with each of her advisees; another let the students initiate meetings, but checked in with each student at least every two weeks. Several students described close relationships with their advisers, as illustrated in the quote.

"With my Rise [Advisory] teacher, I talk to her every single day [about the progress I'm making in school]. I even text her. That's how my bond is with her. And with teachers, I just have two where I constantly check in for my grades and how I'm doing, and they give me feedback on how I'm supposed to do better and I start improving my things better now."

—STUDENT COMMENT ABOUT ADVISORY SUPPORT

achievement and nonachievement data (such as data on student behavior or socio-emotional outcomes) more than a few times a month, versus approximately monthly in the national sample. NGLC teachers also reported using such data to inform and personalize instruction to a greater extent (see Figure 1). However, a majority of teachers in both samples reported that they had plenty of data but needed help translating those data into instructional steps. There were no differences in teachers' reports of how easy their school data systems were to use.

"I would say we're definitely better off in the sense that we're gathering data constantly: their [students'] homework and the assessments in the lessons, the assessment during the projects. It's just ongoing. It's pretty fluid."

—NGLC TEACHER COMMENT ABOUT USING STUDENT ACHIEVEMENT DATA

More NGLC students reported using technology to track their learning progress. However, students in the two samples reported similar levels of discussion with teachers regarding their learning progress or learning goals. Students in the NGLC sample were somewhat more likely to agree that they kept track of their progress using technology (e.g., by using an online gradebook or portfolio) most of the time or always. There were no differences in how often students reported discussing their learning progress with their teachers or working with their teachers to set personal goals for their own learning, but such practices were not widespread.

Teachers in the two samples reported similar rates of keeping up-to-date documentation of student strengths, weaknesses, and goals. When comparing the NGLC survey results with those from the national sample, we found no differences in several key aspects of learner profiles. Similar proportions of teachers reported using frequently updated, shared documents, either paper or electronic (such as learner profiles and learning plans), to document each student's strengths, weaknesses,

and goals. Among teachers who reported using such documents, their characteristics (e.g., whether they exist for all students or were frequently updated), were similar in the two samples.

Challenges of Learner Profiles

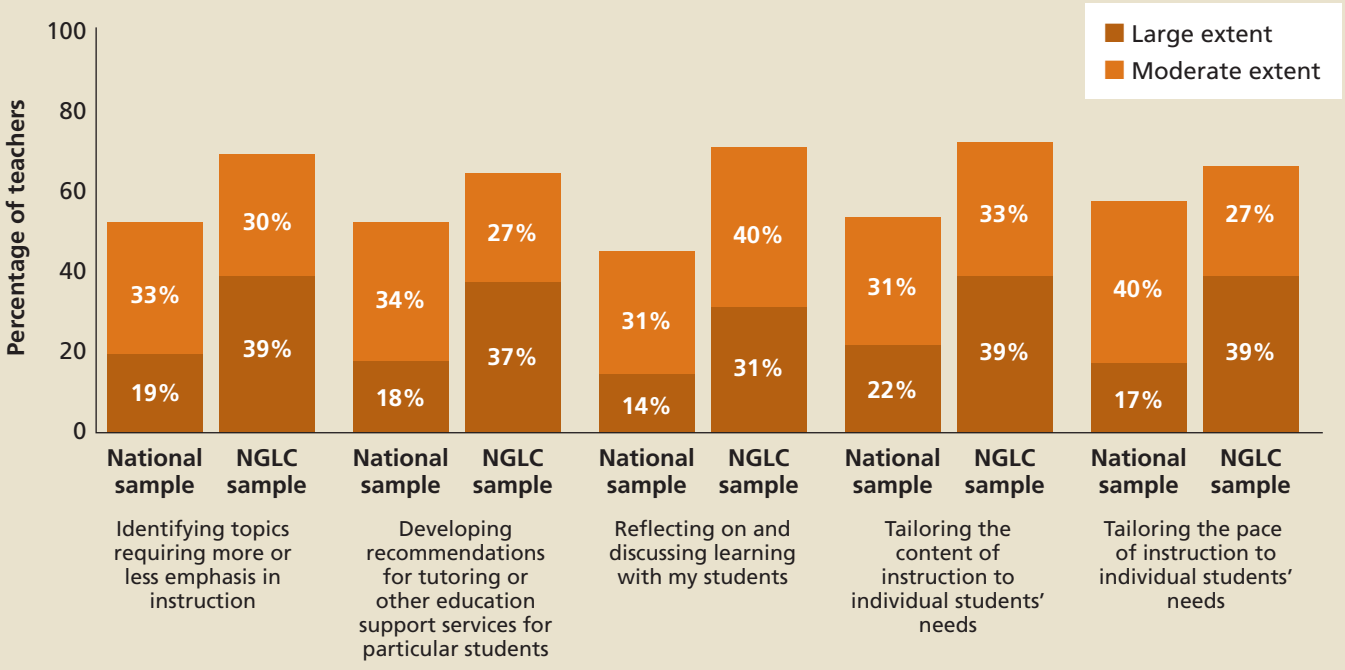
Many NGLC schools struggled to use nonachievement data to inform instructional decisions and goal setting, in part due to challenges measuring those skills and integrating the results with other student data. Using multiple types of student data to inform instruction is a key feature of learner profiles. All NGLC schools collected nonachievement student data, but much of this was done informally. Few schools had robust systems for collecting these data, particularly on socio-emotional skills such as collaboration, critical thinking, or resilience, or using them to inform instructional decisions and understand student progress. Many school administrators told us in interviews that their schools had not yet pulled achievement and nonachievement data together into one cohesive document or system—the data were often tracked in multiple systems. Most of these schools planned to undertake this integration, but had not yet been able to do so.

Data from digital curriculum programs were not always well integrated with other school data systems. Another common challenge, reported by NGLC principals, was that data from the school's various digital curricula and online materials were not well integrated with other data systems (e.g., the learning management system where teachers recorded grades). This increased the burden on teachers who wanted to retrieve and analyze these data.

> *"In terms of nonacademic goals, we teach [students] design thinking, so we do want them to demonstrate the designer skills. And those are things like critical thinking and collaboration. We haven't yet figured out the best ways to measure those . . . "*
>
> —NGLC PRINCIPAL, ON MEASURING NONACHIEVEMENT SKILLS

FIGURE 1

Extent to which teachers used student achievement data



Note: N = 212–214 NGLC teachers; N = 525 national teachers; survey question: "This year, to what extent have you used student achievement/mastery data for each of the following purposes?" Response choices were on a scale of 1 ("My school doesn't do this") to 5 ("Used to a large extent").

Personal Learning Paths

In this section, we discuss:

- the use of a variety of instructional activities, including tailored support
- students' ability to choose topics and instructional materials
- challenges of implementing personal learning paths.

Personal learning paths *allow for flexibility in the specific path students take through content to enact their educational plan, while still holding them to high expectations. Within parameters set by teachers, students can make choices about the content or structure of learning, and the school offers a variety of instructional approaches and curriculum materials, including support for meaningful learning experiences outside of school. Time is available during the school day for one-on-one academic support tailored to students' learning needs, whether for remediation, help with grade-level content, or enrichment.*

KEY TAKEAWAYS ON PERSONAL LEARNING PATHS

NGLC schools implemented more individual support than the national sample.

- NGLC schools appeared to dedicate more time to one-on-one, tailored support of student learning.

Schools in the two groups were more similar on other aspects of implementation.

- Teachers and students in both samples reported relatively low levels of student choice of topics and materials.
- Teachers in both samples reported similar levels of tailoring instruction to student needs, although NGLC students reported slightly higher rates than students nationally.
- Teachers in both groups agreed that their curriculum materials were of high quality.

There were several challenges to implementation.

- NGLC teachers perceived limited time to develop personalized lessons to be the biggest obstacle to implementing personal learning paths.
- Finding high-quality standalone technology-based materials was a challenge.
- Teachers perceived tension between offering student choice and the need to address standards.
- Extensive choice can make student collaboration challenging.

Summary

Most NGLC schools implemented a variety of instructional approaches and focused on one-on-one academic supports tailored to each student's learning needs. Highly personalized approaches, such as flexible paths through content and extensive student choice in the content or structure of learning were not common in either group, most likely because they can be time-consuming for teachers to develop and manage. Teachers also reported that the need to meet standards constrained the amount of choice they could offer to students, which also likely limited implementation of highly personalized approaches.

To implement a variety of instructional approaches, NGLC schools reported adjusting instructional time to focus on coaching and individual supports for students to a greater extent than teachers did in the national sample, a difference that was perhaps facilitated by the fact that the NGLC schools built one-on-one supports for students, such as an advisory period, into the school schedule. In schools that offered choice in path and content, students often worked on different topics and assignments than their peers. While many students enjoyed the flexibility such choices offered, others observed that it made seeking help from (and collaboration with) peers difficult, because students were all working on different things. Teachers in both groups reported that their curriculum materials were of high quality. Interview, log, and survey data suggest that NGLC schools used a combination of standalone tech-based programs and teacher-developed curriculum materials; we do not have comparable data for the national sample.

Personal learning paths: How do NGLC and national practices compare?

NGLC schools appeared to dedicate more time to one-on-one, tailored support of student learning.

Surveyed NGLC teachers reported using individual tutoring, coaching, and support for a greater proportion

VIGNETTE: What do personal learning paths look like?

School B is an urban charter middle school that serves grades 6 through 8 and had been implementing its PL model for three years at the time of our visit. At school B, students could take a flexible, personalized path through content via a "playlist"—a list of a variety of activities (e.g., readings, videos, practice problems, assignments) identified by the instructor and designed to help students learn a particular standard or skill. In these playlists, students were placed at the appropriate level of content based on a standardized test. Students used the playlists to choose which activities to complete as part of their course work, as described in the quote.

Often, students within a class were working on a variety of different standards, and those who were working on a common set of standards were often grouped together for projects or group work. Students were exposed to a variety of instructional approaches, a strategy consistent with personal learning paths. For example, teachers utilized independent work with and without technology, group and independent projects, the playlists, and one-on-one and small-group work with the teacher. Instructional materials included, for example, online curricula, online games, hands-on projects, and textbooks. Teachers had autonomy to vary instructional approaches and materials in their classrooms as needed. Most class periods included time for one-on-one academic support—teachers would confer with some students while others were working—and a 30-minute period at the end of the day was reserved for teachers to work with selected students for additional one-on-one support.

"First, in order to see what standard we're working on we go to Canvas. Canvas shows all of our classes and once you click on the class, it has a list of the standards and [learning levels] students are on. Every [learning level], you go to it and click on the standard you're working on and work on the assignments. Some people are on different [learning levels] so it's based on what they're working on and there are links to activities."

—STUDENT COMMENT ABOUT PLAYLISTS

of the lesson, while teachers in the national sample reported spending a greater proportion of class time on large-group instruction. These results are shown in Figure 2. The administrator interview and site visit data suggest that one-on-one academic support was built into the daily school schedule in all the NGLC schools. In some schools, this took the form of an advisory period where students could receive individualized support from teachers, from peers, or from independent practice. Other schools scheduled "intervention time" as a class period, where students sought help in the subjects in which they were struggling. Still others built one-on-one supports, such as independent practice with teacher support, into each class period. Although we cannot be certain that schools in the national sample did not schedule time for teachers to provide individualized support, the results presented in Figure 2 suggest that there are differences in the structure of learning time between the groups of schools in the two samples.

Teachers and students in both samples reported relatively low levels of student choice of topics and materials. Survey results suggest that student choice in the content and structure of their learning was a feature of both NGLC schools and schools nationally, but we did not find the degree of student choice to be extensive. A majority of teachers in both groups indicated that it was rare for students to choose their own instructional materials or the topic of the class focus (responding that it occurred "not at all" or "to a small extent"). Although NGLC students reported slightly more choice, less than one-third of students in both groups reported that they frequently made their own choices (responding "most of the time" or "always" on items shown in Figure 3). The student focus groups suggested that it was

". . . blended learning is when you have half of the class on the laptop while you're doing lessons and targeting personalized learning. They [students] get personalized learning from the computer and from me. With reading rotations while they are on the computer I get to have my guided reading done while some are in their stations. It helps a lot."

—NGLC TEACHER COMMENT ABOUT VARIETY IN INSTRUCTIONAL APPROACHES

somewhat more common for teachers to offer a choice in how students could complete specific assignments, as illustrated by the student comment.

Teachers in both samples reported similar levels of tailoring instruction to student needs, although NGLC students reported slightly higher rates than students nationally. About two-thirds of teachers in both groups reported that they adapted content and provided a variety of instructional materials to suit individual students' needs to a small or moderate extent. For example, teachers reported using instructional approaches such as teacher-led large- and small-group instruction, individual tutoring, small-group collaboration and projects, or independent practice with and without digital content. NGLC students were slightly more likely than the national sample to report that they had opportunities to learn in different ways within a single lesson, such as listening to the teacher, working in small groups, or working by themselves.

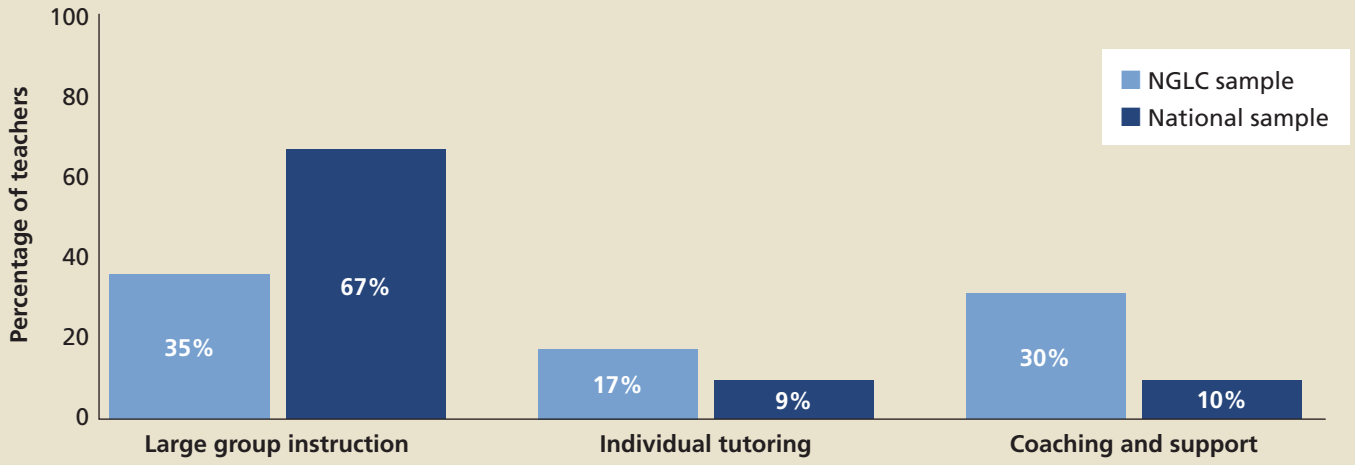
> *". . . it's the same topic, but you can choose to complete the task any way you want to. We just had this assignment called* The Pit and the Pendulum. *And, basically, [the teacher] split it up into two things. After we finished reading it you can compare the book to anything media that compares to the book. . . . And then, also, there was the second part to it where you can build off the story and you can write from the torturer's perspective . . ."*
>
> —NGLC STUDENT COMMENT ABOUT CHOICE

Teachers in both groups agreed that their curriculum materials were of high quality. Access to high-quality curriculum materials is a key support for creating personal learning paths. NGLC teachers agreed with those in the national sample that their curriculum materials were of high quality and met the learning needs of all of their students. These data were collected at a time when teachers nationally were relying heavily on materials that they developed or found on the web as they attempted to align their curricula with new state standards (Opfer, Kaufman, and Thompson, 2016).

Challenges of Personal Learning Paths

NGLC teachers perceived limited time to develop personalized lessons to be the biggest obstacle to implementing personal learning paths. Providing personalized pathways and activities for students is a key feature of personal learning paths, but one that can be time-consuming for teachers. A majority of NGLC teachers reported that "an inadequate amount of time to prepare personalized lessons for all students" was a major or minor obstacle to PL implementation, and half reported that "excessive amounts of time I need to spend developing personalized content" was a major or minor obstacle. Similarly, in site visit and administrator interviews, many NGLC teachers and administrators

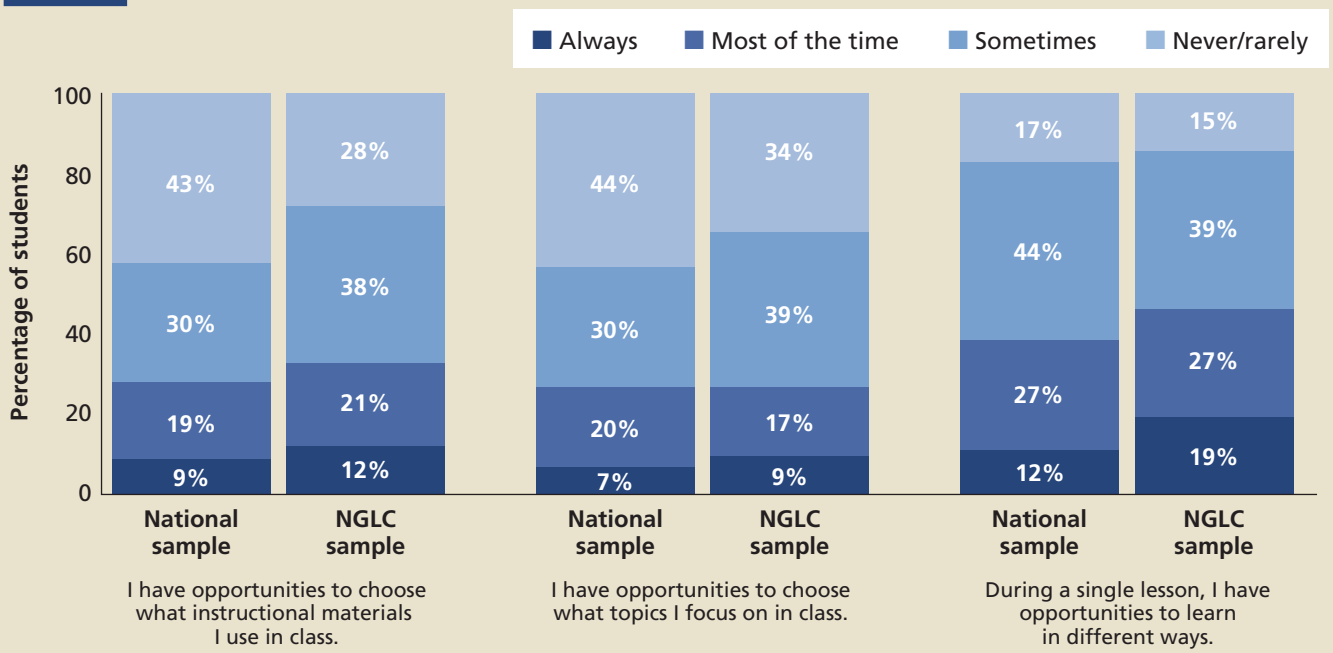
FIGURE 2 Teacher reports of activities used for more than a quarter of class time during a typical lesson



Notes: N = 209–214 NGLC teachers; N = 525 national teachers; survey question: "During a typical class, for what percentage of the time do you utilize the following activities with students?" Respondents wrote percentages for each activity in open-ended text boxes.

FIGURE 3

Student reports of choice in various aspects of instruction



Notes: N = 4,785–4,835 NGLC students; N = 864 national students; survey question: "The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all of your classes in math, English/reading, science, and social studies this year, and mark the response that indicates your typical experience." Response choices were on a five-point scale from 1 ("never") to 5 ("always").

mentioned time as an important obstacle to implementing highly flexible personal learning paths with frequent opportunities for student choice. Clearly, it is time-consuming for teachers to develop personalized lessons in NGLC schools. One theoretical hope is that technology can help make implementation of PL—and personal learning paths in particular—more efficient. While technology has made many features of personal learning paths possible, such as using a playlist, the work of finding (or creating) and organizing high-quality content and assignments often remains in the teachers' hands.

Finding high-quality standalone technology-based materials was a challenge. Survey, log, and interview data suggest that staff in the NGLC schools pieced together their tech- and nontech-based curriculum and instructional materials and used a combination of standalone tech-based programs and teacher-developed material. In interviews, teachers and administrators said they rarely relied on one or two all-encompassing tech-based curriculum products because it was difficult to find ones that were of high quality and effective in their school context. About half of surveyed teachers reported that they often pulled materials from multiple sources or developed them themselves. NGLC teachers reported that they searched for or created about half of their curriculum materials to supplement the curriculum provided to them. Few products were common across schools. Sixty-two different online or digital sources of curriculum materials and assessments were reported across the 40 schools. Only eight of those were mentioned by more than one school, and the two most popular products were mentioned by nine schools each.

Teachers perceived tension between offering student choice and the need to address standards. Although some NGLC schools offered students a high degree of flexibility in the paths they could take through content, in most schools students did not seem to have many opportunities to choose the content or structure of their learning. For instance, several teachers we interviewed reported that offering extensive student choice conflicted with the need to address grade-level standards, and reported that what students learned was dictated by the appropriate standards (i.e., subject matter and grade level) with little variation, thus limiting the extent to which teachers could offer choices to students.

Extensive choice can make student collaboration challenging. Students at some NGLC schools could choose which topics to work on within a given content area and which activities to complete. In focus groups, several students said they generally liked having the flexibility to work on different topics at a different time from their classmates. But they also said it posed challenges for collaboration. As one student said, " . . . sometimes it's really good to have everyone do the same topic because then everyone can help anybody; and when you all have different topics, it's like she's doing that and he's doing that, so we can't talk about it, so it depends."

Competency-Based Progression

In this section, we discuss:

- teacher use of competency-based practices
- student experience with competency-based practices
- challenges of competency-based progression.

Competency-based progression *enables personalized paths to run their natural course by removing external constraints on what material each student works on, when, and for how long. Each student's progress toward clearly defined goals is continually assessed, and assessment occurs "on demand" when a student is ready to demonstrate competency. Assessment may take a variety of forms, such as projects or presentations, as well as more-traditional tests or quizzes. A student advances at his or her own pace and earns course credit (if applicable) as soon as he or she demonstrates an adequate level of competency.*

KEY TAKEAWAYS ON COMPETENCY-BASED PROGRESSION

NGLC schools appeared to differ from the national sample in some respects.

- Although most teachers in each group reported using competency-based practices, NGLC teachers and students reported higher levels of such practices.
- NGLC teachers were more likely than those in the national sample to require students to get through a certain amount of material.

There were several challenges related to competency-based progression.

- Many NGLC teachers said that allowing students to progress at their own pace through content was challenging when students did not complete work at an acceptable pace.
- Competency-based grading systems were difficult to explain to stakeholders and did not fit with traditional reporting practices.
- NGLC schools did not award credit for partial mastery of a course in a way that could be transferred to other schools.

Summary

A majority of teachers in both groups reported using competency-based practices to a moderate or large extent. Teachers and students in the NGLC schools reported that competency-based practices were common, allowing students to work at different paces and on different topics or skills at the same time. While this finding is encouraging, implementing competency-based progression is not without challenges. Some teachers reported that organizing students into groups for the larger performance tasks could be difficult, because students were in different places in learning the material. In addition, principals and teachers said that competency-based grading systems were difficult to explain to stakeholders and did not fit with traditional reporting practices.

Competency-Based Progression: How do NGLC and national practices compare?

Although most teachers in each group reported using competency-based practices, NGLC teachers and students reported higher levels of such practices. Competency-based practices include enabling students to work at various paces and on different topics than their classmates, giving them opportunities to review or practice new material until they really understand it, requiring them to demonstrate that they understand a topic before moving on to a new topic, and enabling them to track their own progress. A majority of teachers in both the national and NGLC samples reported using competency-based practices to a moderate or large extent. However, the NGLC teachers were more likely to report using these practices to a great extent (Figure 4). In addition, NGLC students were more likely than students nationally to report that they always experienced practices consistent with competency-based progression (Figure 5). Overall, though, neither group of students perceived these practices to be very common.

NGLC teachers were more likely than those in the national sample to require students to get through a certain amount of material. This result is shown in Figure 4. Since the amount of time students

VIGNETTE: What does competency-based progression look like?

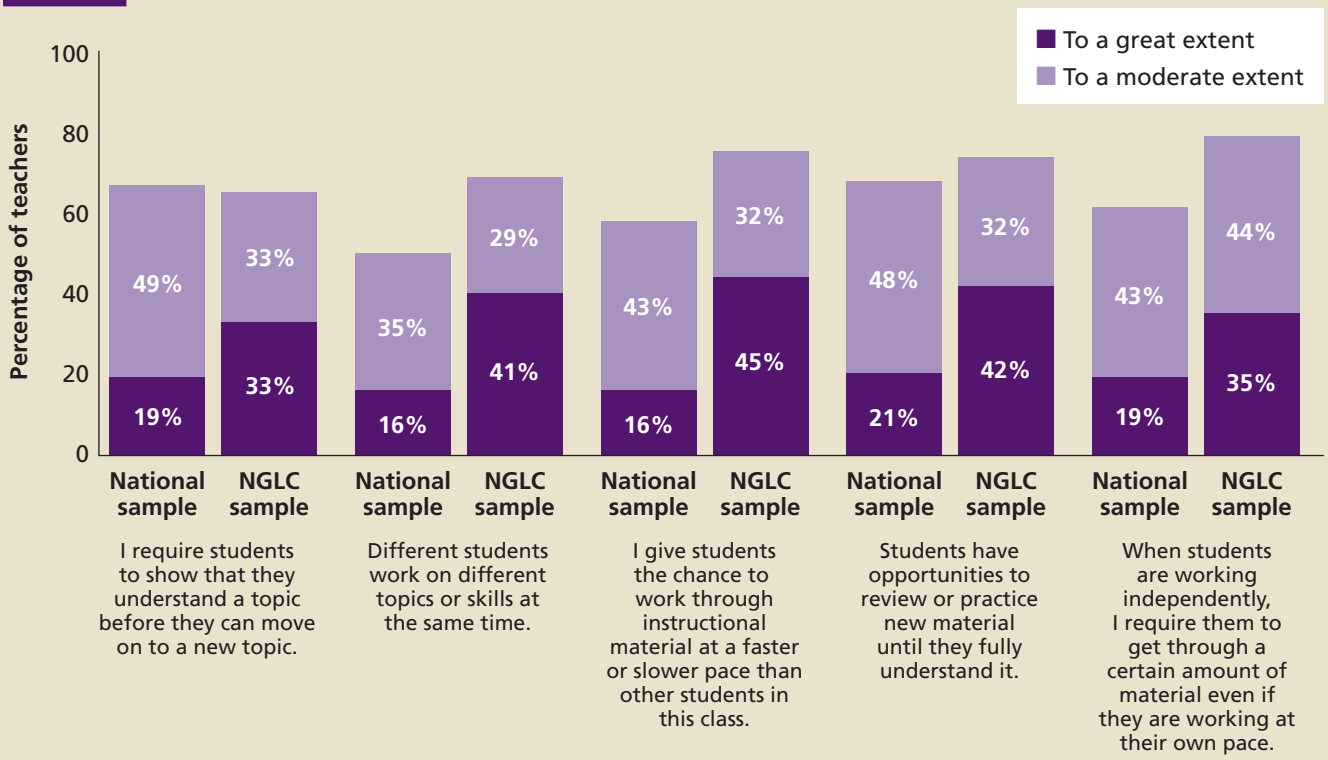
School C, a charter high school with 9th grade students, was broadly implementing competency-based progression during its first year of operation. Students were aware of the goals and standards they were supposed to learn at the start of the year, and they met with teachers during the year to plan how they were going to meet those goals. According to the principal, teachers shared the goals of each course with students at the outset and students "project managed" to get the work done and track their progress.

According to school staff, the curriculum was anchored in the Common Core State Standards and the Next Generation Science Standards. As students worked through the standards, they were periodically tested through brief assessments they called "comprehension checks," as well as longer, more-complex "performance task" assessments. When those were accomplished with a score of at least 75 percent, students moved on to the final assessment for that standard.

Comprehension checks and performance tasks could take several forms, but most often were a quiz or a small project. Students tackled the comprehension checks when they were ready ("…at different times…it's one of the beauties of the school," according to one student) and moved through the content at their own pace. Students were grouped in the same classes "as other people who are on the same pace with us," as one student put it. Students progressed to the next standard upon demonstrating mastery.

FIGURE 4

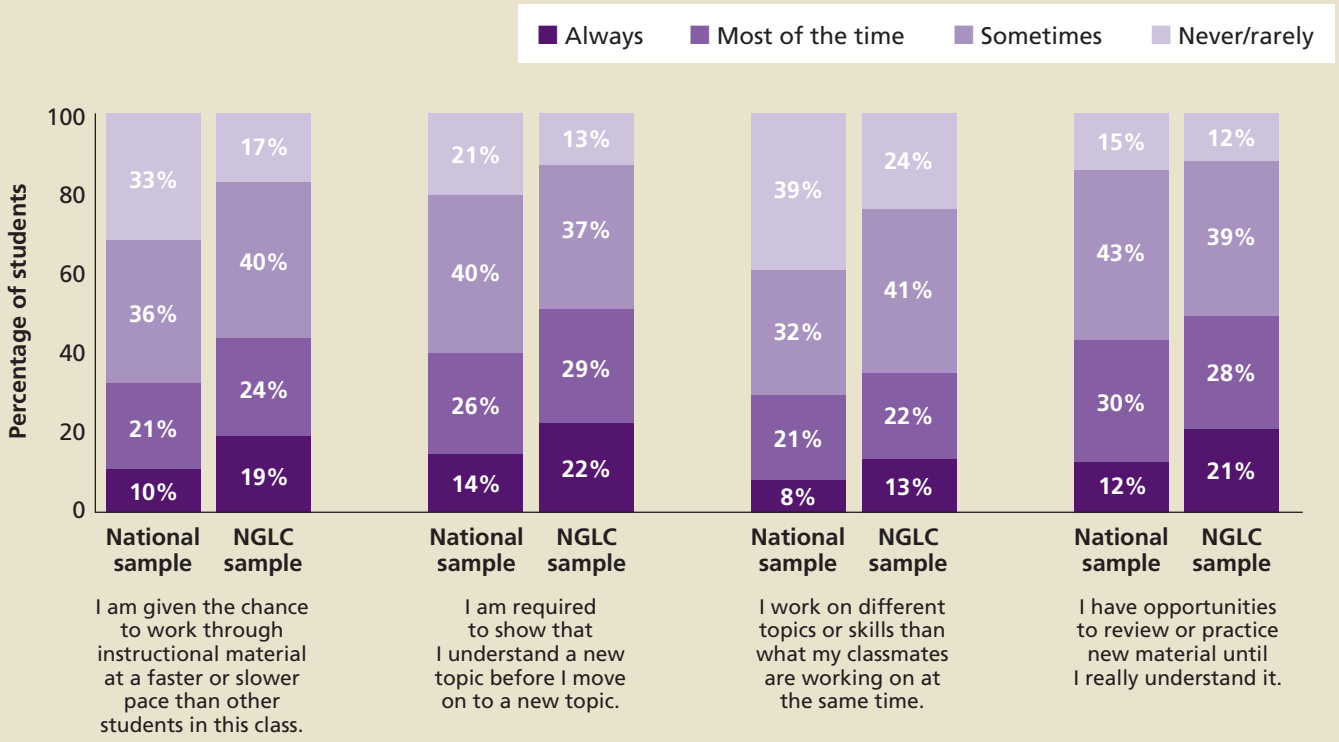
Teacher reports of competency-based learning practices



Notes: N = 210–212 NGLC teachers; N = 525 national teachers; survey question: "Please indicate the extent to which you agree with each of the following statements about your curriculum and instruction." Response choices were on a four-point scale from 1 ("not at all") to 4 ("to a great extent").

FIGURE 5

Student reports of competency-based learning experiences



Notes: N = 4,784–4,794 NGLC students; N = 864 national students; survey question: "The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all of your classes in math, English/reading, science, and social studies this year, and mark the response that indicates your typical experience." Response choices were on a five-point scale from 1 ("never") to 5 ("always").

worked at their own pace was not the same between the two groups, this finding is difficult to interpret. One possibility is that NGLC teachers wanted to ensure that students maintained focus and continued to work hard in a self-paced environment. According to most of the teachers we interviewed, part of the rationale was a belief that students may need to develop the skills necessary to work at their own pace. Many NGLC school leaders told us that they were taking a "gradual release" approach, in which students initially received lots of support and structure, which decreased over time to allow students to take greater responsibility for their own learning. Since most of the NGLC schools were in their first or second year of implementation, it is possible that we observed a greater degree of support and structure than will be present in future years. An alternative explanation for requiring students to cover certain material is that teachers wanted to ensure coverage of curriculum content that the students would not have covered at their own pace. Some school leaders and teachers said they felt it necessary to set a pace to help students fill in gaps in their learning *and* access grade-level content in preparation for state tests, echoing similar comments about offering students choice in their personal learning paths.

Challenges of Competency-Based Progression

Many NGLC teachers said that allowing students to progress at their own pace through content was challenging when students did not complete work at an acceptable pace. Teachers and administrators at many NGLC schools reported that allowing students to progress through content at their own pace was challenging for several reasons. According to teachers, many students did not know how to organize their time so they would complete their work at a sufficient pace. For example, at the end of the second semester some students still had not completed work they were expected to do during the first semester. Many schools addressed this challenge by using pacing guides, or by

specifying a minimum amount of work that students must complete in a certain time frame.

Competency-based grading systems were difficult to explain to stakeholders and did not fit with traditional reporting practices. Teachers and principals reported that competency-based grading systems were often challenging to explain to parents and community members. Principals also reported that competency-based grades had to be converted into traditional "grades" that were acceptable for state-level reporting and college applications, a challenge encountered by all the NGLC schools implementing this strategy.

NGLC schools did not award credit for partial mastery of a course in a way that could be transferred to other schools. One of the goals of competency-based systems is that students would be awarded credit when they had demonstrated mastery of the material. Ideally, credit would be awarded in increments smaller than a course, and the credit would be transferrable, eliminating the need for students to repeat lessons or courses if they transfer schools. Like many schools, NGLC schools awarded credit for mastery when students completed a course. Where NGLC schools awarded mastery of material at a finer granularity than a full course (such as individual learning standards), students were not able to take these credits with them when they transferred to another school, according to principals. As a result, students who transferred schools would likely have to cover the material again. This challenge may stem from a lack of widely accepted standards for how to track completion of material in increments of less than a full course.

Flexible Learning Environments

In this section, we discuss:

- flexible use of school resources such as staff, space, and time
- use of technology in instruction
- frequency of adjusting student groups based on data
- challenges of flexible learning environments.

Flexible learning environments *imply that the school adapts the use of resources such as staff, space, and time to best support personalization. For example, elements of the learning space—size, classroom organization, and furniture—are designed to support implementation of PL. The structure of learning time and student grouping strategies are flexible, responsive to student needs, and driven by data where appropriate. Technology is a key aspect of the school model and is available to all students; often schools provide a device to each student.*

KEY TAKEAWAYS ON FLEXIBLE LEARNING ENVIRONMENTS

NGLC schools appeared to differ from the national sample in some respects.

- NGLC teachers reported more-flexible use of resources such as space, staff, and instructional time to support PL.
- Although both groups reported that technology played a primary role in instruction, NGLC teachers reported greater reliance on technology-based instructional materials than teachers nationally.
- Obstacles to teaching with technology were reported to be less prevalent in NGLC schools.
- Although both groups considered student grouping based on data to be a key strategy, NGLC teachers reported adjusting those groups more frequently.

NGLC schools were similar to the national sample in emphasizing student grouping and use of technology (noted above).

- Teachers and students in both groups reported positive opinions about the school environment.

There were several challenges related to flexible learning environments.

- Creating flexible learning spaces in traditional school buildings was challenging.
- NGLC schools experienced barriers to implementing flexible learning environments at the school level, but practices were more flexible at the classroom level.

Summary

NGLC schools used space, staff, and time in ways that were different from schools in the national sample. These practices included creating learning spaces that were open and flexible, using a variety of activities that were based on the needs of the student or the demands of the lesson, using student achievement data to assign students to groups, and, among teachers who reported grouping students by ability level, changing those groups more frequently.

The role of technology in instruction was similar in both samples, as was the use of data to assign students to groups. While these findings are encouraging, some NGLC principals, teachers, and students reported that creating and using flexible spaces in traditional school buildings was challenging: Such spaces were often noisy, making it difficult for students to concentrate. Some aspects of flexible scheduling also proved challenging for NGLC schools: Schools experienced barriers to flexible scheduling at the school level but used time flexibly at the classroom level, and student grouping was more flexible within classes than schoolwide. Teachers and students in both groups reported similarly positive perceptions of the school environment, which could enable flexible use of resources in ways that support PL.

VIGNETTE: What do flexible learning environments look like?

School D, a charter school serving grades 6–8, had been implementing its PL model for one year at the time of our visit. The school was in a converted office building, with classrooms made out of modular walls that did not reach the ceiling and could be rearranged, and several large open spaces where students could work independently or in groups, and where the school gathered for "design challenges" (complex, interdisciplinary, long-term projects), and other whole-school events.

A 5-week "trimester" allowed students who were struggling to solidify their skills, and permitted students who were on track to take interdisciplinary classes that went beyond the regular curriculum (such as public speaking, coding, and "myth busters"). By design, many teachers were cross-certified, enabling administrators to be flexible in how they organized classes, and to play to teachers' strengths. Many projects were team-taught.

Classes were organized in a block schedule, although class length fluctuated as the school experimented with the schedule. The schedule could be rearranged easily, even on short notice, to accommodate projects and whole-school design challenges. In classrooms, the structure of learning time was flexible, teachers had discretion to use the time as they saw fit, and students experienced a variety of instructional approaches and activities depending on the lesson.

Students were grouped by learning level schoolwide. Administrators considered standardized test data and consulted with parents and students to make student grouping decisions. In classrooms, grouping was more fluid and often dependent on the lesson requirements. The school had a one-to-one technology model. Students used their own laptops or were supplied with Chromebooks, which they could take home. Students reported using technology "constantly" to monitor their progress, take tests, work on projects, communicate with teachers, do research, and complete assignments. In addition, students could attend classes virtually, using Adobe Connect, during unplanned school closures, such as snow days.

"Who teaches what depends on certification area or talents working with low-achieving or high-achieving kids . . ."
—PRINCIPAL COMMENT ABOUT STAFFING

Flexible learning environments: How do NGLC and national practices compare?

NGLC teachers reported more-flexible use of resources such as space, staff, and instructional time to support PL. NGLC teachers were somewhat more likely to report that their schools had large open spaces and comfortable furniture that could easily be rearranged to facilitate PL. In addition, NGLC teachers were more likely to report that "co-teaching or job-share" best described their teaching arrangement, although such arrangements were not common overall.[1] NGLC teachers were less likely to report that scheduling constraints were an obstacle to implementing PL in their schools; and many NGLC administrators we interviewed reported that their schools had flexible schedules that were intended to facilitate PL. Many NGLC administrators also mentioned that teachers generally had flexibility to use their classroom time in the way that was best suited to the lesson and the needs of the students. Teacher log data confirm that NGLC teachers adjusted their instructional activities to suit the needs of the lesson or the student, though comparative log data are not available for a national sample.

NGLC teachers reported greater reliance on technology-based instructional materials than teachers nationally. Teachers in both groups were equally likely to report that technology played a primary role in instruction, but there is some evidence to suggest that NGLC teachers relied on technology-based instructional materials for some activities to a greater extent than teachers in the national sample. For example, NGLC teachers reported that students were engaged in independent practice with software for a larger

[1] *Co-teaching* or *job-share* was defined as, "I am one of two or more teachers who are jointly responsible for teaching the same subject(s) to a group of students (for example, in the same classroom), all or most of the day and/or in a majority of classes."

proportion of the lesson. In addition, when students were using technology, NGLC teachers reported that students more often used structured online curriculum materials; watched videos, animations, or simulations; solved multi-step, open-ended problems; and received immediate feedback on problem solutions, as shown in Figure 6.

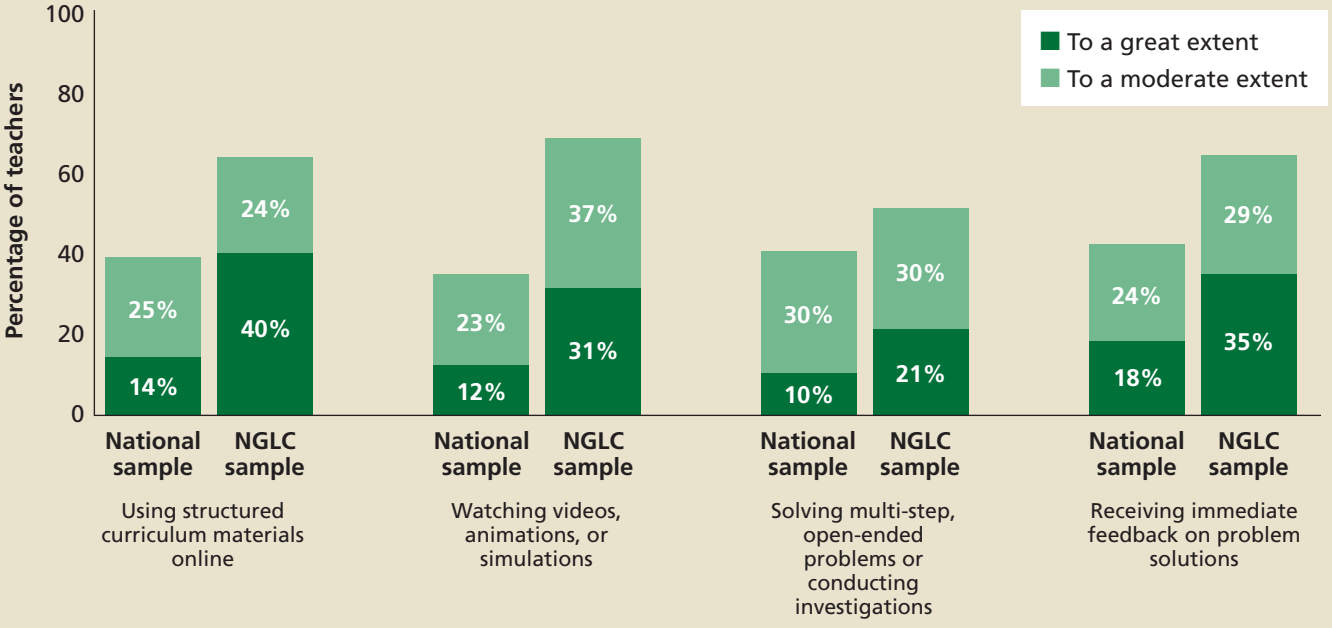
Obstacles to teaching with technology were reported to be less prevalent in NGLC schools. NGLC teachers were less likely to perceive the following operational and logistical obstacles to promoting student learning using technology:

- an inadequate number of devices (e.g., laptops)
- problems with hardware
- inadequate bandwidth
- lack of opportunities to participate in professional development
- lack of flexibility in deciding how to use technology in instruction
- lack of support from technology specialists
- inadequate opportunities for teachers to provide input on how technology is used.

Although both groups considered student grouping based on data to be a key strategy, NGLC teachers reported adjusting those groups more frequently. Similar proportions of teachers in the NGLC and national samples reported using student achievement and nonachievement data to assign students to groups within their classes. Similar proportions also reported grouping students of similar ability levels together. However, there is some evidence that NGLC teachers adjusted student groupings more frequently. Among teachers who reported grouping students of similar ability levels together, NGLC teachers changed groups more frequently: 29 percent reported changing groups weekly, compared with 4 percent of teachers in the national sample.

Teachers and students in both the national and NGLC schools reported positive opinions about the school environment. Large majorities of teachers in both samples agreed that administrators and teachers were focused on improving student learning, were supportive, and that teachers collaborated well with one another. NGLC teachers were more likely to report high levels of administrator support and trust, and that teachers were highly focused on improving student learning, as shown in Figure 7. Teachers' perceptions of students were similar and largely positive in both samples, with majorities agreeing that students were respectful of other students and staff and were

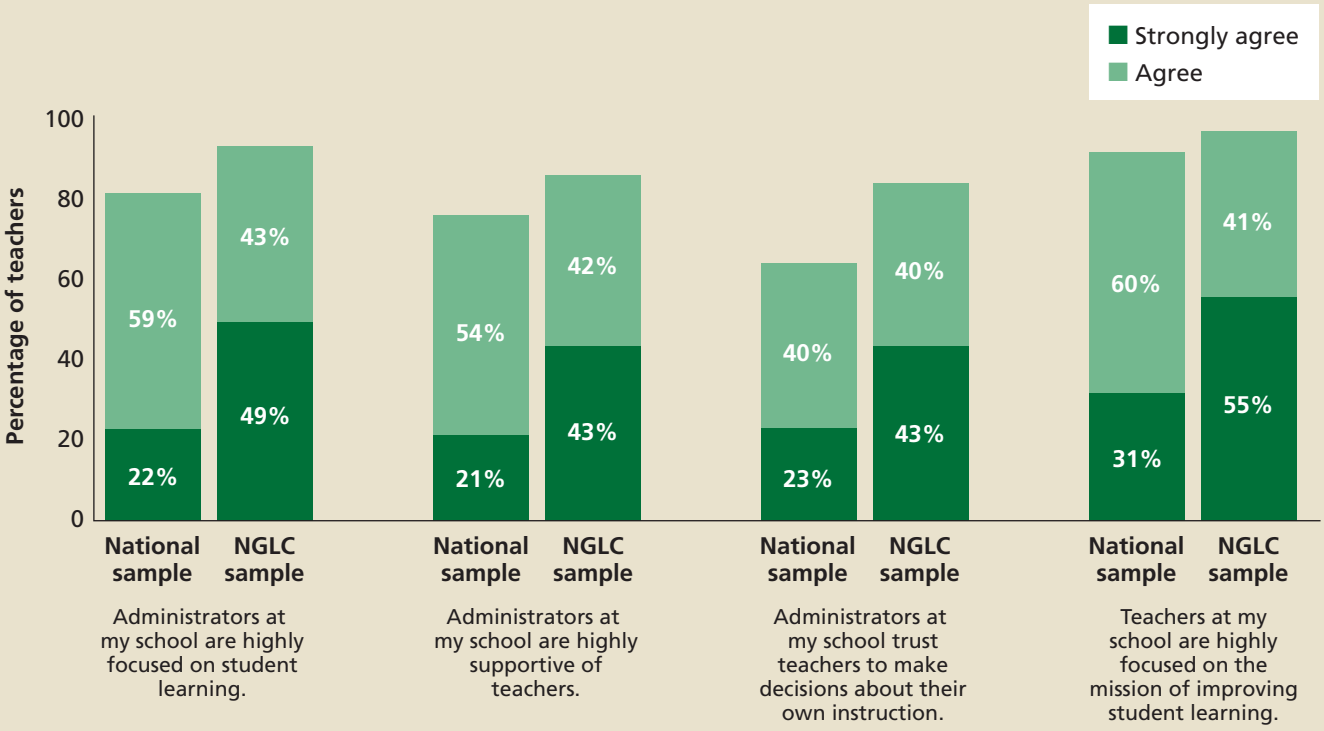
FIGURE 6 Teacher reports of the extent to which students engaged in certain activities while using technology



Notes: N = 213–216 NGLC teachers; N = 525 national teachers; survey question: "For this question, we are interested in the activities students are engaged in when they are using technology. Please indicate the extent to which students are engaged in the following types of activities." Response choices were on a four-point scale from 1 ("not at all") to 4 ("to a great extent").

FIGURE 7

Teacher agreement with items related to school environment, NGLC and nationally



Notes: N = 228–230 NGLC teachers; N = 525 national teachers; survey question: "Rate your level of agreement with each of the following statements about your school." Response choices were on a four-point scale from 1 ("disagree strongly") to 4 ("agree strongly").

motivated to achieve. However, in both groups, about half of teachers did report that certain factors were obstacles to implementing PL, such as too many students in classes, too much diversity in achievement levels, high levels of absenteeism, disciplinary problems, motivation, behavior, or attendance.

Students had similarly positive perceptions of their school environment. Large majorities of students in the national and NGLC samples reported positive feelings about their schools and learning environments, and the two groups were equally likely to agree that they felt supported by their teachers in their school work and in preparing for the future, as shown in Figure 8. Although a large majority of NGLC students expressed positive opinions, they were somewhat less likely to report that they felt safe, comfortable being themselves, and an important part of the school community.

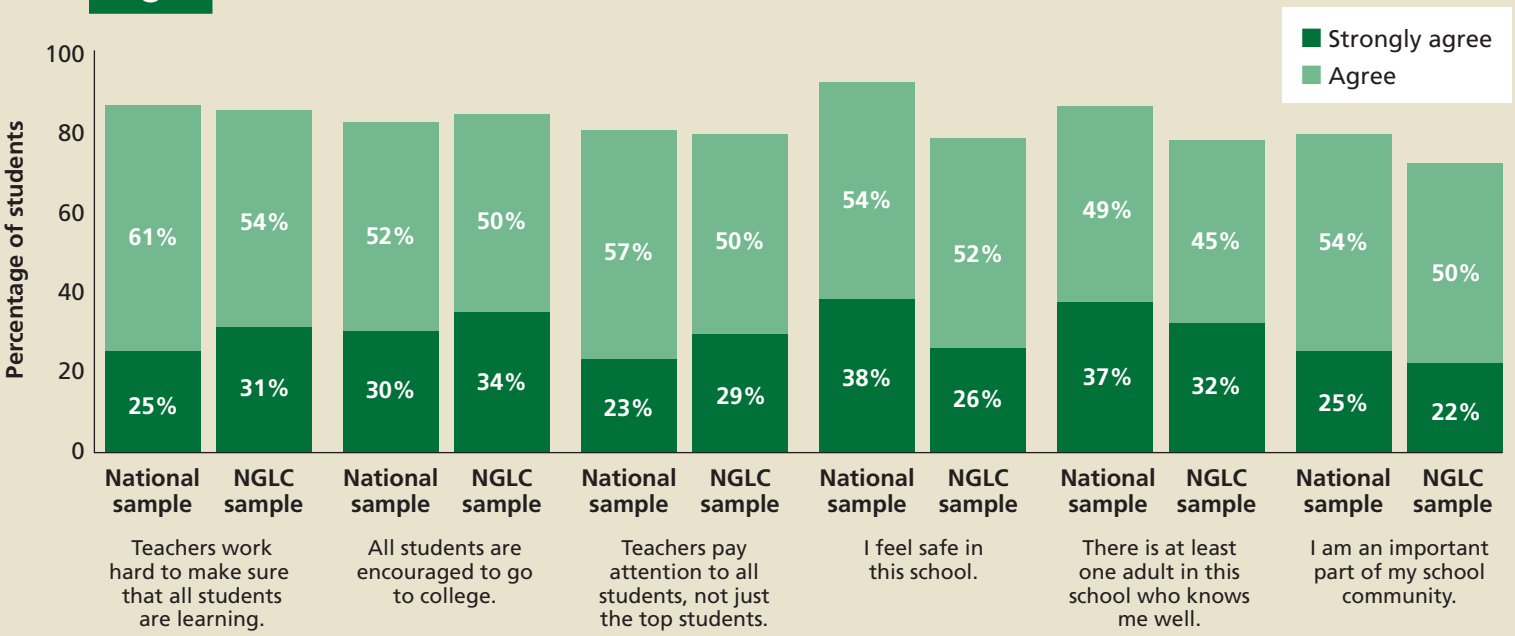
Challenges of Flexible Learning Environments in NGLC Schools

Creating flexible learning spaces in traditional school buildings was challenging. Most of the NGLC schools were located in traditional school buildings that often could not be reconfigured to use space in flexible ways. Nonetheless, a majority of administrators reported that their schools contained some flexible learning spaces. Where nontraditional spaces existed, using them was not without challenges. In particular, staff and students in such schools reported that open spaces were noisy, making it difficult to focus on instruction.

NGLC schools experienced barriers to implementing flexible learning environments at the school level, but practices were more flexible at the classroom level. Principals reported that flexible grouping was rarely used at the school level, and in most schools students were grouped by traditional grade level. Teacher survey and interview data indicate that student grouping was more flexible within classes than schoolwide. Teachers reported that data-based student grouping strategies were used frequently at the classroom level, where students were sometimes grouped homogeneously and sometimes heterogeneously, according to the goals of the lesson.

FIGURE 8

Students' opinions about their school environment



Notes: N = 4,629–4,665 NGLC students; N = 864 national students; survey question: "How much do you agree with the following statements about your school?" Response choices were on a four-point scale from 1 ("strongly disagree") to 4 ("strongly agree").

However, implementing flexible learning time at the school level could also take the form of flexible scheduling, where each student would have a unique schedule that changed as often as weekly based on data about their learning needs. Only a few NGLC schools took this approach, and in those that did, the principals reported that it was difficult for logistical reasons. Staff at these schools reported that they struggled to create classes of a reasonable size, or ensure that there were enough teachers available to supervise students working independently. Some school leaders found that the work of creating a new schedule each week was too burdensome. Teachers, however, reported that they were empowered to use their classroom time flexibly, which in most schools meant using a variety of instructional strategies in accord with the needs of the lesson or the student.

> *"We started off with teachers creating a different schedule for the students every week. It just took too much time because they were doing it all by hand . . ."*
>
> —PRINCIPAL COMMENT ABOUT FLEXIBLE SCHOOLWIDE SCHEDULING

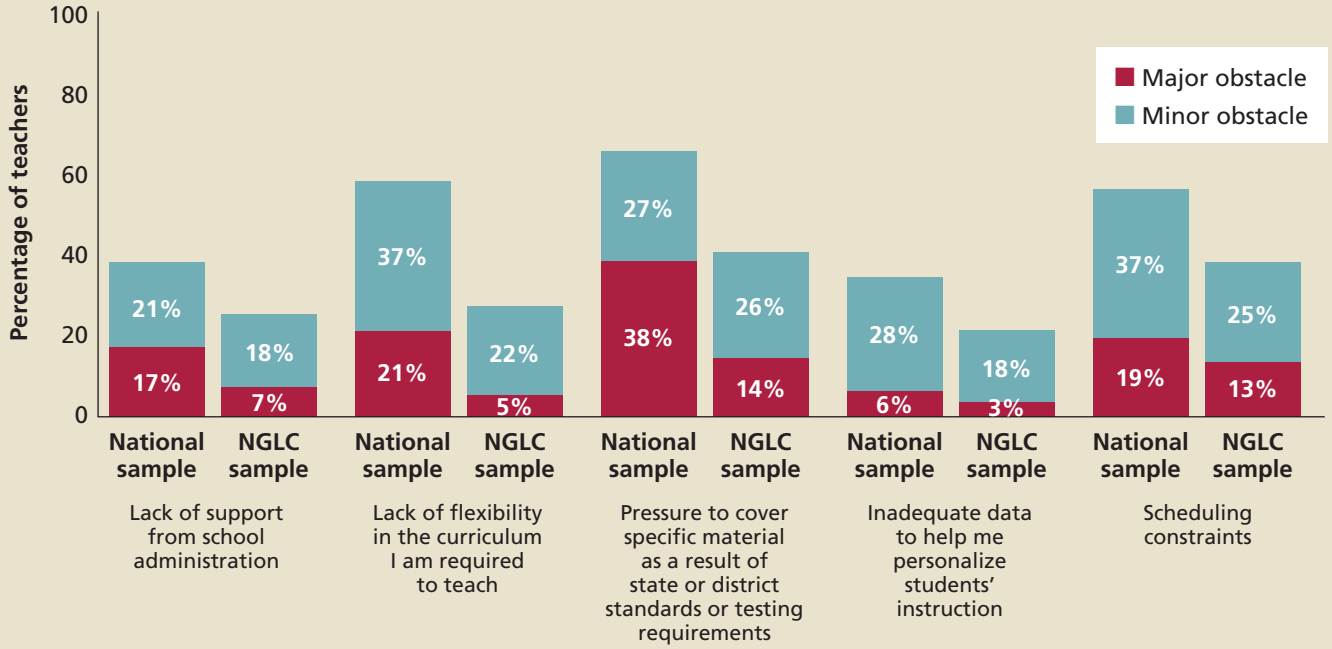
General Challenges to Implementing Personalized Learning

The teacher survey included several questions that addressed challenges to implementing PL. These questions asked about implementing PL in general and were not specific to any one of the four strategies; we therefore discuss them briefly in this section.

NGLC teachers were less likely than teachers nationally to report operational obstacles to implementing PL, such as scheduling constraints. We examined teacher perceptions of obstacles to implementing PL in the NGLC and national samples. Although some of these conditions were not perceived as obstacles by a majority of teachers, teachers in the NGLC sample were less likely to report that environmental and operational factors, such as lack of administrator support, pressure to cover specific material, lack of data, lack of flexibility in curriculum, and scheduling constraints, were obstacles. Pressure to cover specific material and lack of flexibility in the curriculum seemed to be the largest obstacles for teachers in the national sample, as shown in Figure 9.

FIGURE 9

Teacher reports of obstacles to implementing PL



Notes: N = 217–219 NGLC teachers; N = 525 national teachers; survey question: "Please indicate the extent to which each of the following conditions is an obstacle to your efforts to promote personalized learning for students. If the condition does not exist in your school, please mark 'not applicable.'" Response choices were on a four-point scale from 1 ("not applicable; condition does not exist in my school") to 4 ("condition exists and is a major obstacle").

How Did Charter and District NGLC Schools Compare in Their Implementation of Personalized Learning?

This section compares charter and district implementation of

- learner profiles
- personal learning paths
- competency-based progression
- flexible learning environments.

With early signs that PL holds promise for positive effects on student achievement, there has been considerable enthusiasm about scaling up its implementation. Although many of the earliest adopters of PL have been charter schools, successful scale-up of this approach will inevitably include district-operated public schools, which serve the vast majority of K–12 students in the United States. A manifest question is whether the positive results seen thus far in samples that are dominated by charter schools are likely to generalize broadly. For example, charter schools comprised 92 percent of the sample that produced favorable results in the Pane et al. (2015) study. Are there attributes of charter schools that are particularly conducive to implementing a somewhat radical innovation like PL, or should we expect scale-up in districts to proceed with similar results to those seen in these charters? We use the limited data available in the current study to conduct a preliminary exploration of this topic. Our small sample consists of one-fourth district-operated and three-fourths charter schools. Here, we examine implementation similarities and differences we observed between district and charter schools in our sample, and we examine achievement outcomes along the same dimension in the next chapter. *Although these analyses do not enable strong conclusions due to small sample sizes, and thus should be interpreted with great caution,* they may offer some observations that warrant consideration by stakeholders interested in the scale-up of PL.

KEY TAKEAWAYS FROM THE COMPARISON OF CHARTER AND DISTRICT SCHOOLS

Implementation of PL in Charter and District Schools Within the NGLC Sample

Charters appeared to have higher levels of implementation than district schools in some ways.

- Teachers reported more-frequent receipt and more-extensive use of actionable student data.
- More teachers reported adapting course content to meet students' needs to a great extent.
- Teachers reported using small-group instruction for larger portions of the lesson.
- Teachers were more likely to agree that their curriculum materials were of high quality.
- Teachers and students reported more-extensive use of competency-based practices.
- Flexible use of space and staff was more prevalent.
- Teachers reported incorporating more technology into instruction and fewer obstacles to doing so.
- Teachers reported greater use of data to group students.
- Teachers and students reported more-positive perceptions on some dimensions of school environment.
- Teachers were less likely to report that student factors, such as discipline, were major obstacles to PL.

District schools appeared to have higher levels of implementation than charters in some ways.

- District schools appeared to have more-comprehensive learner profiles.
- Teachers reported changing student groupings more frequently.

Charter and district schools appeared to be similar in several ways.

- Teachers reported similar levels of use of learner profiles.
- Opportunities for student choice were uncommon.
- Teachers were equally likely to assign students to classes and groups by age or achievement.
- Teachers reported using classroom time flexibly and incorporating a variety of activities.
- Teachers and students reported positive perceptions of the school environment.

Summary

In general, charter schools tended to display more-extensive implementation of many aspects of PL. District schools displayed less-extensive implementation and tended to look more similar to the national sample, suggesting lower implementation of novel PL practices. Charter teachers reported greater use of key aspects of learner profiles, such as more-frequent receipt and use of student data, and greater adaptation of course content to meet students' needs. Charter teachers and students reported a greater extent of using and experiencing competency-based practices, such as being able to work on different topics than others and at their own pace. Key components of flexible learning environments, such as flexible use of staff and space, and use of technology, were reportedly more common in charter schools.

Reported use of learner profile documents was the same in both groups, although charter teachers' responses suggest that their learner profiles were less comprehensive. As with the national and district samples, opportunities for student choice were uncommon, although charter teachers reported using small-group instruction more frequently. Flexible use of class time was common in both groups, as was using a variety of instructional strategies. Teachers and students in both groups reported similarly positive perceptions of the school environment.

Due to numerous limitations, these findings should be interpreted with great caution.

Learner Profiles

Charter teachers reported more-frequent receipt and more-extensive use of actionable student data. In both groups, majorities of teachers reported receiving a variety of achievement and nonachievement data at least a few times per month. However, charter teachers reported receiving such data more frequently: approximately weekly. Teachers in both groups also reported using student achievement data for activities related to personalization, but charter teachers reported more-extensive use of such data for many instructional activities. Charter teachers were also slightly more likely to agree that their school data systems provided them with actionable data. Majorities of teachers in both district and charter schools agreed that data systems were easy to use and provided them with real-time, actionable data, but charter teachers expressed stronger agreement, as shown in Figure 10.

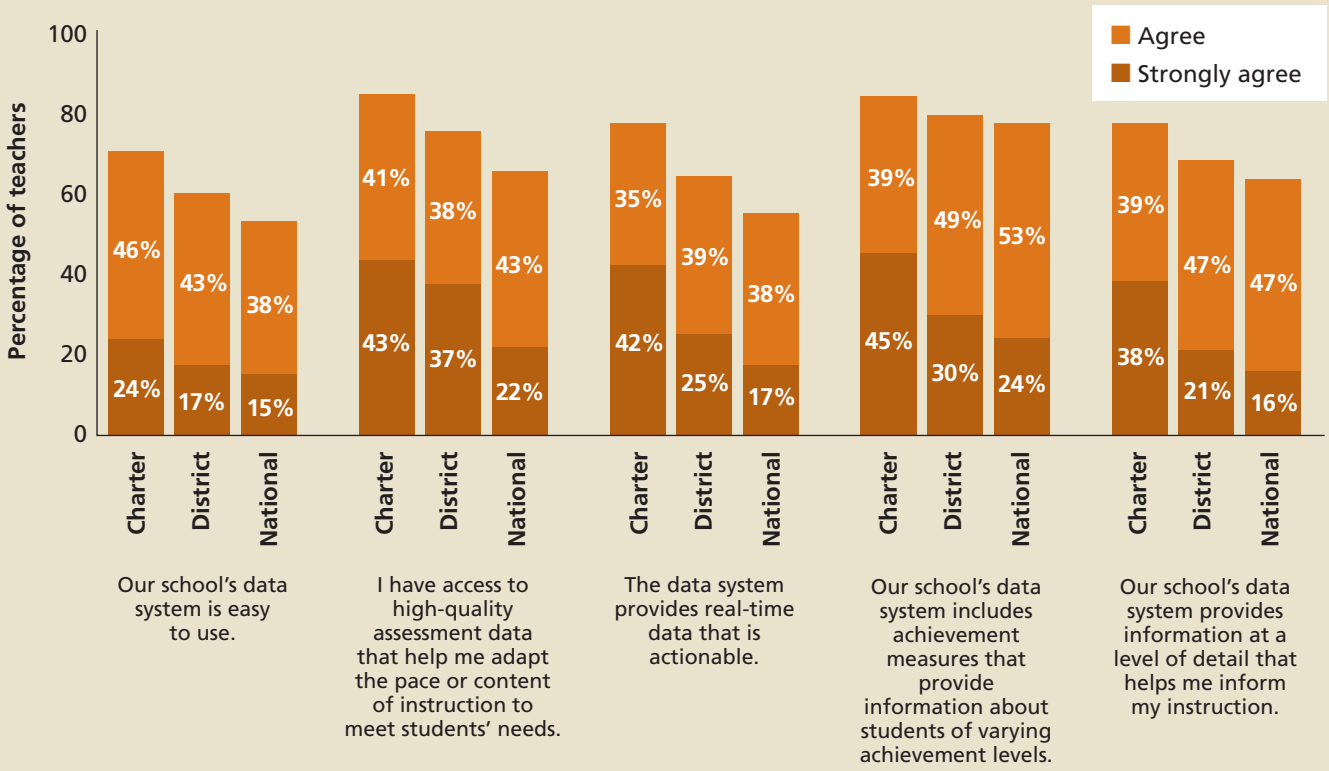
Teachers in both groups reported similar levels of use of learner profiles, but district schools appeared to have more-comprehensive learner profiles. About half of charter and district teachers reported that their schools used learner profile documents. Among teachers who reported that their schools used learner profiles, district and charter teachers were equally likely to report that the profiles were frequently updated and set forth a plan for students to accomplish their learning goals. Charter teachers were less likely to report that the profiles were comprehensive and available for every student.

Personal Learning Paths

More charter teachers reported adapting course content to meet students' needs to a great extent. Surveyed teachers in both groups reported a limited amount of student choice in the instructional materials and topics students used and focused on, with about one-third of teachers responding that they provided such choice. As described above, NGLC teachers tended to make efforts to adapt course content to meet students' needs by providing additional assignments, resources, and activities for remediation or enrichment. Charter teachers reported using these approaches to a greater extent than district teachers.

Opportunities for student choice were uncommon in both groups, but charter teachers reported using small-group instruction for larger portions of the lesson. Students in both groups reported that they were not given a great deal of choice in the topics or materials they used in their classes. Less than half of students reported that it was very or mostly true that their teachers took their experiences and interests into account when deciding what they would work on. Similarly, about one-third of teachers in both groups reported that students had opportunities to choose the instructional materials they used in class. However, there is some evidence that charter teachers attempted to address individual students' needs. Charter teachers were less likely to report using large-group instruction for

FIGURE 10 **Charter and district teachers' opinions of their schools' data systems, with national sample results included as a reference**



Notes: N = 151–153 charter teachers; N = 61–63 district teachers; N = 525 national teachers; survey question: "Please indicate your level of agreement with each of the following statements." Response choices were on a five-point scale of 1 ("not applicable") to 5 ("strongly agree"). For each item and group, 3 to 8 percent of teachers indicated the item was not applicable.

large portions (i.e., more than 40 percent) of the lesson, and more likely to report using small-group instruction.

Charter teachers were more likely to agree that their curriculum materials were of high quality. Access to high-quality curriculum materials is a key enabler of personal learning paths. Majorities of charter and district teachers agreed that their materials were of high quality and met the needs of all of their students, though charter teachers expressed slightly stronger agreement. Charter teachers also reported that a greater proportion of their curriculum materials were provided to them.

Competency-Based Progression

Charter teachers and students reported greater use of competency-based practices. Although a majority of teachers in both groups reported using competency-based practices to some extent, larger proportions of charter teachers reported using these practices to a great extent, as shown in Figure 11. In most cases, fewer district teachers reported using these practices to a great extent than did charter teachers, although this percentage was still higher than in the national sample (Figure 11). Daily instructional logs seem to confirm this pattern: teachers reported using competency-based learning practices for more than a small portion of the lesson but less than half of the lesson. Charter teachers reported using competency-based practices for a larger portion of the lesson overall, and that more of the lesson involved content students could experience at different levels of depth. Charter students also reported experiencing some competency-based practices to a greater extent than district students and students nationally. Charter students were slightly more likely to report that they were required to demonstrate understanding of a topic before they could move on to the next one, and that they had opportunities to practice or review until they fully understood the material. Charter and district students were equally likely to report that they could work on different topics or skills than their classmates at the same time and work at a different pace than other students in the class. District students reported experiencing most of these practices to a similar extent as students nationally, as seen in Figure 12.

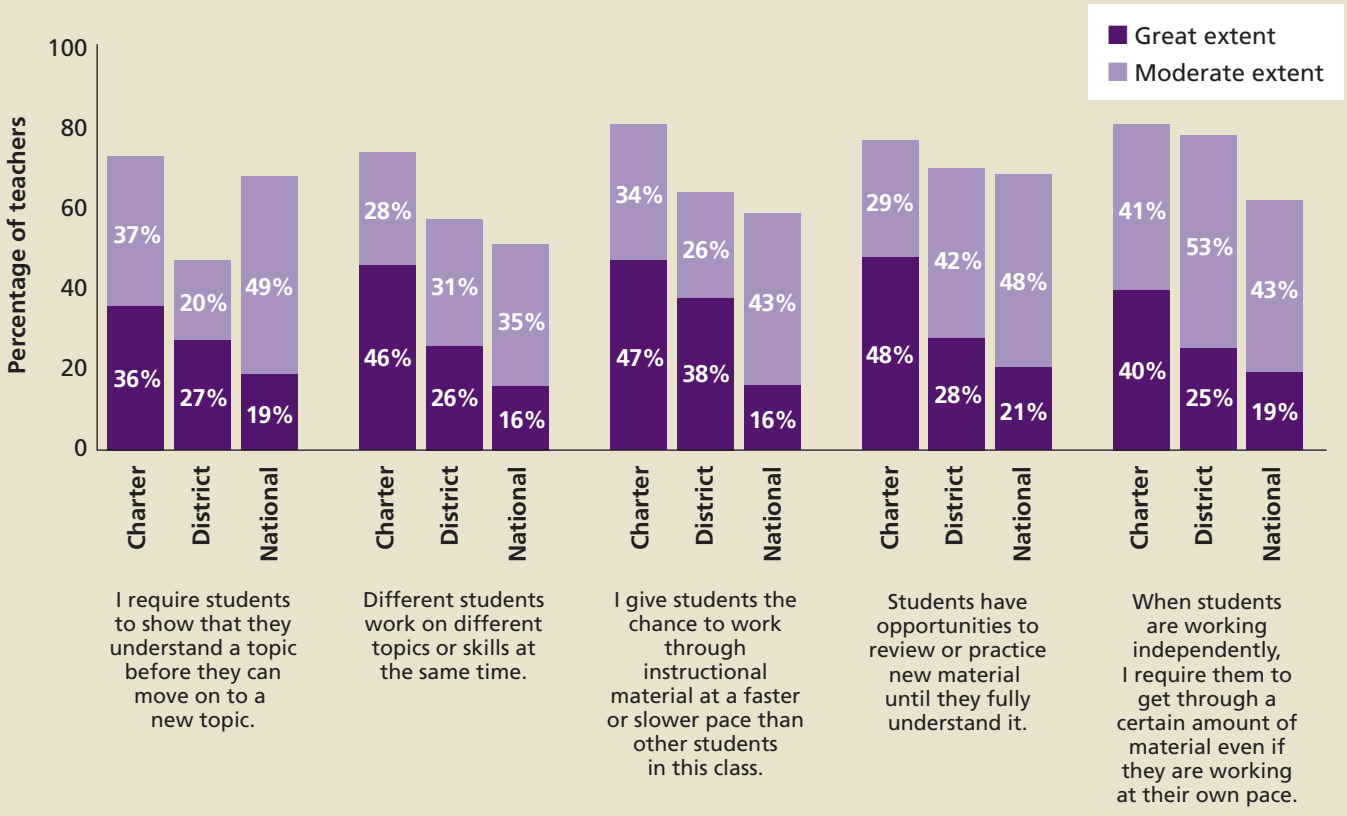
Charter and district teachers were equally likely to assign students to classes and groups by age or grade level. Teacher logs of daily practices show that students were assigned to their class by age or grade level in about two-thirds of lessons; assignment to classes or groups by achievement level was relatively uncommon. When students were grouped by achievement level, teachers reported using homogeneous groups in more lessons than heterogeneous groups. This is consistent with what many administrators told us in interviews.

Flexible Learning Environments

Flexible use of space and staff was more prevalent in charter schools than in district schools. Charter teachers were somewhat more likely to report that their school had nontraditional instructional spaces (e.g., space with comfortable furniture, large open instructional spaces, open common areas for student use, and breakout rooms) and that those spaces facilitated PL practices. Charter teachers were also more likely to report that "co-teaching or job-share" or "working under the supervision of another teacher" best described their teaching arrangement, although such teaching arrangements were not common overall. Principals of several NGLC charter schools reported that noncertified staff acted as advisers and mentors to students; we did not hear about similar roles in most district schools. Together, these findings suggest that charters were employing staff in unconventional roles to a greater extent than district schools. Finally, charter teachers were somewhat less likely to report that scheduling constraints were an obstacle to implementing PL practices.

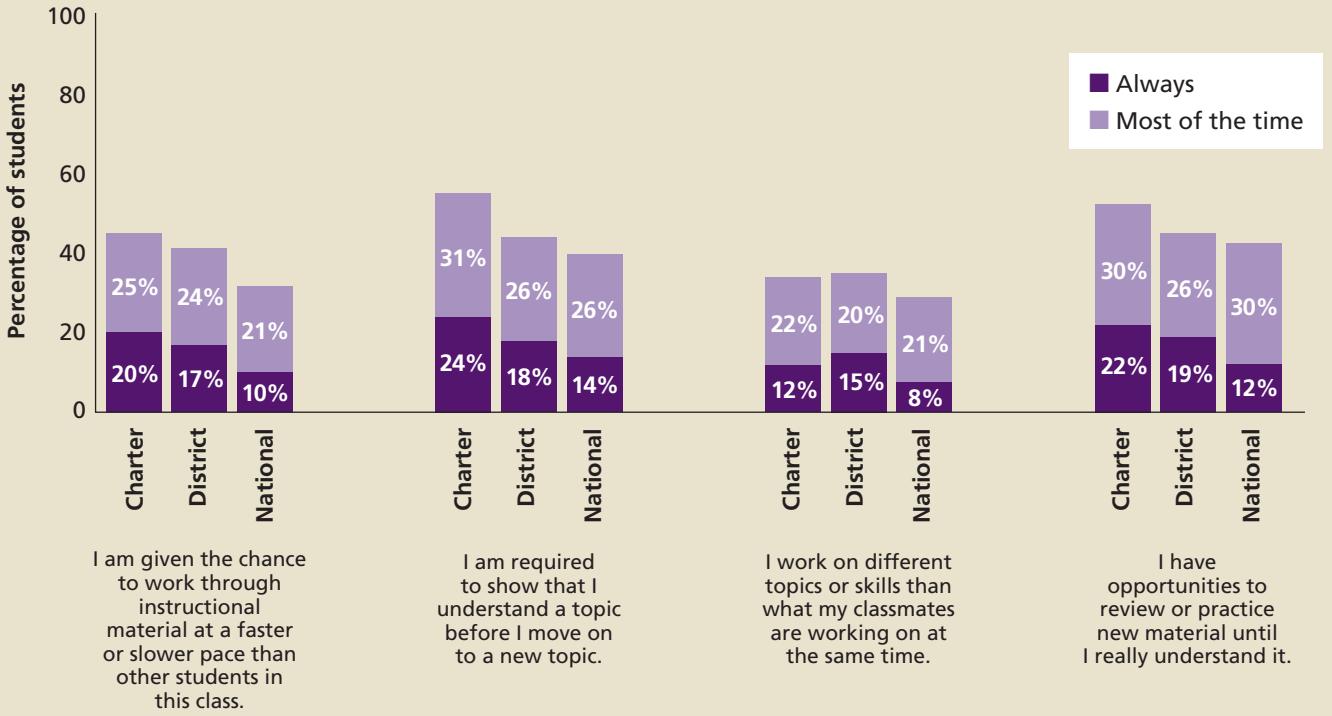
Charter teachers reported greater use of data to group students, but district teachers reported changing groupings more frequently. A majority of teachers in both groups reported using achievement data to assign students to groups within their classes, but charter teachers reported doing so to a greater extent. Charter teachers also reported grouping students of similar ability levels together somewhat more frequently, a finding that is consistent with charter teachers' greater reported use of competency-based instructional practices. However, among the teachers who reported grouping

FIGURE 11 Charter and district teachers' implementation of competency-based learning practices, with national sample results included as a reference



Notes: N = 149–150 charter teachers; N = 59–61 district teachers; N = 525 national sample; survey question: "Please indicate the extent to which you agree with each of the following statements about your curriculum and instruction." Response choices were on a four-point scale from 1 ("not at all") to 4 ("to a great extent").

FIGURE 12 **Charter and district students' experiences of competency-based learning practices, with national sample results included as a reference**



Notes: N = 2,964–2,978 charter students; N = 1,727–1,732 district students; N = 864 national students; survey question: "The following questions ask about your classroom experiences. When you answer them, please think about your experiences with all of your classes in math, English/reading, science, and social studies this year, and mark the response that indicates your typical experience." Response choices were on a five-point scale from 1 ("never") to 5 ("always").

students by ability level, district teachers reported changing groupings more frequently.

Teachers in both groups reported using classroom time flexibly and incorporating a variety of activities. Charter and district teachers reported similar use of instructional time and instructional activities. Teacher-led small-group instruction and small-group collaboration were used for more of the lesson than large-group instruction. Teachers' reports of their daily practice confirm that a variety of activities were used across lessons, and for a small portion of the lesson, suggesting that teachers changed activities frequently based on the needs of the student or the requirements of the lesson.

Activities used by teachers in both groups

- ✔ *students working independently with and without technology*
- ✔ *students working with teacher support with and without technology*
- ✔ *students working in small groups with and without technology.*

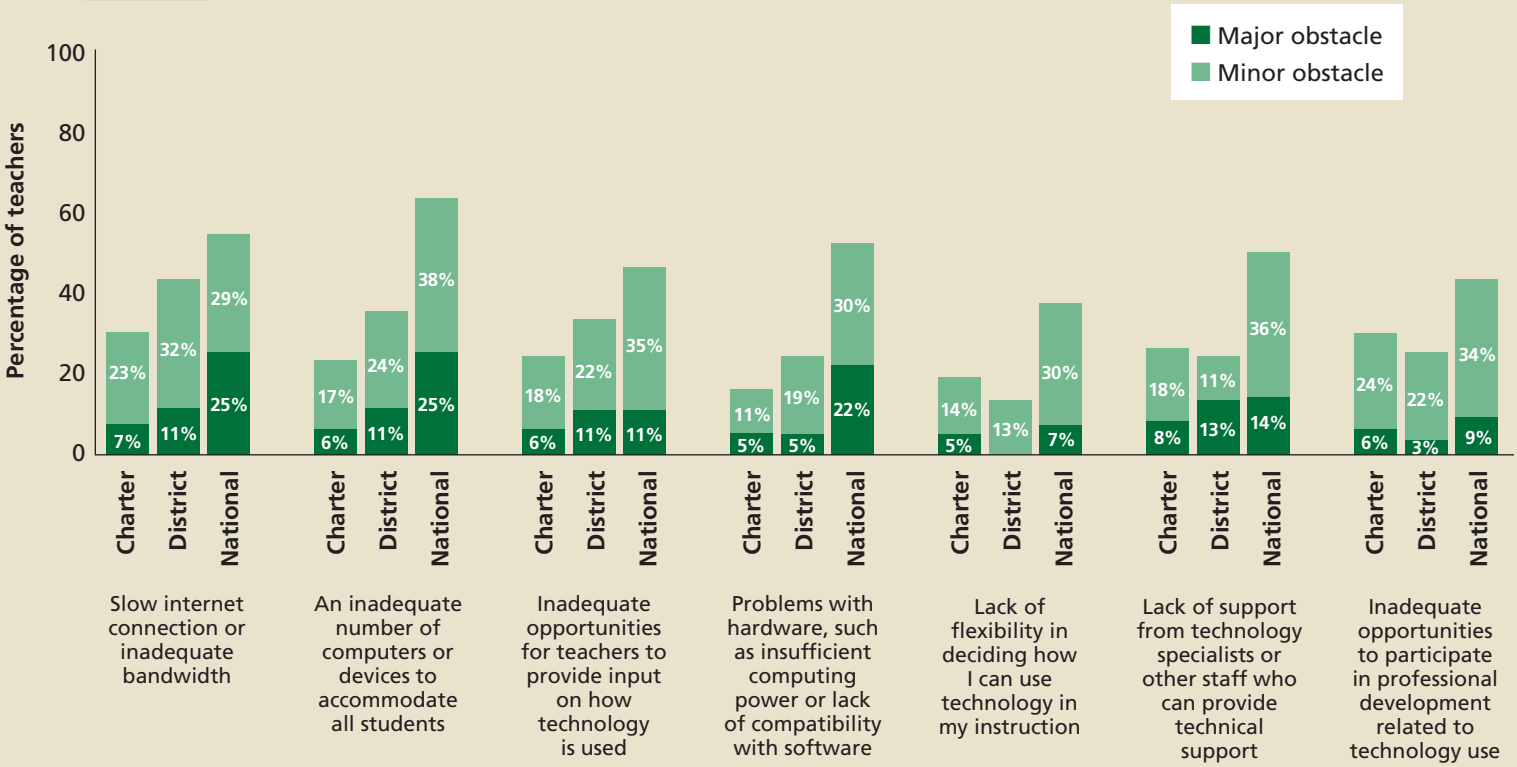
Charter teachers reported incorporating more technology into instruction and fewer obstacles to doing so. Charter teachers were more likely to report that technology played a primary role in instruction and that students used structured curriculum materials online to a moderate or great extent. The proportion of time teachers reported using technology in the classroom was similar in the two groups. Regarding technology obstacles, the two groups reported similar levels of logistical issues (e.g., lack of opportunity to participate in professional development), but district teachers reported more hardware and infrastructure issues, as shown in Figure 13.

Teachers and students in both groups reported positive perceptions of the school environment, with charter teachers reporting more-positive perceptions on some dimensions. Large majorities of NGLC teachers in both groups expressed positive opinions about the school environment. However, charter teachers were more likely to strongly agree that their school had high levels of administrator support and trust, and that teachers were highly focused on improving student learning. Charter teachers also reported more-positive perceptions of family involvement. Students in NGLC charter and district schools had similarly positive

perceptions of their school environment. Large majorities of students reported positive feelings about their schools and learning environments. The two groups were equally likely to agree that they felt supported by their teachers in their school work and in preparing for the future, and that they felt safe, supported by teachers, comfortable being themselves, and an important part of the school community.

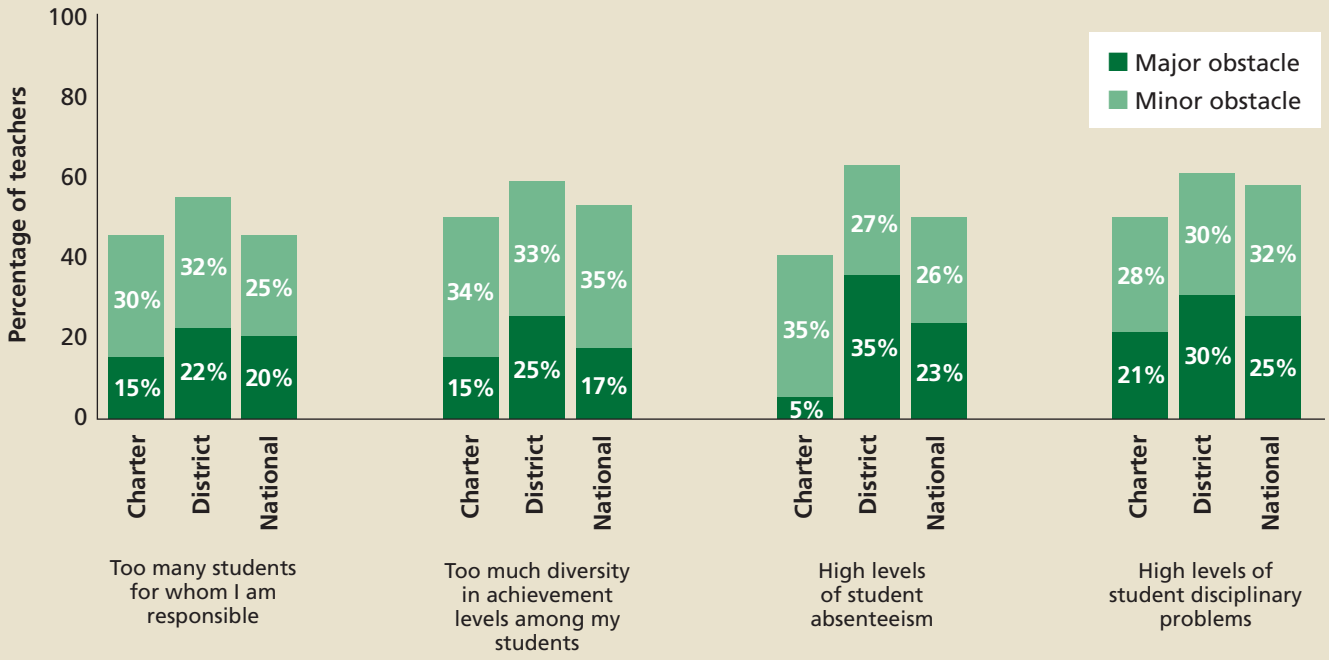
Charter teachers were less likely to report that student factors, such as discipline, were major obstacles to PL. Although majorities of charter and district teachers reported positive perceptions of students, charter teachers were more likely to agree that students in their school were motivated to achieve, and less likely to report that student characteristics and behavior were obstacles to implementing PL, as shown in Figure 14. In most cases, charter teachers perceived that these student-related factors were obstacles at similar rates as teachers nationally. Student absenteeism appears to be a notable exception; charter teachers were much less likely to report that this was a major obstacle than district teachers and teachers nationally.

FIGURE 13 Charter and district teacher reports of obstacles to using technology, with national sample results included as a reference



Notes: N = 160–163 charter teachers; N = 62–63 district teachers; N = 525 national teachers; survey question: "Please indicate whether the following conditions exist in your school and the degree to which each is an obstacle to your efforts to promote student learning using technology such as computers, smartphones, or tablets. If the condition does not exist in your school, please mark 'not applicable.'" Response choices were on a four-point scale from 1 ("not applicable; condition does not exist in my school") to 4 ("condition exists and is a major obstacle").

FIGURE 14 **Charter and district teacher perceptions of student-related obstacles to PL, with national sample results included as a reference**



Notes: N = 153–155 charter teachers; N = 62–63 district teachers; N = 525 national teachers; survey question: "Please indicate the extent to which each of the following conditions is an obstacle to your efforts to promote personalized learning for students. If the condition does not exist in your school, please mark 'not applicable.'" Response choices were on a four-point scale, from 1 ("not applicable; condition does not exist in my school") to 4 ("condition exists and is a major obstacle").

How Did NGLC Schools Affect Student Achievement?

Now our attention shifts from implementation of PL to an examination of achievement effects. To measure achievement effects over one academic year, the study analyzed mathematics and reading scores for all students in the NGLC schools who took the NWEA MAP assessments in fall 2014 and spring 2015. Similarly, to measure achievement effects over two academic years, the study analyzed scores for students who took the assessments in fall 2013 and spring 2015. MAP is an online adaptive test in which the test software 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 can provide accurate information on a broad range of student ability from kindergarten to grade 11, including how much progress a student makes over the course of a school year. Each NGLC student in the achievement analysis was matched to a set of similar students to form a "virtual comparison group" (VCG). More details are available in Appendix B.

Of the 40 NGLC schools, 32 had MAP data available for the one-year span, representing approximately 5,500 students, and 16 had data available for the two-year span, representing about 1,800 students. While these schools and students were all included when estimating effects for the NGLC schools in the aggregate, school-level results are not reported where data were available for fewer than 30 students. All of these schools implemented PL schoolwide during the years they are included in the analyses. *As discussed above, these results are not directly comparable to achievement analyses presented previously in Pane et al. (2015) because the current sample is composed of mostly secondary schools relatively new to implementing PL, whereas the sample for the prior report, on average, had greater PL experience and a majority of elementary schools.*

Overall Results

In this report, we focused our analysis of treatment effects primarily on the NGLC schools for which we have data for the most recent year (2014–15). This one-year span has the greatest number of schools and students,

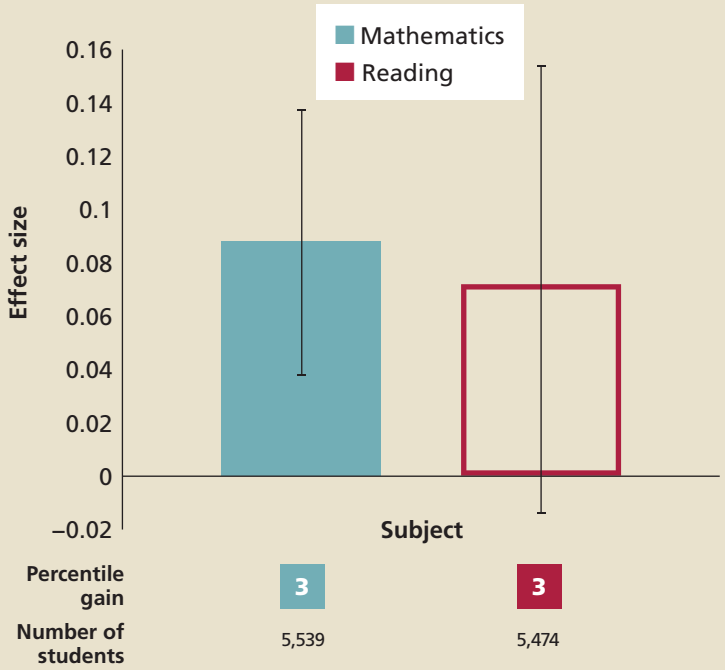
We report achievement effects of PL using effect sizes, a standard way researchers measure the impact of an educational strategy. This allows researchers to make comparisons across research studies. To assist with interpretation, we also translate the effect sizes into the percentile rank of a PL student who would have performed at the median (50th percentile) if they had been in a non-PL school.

but also includes some schools that were in their first year of implementing PL. We estimated positive treatment effects of approximately 0.09 in mathematics and 0.07 in reading, as shown in Figure 15. Only the mathematics estimate is statistically significant. These effect sizes translate to gains of about 3 percentile points; specifically, a student who would have performed at the median in the comparison group is estimated to have performed 3 percentile points above the median in an NGLC school in both subjects.

The average fall and spring student national percentile ranks are shown in Figure 16. Here, instead of using statistical models with control variables to compare NGLC students with a matched set of students (the VCGs), we simply compare their average performance to national norms for their grade. The figure shows that students started the year significantly below national norms in both mathematics and reading, and gained a modest amount during the school year. In mathematics, students gained about two percentile points but remained significantly below national norms; in reading, students also gained about two percentile points and were performing approximately at national norms by spring.

FIGURE 15

Analyses show positive effects for NGLC schools for the 2014–15 academic year

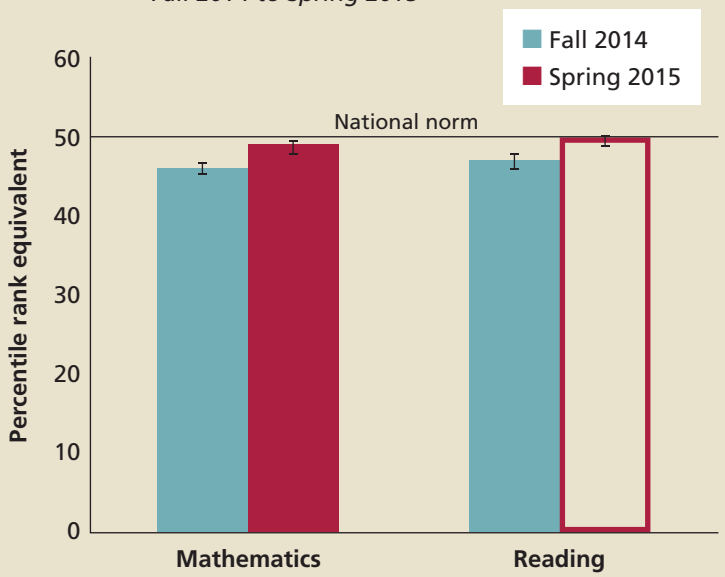


Note: Solid bar indicates statistical significance ($p < 0.05$) after adjustment for multiple hypothesis tests. Percentile gains translate the treatment effect sizes into the amount of improvement experienced by the median student.

FIGURE 16

Students started below national norms and approached them by end of the academic year

Fall 2014 to Spring 2015



Note: Solid bars indicate statistically significant differences from national norms ($p < 0.05$) after adjustment for multiple hypothesis tests.

Results Over Time

Although our primary analysis focuses on a single year of academic growth over the 2014–15 academic year, there were 16 NGLC schools that had been in operation the prior year, 2013–14, and administered the MAP assessment in both academic years. To examine growth trajectories in those schools, we restricted the sample to students with test scores in fall and spring of both academic years, and examined their scores relative to national norms. The results are shown in Figure 17. In both mathematics and reading, cumulative growth over the two years is evident. Students started significantly below national norms, gained ground after one academic year, and gained further ground the second academic year, placing them above national norms (though not statistically significantly above) at the end of two years. The largest gains on average appeared to occur in the second year, suggesting that PL systems may require some experience before operating at their fullest potential.

Results by School

Figure 18 displays treatment effect estimates for each NGLC school for the 2014–15 academic year. In each subject, we included only schools for which we had data on at least 30 students (dropping one school in mathematics). Where the estimates were statistically significant, the bars are solid. K–8 schools are colored red, elementary schools purple, middle schools orange, and high schools blue. Superscripts next to the school number indicate a district-run school. Overall, a slight majority of schools were estimated to have positive effects, though they are not always significant. Middle schools have strong representation among the schools with significant positive estimates, and many of the district schools have negative estimates. Figure 19 displays treatment effect estimates for each NGLC school in operation for the two-year span of 2013–15, and with data from at least 30 students (dropping three schools in both mathematics and reading). In both mathematics and reading, about half of the schools have positive treatment estimates.

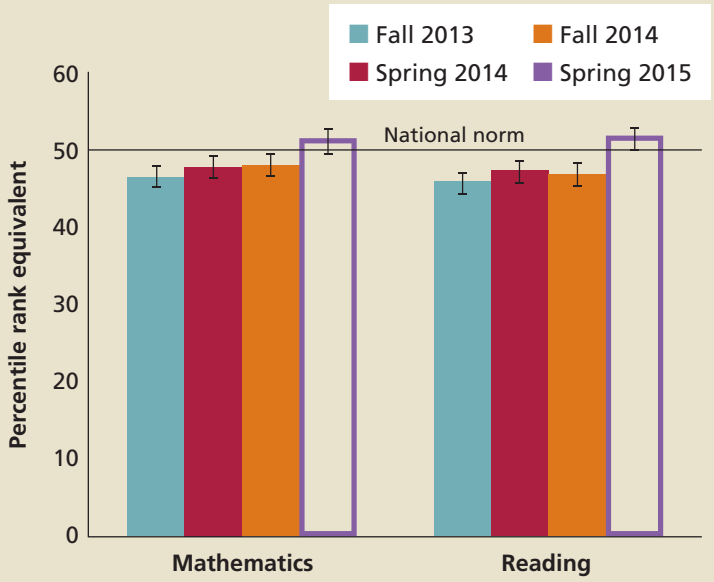
Effects for Subgroups

PL could have greater or lesser effects for various subgroups of students. In this section, we examine effects by starting achievement level, grade level, gender, and whether the school is a district-operated or a charter school. We lacked data to examine race/ethnicity or high-poverty subgroups.

FIGURE 17

Percentile rank changes over two academic years

For the 16 schools in the study that started implementing PL in 2013 or earlier, restricted to the students present for all four tests



Note: Solid bars indicate statistically significant differences from national norms ($p < 0.05$) after adjustment for multiple hypothesis tests.

Effects for Groups at Different Starting Achievement Levels

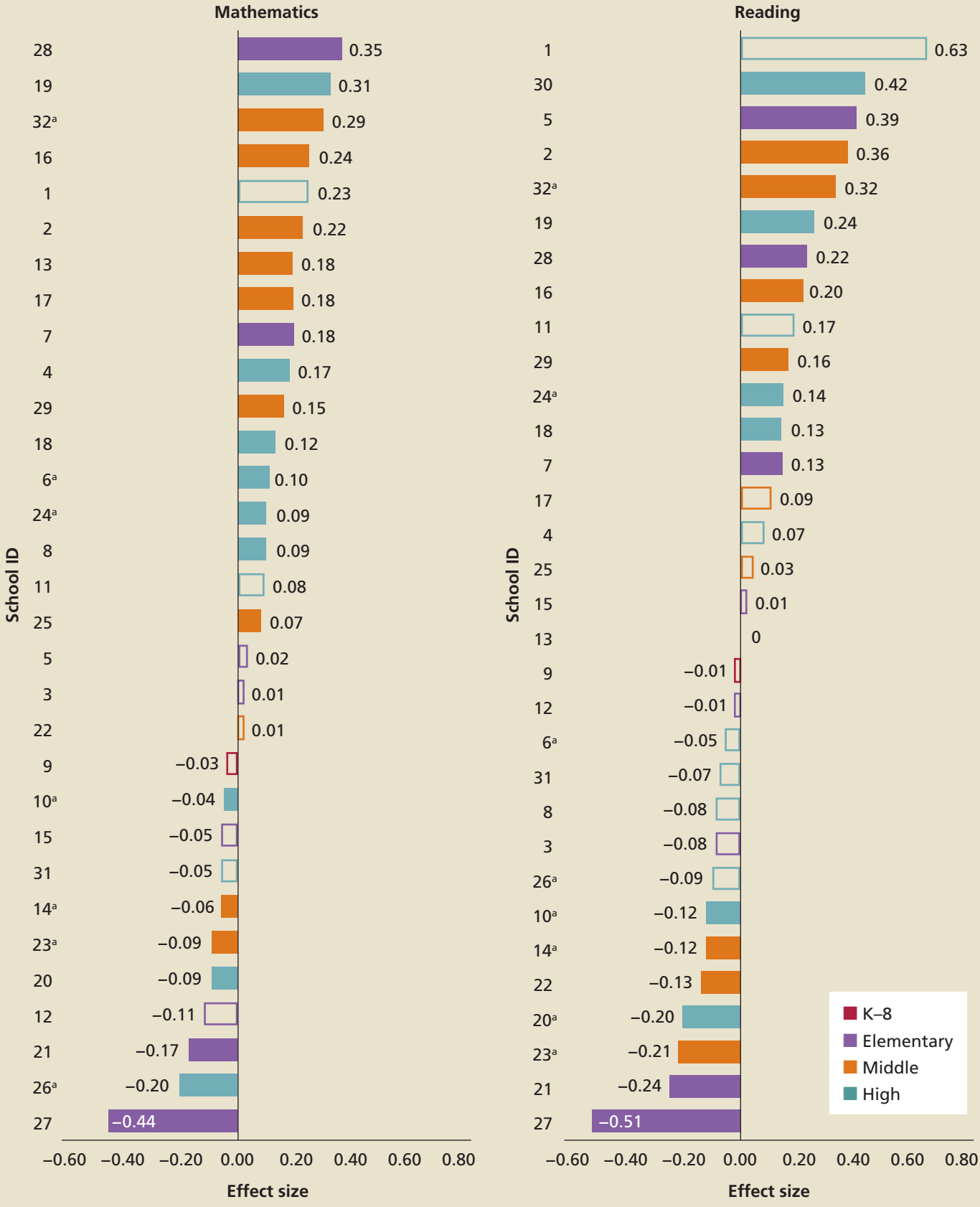
First, we examine the distribution of students in the NGLC sample based on their scores at the beginning of the 2014–15 school year. We defined five levels (quintiles) based on national norms, such that an equal number of students nationally are in each group, with each higher quintile containing students with higher fall scores. If NGLC students started out similar to national norms for their grade level, we would expect each group to hold about 20 percent of the sample. However, as we saw in Figure 16, the NGLC sample as a whole was below national norms in fall 2014. Figure 20 shows that about one-quarter of the NGLC students were in the lowest-achieving group in both subjects.

As a way of examining achievement effects within each of these groups, we calculate the fraction of NGLC students who surpassed their VCGs in raw score growth. If the NGLC sample grew similarly to students nationally, we would expect about 50 percent of NGLC students to surpass their VCGs, and 50 percent not to, simply due to random fluctuations. Figure 21 shows that across the

FIGURE 18

Treatment effect estimates for the 2014–15 academic year, by school

2014–15 effect sizes by school

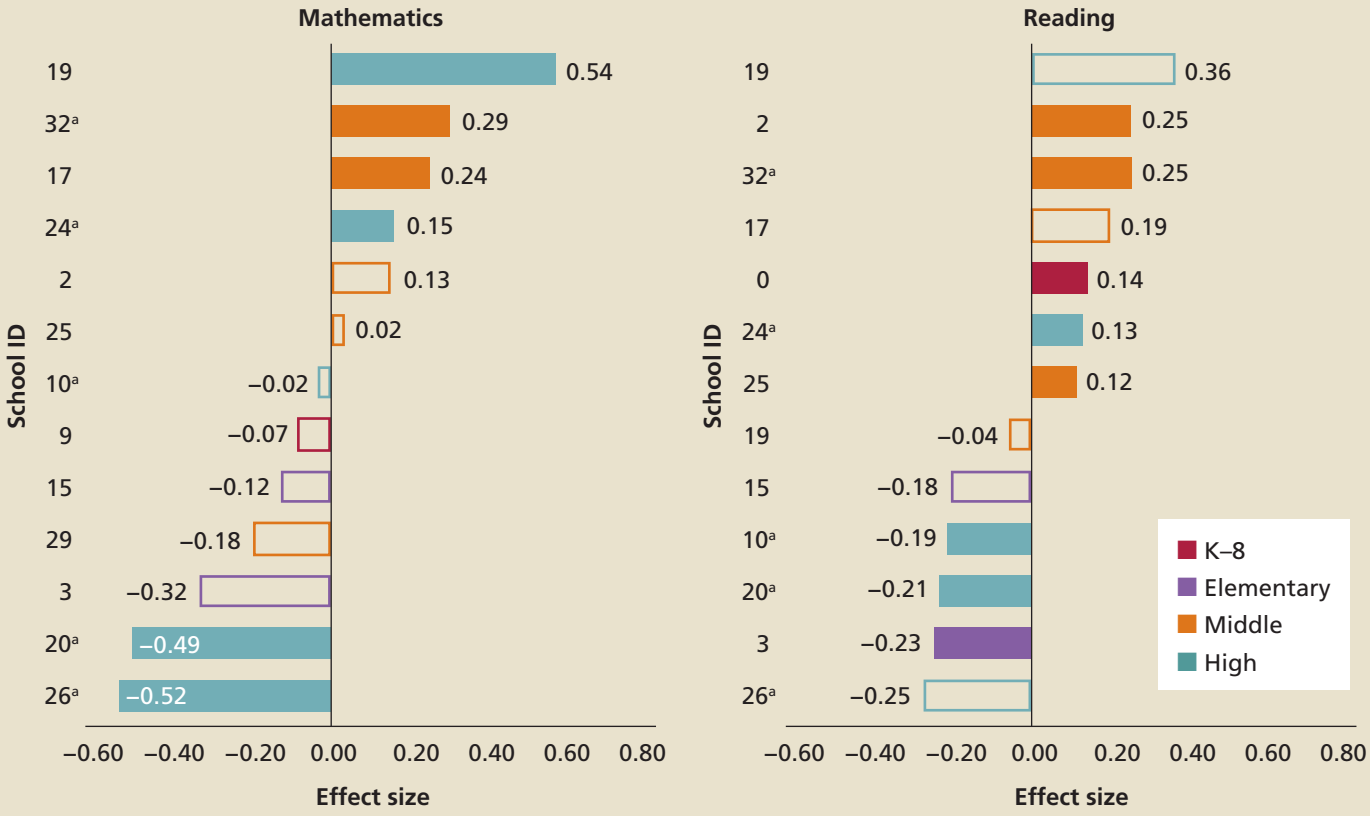


Note: Solid bars indicate significance at the $p < 0.05$ level after adjustment for multiple hypothesis tests.
[a] Indicates district schools.

FIGURE 19

Treatment effect estimates for the 2013–15 two-year time span, by school

2013–2015 effect sizes by school

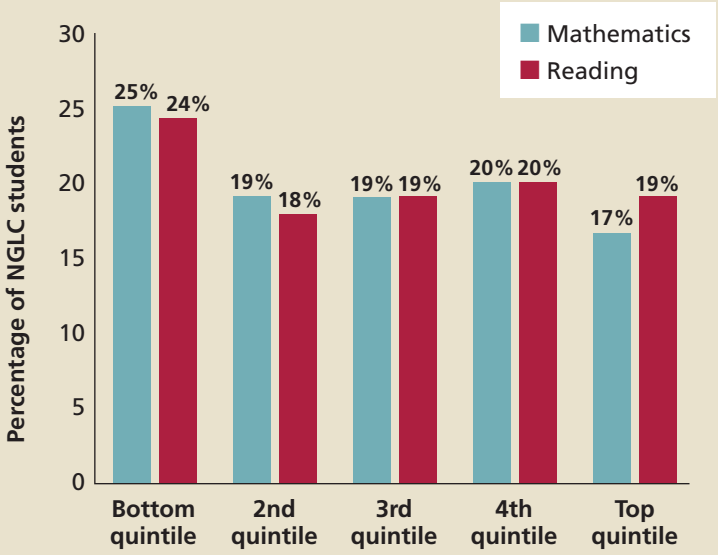


Note: Solid bars indicate significance at the p < 0.05 level after adjustment for multiple hypothesis tests.
ª Indicates district schools.

FIGURE 20

Percentages of students in each quintile of starting achievement based on national norms

Fall 2014



achievement spectrum, more than half of NGLC students surpassed their comparison students. This suggests that PL is benefiting students of all ability levels. For the lowest four quintiles, approximately 60 percent of NGLC students surpassed their VCGs in both mathematics and reading; for the highest quintile, the percentages are in the mid-50s.

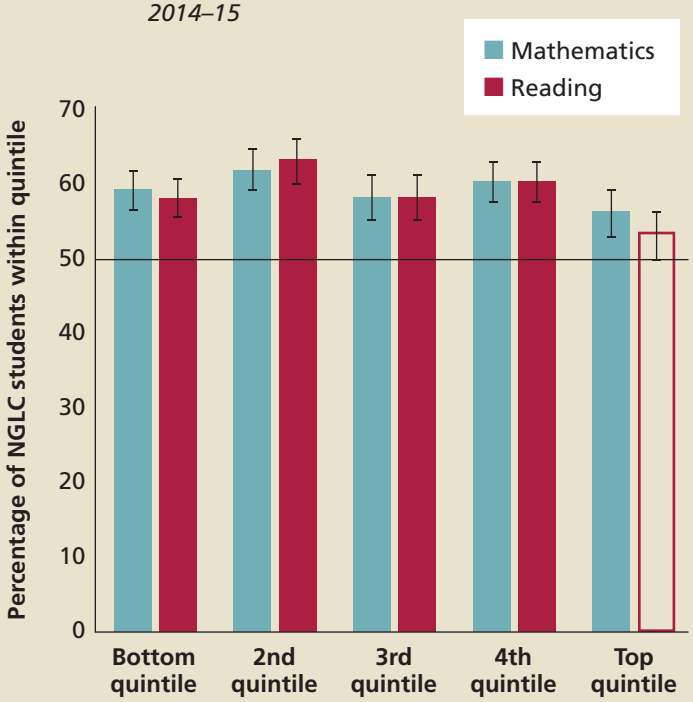
Effects by Grade Span

Figure 22 displays estimated treatment effects for three grade spans, K–5 (elementary), 6–8 (middle), and 9–12 (high). With the exception of elementary reading, the estimates are positive for both subjects, though only the middle school mathematics estimate is statistically significant. The largest estimates are for the middle school grades. These results contrast with those presented in Pane et al. (2015), where elementary schools performed the strongest, probably due to differences in the two samples.

FIGURE 21

Percentages of students surpassing their virtual comparison groups by quintiles of starting achievement

2014–15



Note: Solid bars indicate statistically significant differences from national norms ($p < 0.05$) after adjustment for multiple hypothesis tests.

Effects by Gender

We did not find evidence of differing PL treatment effects by gender. As mentioned above, limitations of the data do not enable us to examine other demographic subgroups, such as race/ethnicity or socioeconomic status.

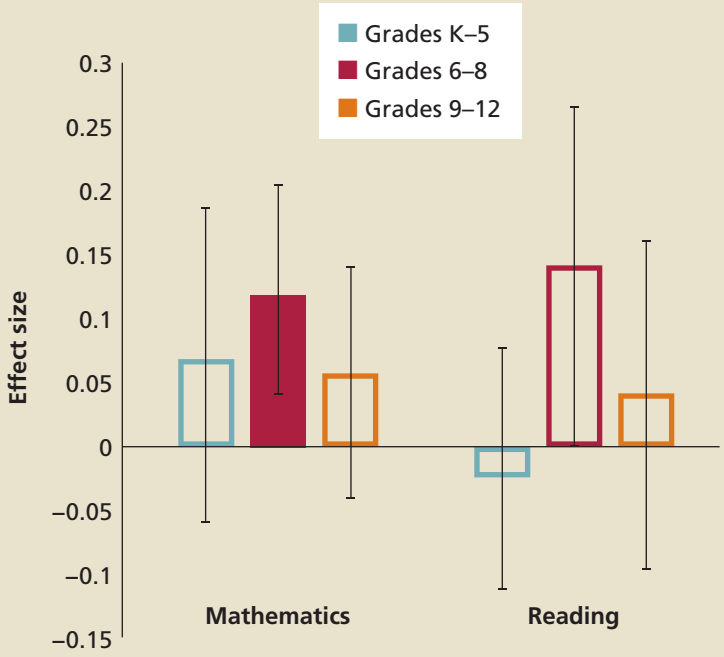
Effects in District-Operated and Charter Schools

Among the NGLC schools with achievement data, eight are operated by school districts and the remaining 24 are charter schools. The district schools are all high schools and middle schools, whereas the charter schools include schools of all levels. To review, Figure 18 showed that many, though not all, of the district schools had negative estimates of treatment effects, and Figure 22 showed that estimated treatment effects vary depending on the grade levels served by the schools.

The analysis here calculates the average treatment effects for district and charter schools separately. As shown in Figure 23, the charter schools performed similarly in both mathematics and reading, with estimated effects near 0.10 (only the mathematics estimate is significant). The district schools have smaller estimates—about half as large in mathematics and near zero in reading. Due

FIGURE 22 **Estimated effects for elementary, middle, and high-school grade bands**

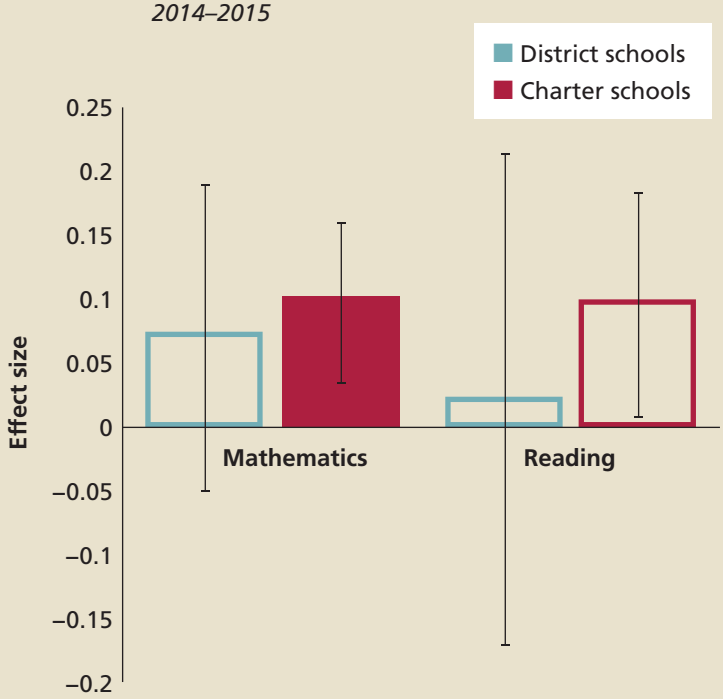
2014–2015



Note: Solid bar indicates statistical significance ($p < 0.05$) after adjustment for multiple hypothesis tests.

FIGURE 23 **Estimated effects for district-operated and charter schools**

2014–2015



Note: Solid bar indicates statistical significance ($p < 0.05$) after adjustment for multiple hypothesis tests.

to small samples, the district estimates are particularly imprecise. This imprecision, along with the differences in the grade levels served by the two groups, suggests that these trends in district versus charter schools should be treated as suggestive but not conclusive.

These Findings Withstand a Series of Rigorous Sensitivity Analyses

To help evaluate the robustness of the findings discussed in this chapter, we performed a variety of sensitivity tests. These included analyses based on national norms of growth, restricting the VCGs to come from the same school type (charter or district) as the corresponding NGLC school, and examining the effect of test duration on results. The rationale, methods, and results of these tests are discussed in Appendix B. After evaluating the results of these sensitivity tests, we concluded that they support the results presented here and the substantive conclusions we are able to draw.

Implications and Policy Recommendations

Review of Key Findings

- NGLC teachers reported higher levels of implementation than teachers nationally on some aspects of PL.
- NGLC schools looked more like the national sample on some more-difficult-to-implement aspects of PL.
- Barriers to PL implementation included poor integration of data systems, tensions between competency-based practices and meeting grade-level standards, and the time needed to develop personalized lessons.
- Students in NGLC schools experienced positive achievement effects in mathematics and reading, although the effects were only statistically significant in mathematics.
- On average, students overcame gaps relative to national norms after two years in NGLC schools.
- Students at all levels of achievement relative to grade-level norms appeared to benefit.
- Results varied widely across schools and appeared strongest in the middle grades.
- Implementation and effects of PL seemed higher in charter schools than in district schools in the sample.

Implications

Although advocates and reformers have developed PL models, many of the component practices are relatively common nationally, making it difficult to clearly identify what makes a school a PL school. At a theoretical level, PL is very different from the instructional approaches that have been typical in K–12 schooling in the United States. It puts a primary focus on identifying each individual student's strengths, needs, goals, and progress; uses those to provide appropriate and meaningful individualized instructional experiences with the necessary adult supports; removes constraints on what students work on, when, and for how long; and reallocates resources to best support these processes.

In this theoretical conception, schools that are high implementers of PL approaches would look very different from more-traditional schools. In practice, although there were some differences between the NGLC schools and the national sample, we found that schools in our study were implementing PL approaches to a varying degree, with none of the schools looking as radically different from traditional schools as theory might predict. This is due in part to the schools trying various combinations of strategies and features, rather than all of them; to the newness of the schools in our study (most of which had been in existence for less than three years); and to external constraints, such as state or district policies. Despite the lack of clear differences in the practices that teachers reported implementing in NGLC schools and the national sample, it is important to note that all of the NGLC schools had adopted structures and systems to support PL within three years, and all of them reported striving to emphasize personalization in their school designs and operations.

At the same time, many of the core practices of PL were also implemented to some extent in the national sample, which consisted of schools that do not clearly identify themselves as PL schools. These factors make it difficult to draw a clear line separating PL from non-PL

schools. And, although early evidence suggests that these PL approaches may be quite promising for improving achievement for a broad range of students, at this early stage in the development of the innovation it is not clear what PL practices or combination of practices have the greatest impact on students.

There is suggestive evidence that greater implementation of PL practices may be related to more-positive effects on achievement; however, this finding requires confirmation through further research. First, the NGLC sample shows both positive treatment effects and higher levels of implementation than schools in the national sample. Second, within the NGLC sample, there is a trend toward smaller estimated effects for district schools as compared with charters, and this is accompanied by lower levels of implementation of PL practices among district schools. Specifically, the reason students in the NGLC schools are outperforming students in the national comparison group may be because NGLC schools are implementing more PL strategies or implementing the strategies to a greater extent than schools nationally; and this may also explain why charter-school students tend to outperform district students within the NGLC sample. Third, we see increasing effects on student achievement with longer exposure to PL, as shown in Figure 17.

Although consistent with the data, this hypothesis is somewhat speculative and should be interpreted cautiously for several reasons. First, the number of district schools in the implementation sample was small—nine out of 40—and thus the differences we observe may not be broadly generalizable to schools implementing PL. Second, we have limited information about schools in the national comparison sample, such that differences in composition (such as the proportion of charters) could influence the NGLC–national comparisons. We also cannot rule out the possibility that students in the NGLC sample—and in charter schools in particular—are experiencing greater achievement gains for reasons that are not related to the implementation of PL strategies. It is also possible that PL may be more challenging to implement in district schools. We do not have any evidence that could confirm or elucidate this hypothesis, and it is possible that district schools could see the effects of PL if their practices were more consistent with those reported in charters. In addition, our observations about the extent of implementation of PL practices rely largely on surveys of teachers and students, where responses are self-reported. Not only do we have no objective way to confirm their perceptions, but the problem of reference bias (West et al., 2016) could make comparisons among responses difficult to interpret. Finally, we do not have any information about the extent to which schools in the achievement comparison group were implementing PL strategies.

In light of these limitations, this apparent relationship between the extent of PL implementation and student outcomes should be interpreted cautiously. The question of why district schools in the NGLC sample did not seem to be implementing PL approaches to the same extent as charter schools, however, is an important one that merits further investigation.

The positive student outcome effects found in this study may not occur quickly or in all contexts. As policymakers, practitioners, and funders think about how they could use the results of this research to enhance, expand, and support implementation of PL, it is important to keep in mind that the positive student outcome effects found in this study may not occur quickly or in all contexts. The earlier report from this study (Pane et al., 2015), which found statistically significant positive effects for PL schools in mathematics and reading, focused on a sample of 62 schools, many of which were experienced implementers of PL and part of large, well established charter networks. This report focuses on a smaller sample of schools that were newer (most had been open for less than three years) and, although mostly charters, were generally starting new networks or were part of smaller networks. While the effects reported here are generally positive, they are smaller, overall, than those in the earlier report. Furthermore, the effects are smaller for district schools than for charter schools in both reports. Taken together with the implementation data presented here, it is possible that implementing PL in new schools and in district schools may be more challenging than in other contexts. Therefore, as PL strategies become more widely used and studied, it is possible that not all schools will see gains as large as those in the current sample or the sample examined in Pane et al. (2015).

Policy Recommendations

In this study, we found that schools were implementing PL approaches to varying degrees, although all were attempting to implement PL strategies to some degree. This could be because schools chose to implement some PL strategies but not others, or because the schools in this sample, which are predominantly new schools, planned to implement more PL strategies as their schools grew to full capacity. It could also be because there are barriers at the local or state levels that cause variation in the ability to fully implement PL strategies.

Yet the results of this study suggest that PL has positive effects, and there is a lot of momentum in the field to spread it. Given the results of this study, and those described in Pane et al. (2015), we offer the following recommendations for policymakers, implementers, and funders of PL that could help to support broader implementation of PL and enhance our understanding of how to implement it effectively.

For State and District Policymakers

Incorporate flexibility into policies related to course progressions. Some PL strategies involve allowing students to complete specific standards or sections of a course as one way of helping them catch up to grade level, and others rely on allowing students to progress to new content as they demonstrate competency. NGLC teachers and school leaders described using such practices but also requiring students to get through a certain amount of material. In addition, NGLC principals reported that they were not yet able to award credit for mastery of specific standards or sections of a course in a way that would prevent students from having to repeat that material if they transferred to other schools. State or district policies that mandate courses be completed in their entirety, in a specific order, or at a certain grade level can inhibit full implementation of such strategies. Taking a more flexible approach could also include revising policies to allow inclusion of multidisciplinary courses or projects, or revising or eliminating seat-time policies. Such policies, while remaining flexible, should ensure that all students are exposed to rigorous, high-level content and should be monitored to ensure equitable outcomes.

Allow school staff to have some autonomy to design school schedules that support PL. Flexibility to design a schedule that supports the school model and vision of PL, and modify it as needed over the course of the year, can be a key component of successful PL implementation. While some safeguards should be in place to ensure adequate coverage of content in compliance with state standards, this flexibility might entail allowing schools to implement a longer school day or year, customize the length or number of class periods, or develop multidisciplinary classes or project-based classes. District policies that require uniform school schedules could inhibit this flexibility. District leaders and other stakeholders might seek ways to enable flexibility in schools that are willing to innovate.

Enable schools to hire staffs that are the best fit for the school. As we described in Pane et al. (2015), NGLC administrators reported that a key challenge was hiring and retaining teaching staff with both the right level of skills and experience and who were a good fit for the school model. Although it is not yet clear which qualifications are most important to consider when staffing PL schools, having the flexibility to hire staff that support the school model and fostering working conditions that support retention both seem important. Policymakers could consider revising teacher placement or hiring provisions, revising policies to support hiring staff in nontraditional roles, and allowing schools flexibility in some work rules, within legal and contractual limits, to enhance teacher retention.

Ensure that accountability policies value growth and other metrics of student success. As we reported in Pane et al. (2015), several of the NGLC schools reported policy challenges related to implementing competency-based approaches. One example was state tests focused on grade-level standards, which might not be sensitive to learning content above or below grade level, and which could have accountability consequences for teachers or schools. Policymakers should consider refining their systems to use assessments that can accurately measure growth in learning and to use growth-related metrics for accountability. States could take advantage of the flexibility offered by the accountability provisions of the *Every Student Succeeds Act* (ESSA), which allow for the inclusion of statewide academic indicators that measure growth rather than relying exclusively on proficiency. ESSA also permits states to adopt an additional indicator of "school quality or success," which could include measures that might capture useful information about schools adopting PL, such as access to advanced course content or postsecondary readiness. These indicators could provide a broader and more-balanced set of information with which to gauge the performance of schools implementing PL models.

Revise grading policies to incorporate competency-based approaches, and clearly communicate these approaches to students, families, employers, and postsecondary education institutions. Competency-based learning strategies often require nontraditional approaches to grading student work or judging student readiness to progress through content and may need to include ways to assess learning using multidisciplinary coursework or projects. Innovative schools should have the flexibility to develop nontraditional grading systems that support the school model, and policymakers could consider limiting the need to convert grades back to a traditional letter grade for reporting purposes. At the same time, nontraditional grading systems can be challenging to understand, so policymakers could work to ensure that the school has the necessary resources to clearly communicate their grading system to internal and external audiences.

Look to early adopters of PL for examples of large-scale policy change. Policymakers at the state level interested in exploring some of these recommendations could look to states such as Vermont, New Hampshire, and Kentucky, all of which are working to revise state policies to support PL at the local level and implement PL strategies statewide. District-level policymakers could look to districts such as Fulton County, GA, Piedmont, AL, and Horry County, SC, which are working to implement PL strategies districtwide. Policymakers at all levels could also look to charter management organizations such as Rocketship and Summit, which aim to incorporate PL approaches in their schools.

For Implementers at the District and School Levels

Provide teachers with the resources and time to pilot new instructional approaches and gather evidence of how well they work. As we argue elsewhere in this report, it is not yet clear which PL strategies and practices are most likely to positively affect student outcomes. Therefore, it is important to ensure that teachers and school leaders have the flexibility, time, and resources (e.g., funding, support staff, access to experts) to experiment with new instructional approaches, develop a systematic process for collecting and analyzing evidence of their effectiveness, and make changes as needed.

Provide teachers with time and resources to collaborate on developing curriculum and on reviewing and scoring student work. If the school staff prefer to develop their own curriculum materials, it is important to ensure that teachers have the flexibility, time, and resources to collaborate on curriculum development and score student work in ways that are minimally intrusive on their teaching duties. Time to collaborate on scoring student work is particularly important in schools that use mastery-based grading systems, where the system's norms and parameters may still be in development.

Identify a school staff member (or two) who is comfortable with technology and has curriculum expertise to serve as a just-in-time resource for teachers. Some technology resources have the potential

to enable key PL strategies, but integrating technology into instruction can often be challenging for teachers. It is therefore important that schools identify one or two staff members who have the ability to support teachers in troubleshooting technology issues as they arise, creating technology-integrated lessons and projects, accessing and interpreting data from technology-based curriculum materials, and developing classroom management plans to include technology.

Provide resources and support for school staff to help them choose the most-appropriate digital or nondigital curriculum materials. Many NGLC schools reported that finding standalone technology-based curriculum programs of high quality that were well suited for the school context was challenging. As a result, many schools tended to rely on multiple technology-based programs and teacher-developed materials, a situation that can make developing lesson plans time-consuming for teachers. In addition, the lack of curricula designed to meet the needs of students performing at different levels can hinder teachers' efforts to personalize instruction. Ensuring that school staffs have the necessary resources (e.g., time, funding, extra staff) and support (e.g., access to curriculum experts or other means of vetting, adapting, or combining materials) could help ease the burden of curriculum development for teachers, allowing them to focus more time on instruction.

Provide resources and support for school staff to integrate multiple data systems. Although technology is a key enabler of PL, another barrier to widespread, effective PL implementation is that some technologies have not yet developed to the point where they support PL by making some aspects of teaching more efficient. For example, many school data systems in use in PL schools do not yet integrate achievement and nonachievement data, shifting the burden of integrating and interpreting those data onto teachers. Providing resources or support could help ease the burden of data entry and integration for teachers, allowing them to focus more time on instruction.

For Funders

Direct funding to technology developers who will work with teachers and curriculum experts to design technology-based curriculum materials and data systems that will support PL practices. For example, such efforts could include curriculum programs that incorporate multiple paths through content and include high-quality assessments of competency. Ideally, such materials and systems would be adaptable to students at a variety of learning levels and integrated with student information systems to provide a complete picture of each student's goals and progress.

Allocate funding for research that includes stronger experimental designs and that systematically tests specific PL strategies. As funders continue to invest in PL, and administrators continue to adopt the strategy in states, districts, or other groups of schools, intentional program design can enable more-rigorous evaluation methods than were available for the current study. In particular, implementing a well-defined PL model in a sample of schools, with half of the sample randomly assigned to begin immediately, and the other half serving as a control group for a set period of time, can enable rigorous causal estimates of PL effects. Such a design can rule out concerns of selection bias—that factors other than PL are responsible for the effects measured in PL schools. Moreover, a clearly defined model of implementation for PL schools can help to clarify uncertainties about how, and to what extent, PL differs from more-traditional practice, and, if results are positive, enable clear specification of a model for replication and scale-up.

Acknowledgments

The authors are grateful to the PL students, teachers, and administrators who voluntarily participated in project data collection; NGLC staff who facilitated connections with the schools; and NWEA staff who responded to our multiple requests for assessment data and helped us interpret them. We are also grateful to the following RAND staff who contributed to the research: Mollie Rudnick, Courtney Kase, Amanda Edelman, Andrea Bingham, Katharina Best, Christopher Paul, Evan Peet, Kyle Siler-Evans, Gerald Hunter, Melanie Rote, Suzette Gambone, and Stephanie Lonsinger. This document is vastly improved as a result of feedback we received on earlier drafts from Chandra Garber, Paul Hill, Cathy Stasz, Julia Kaufman, Patrick McEwan, and Brad Bernatek and his colleagues at the Bill & Melinda Gates Foundation. Any flaws that remain are solely our own.

References

Bloom, Benjamin S., "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring," *Educational Researcher*, Vol. 13, No. 6, 1984, pp. 4–16.

Brodersen, R. Marc, and Daniel Melluso, *Summary of Research on Online and Blended Learning Programs that Offer Differentiated Learning Options*, Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central, 2017. As of June 1, 2017: https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=4499

Iacus, Stefano M., Gary King, and Giuseppe Porro, "Causal Inference Without Balance Checking: Coarsened Exact Matching," *Political Analysis,* Vol. 20, No. 1, 2012, pp. 1–24.

Opfer, V. Darleen, Julia H. Kaufman, and Lindsey E. Thompson, *Implementation of K–12 State Standards for Mathematics and English Language Arts and Literacy,* Santa Monica, Calif.: RAND Corporation, RR-1529-1-HCT, 2016. As of June 1, 2017: http://www.rand.org/pubs/research_reports/RR1529-1.html

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 1, 2017: 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*, 2016.

Thum, Yeow Meng, and Carl H. Hauser, *NWEA 2015 MAP Norms for Student and School Achievement Status and Growth*, Portland, Oreg.: Northwest Evaluation Association, August 16, 2015. As of June 1, 2017: http://www.sowashco.org/files/department/rea/2015NormsReport_Reading.pdf

VanLehn, Kurt, "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems," *Educational Psychologist*, Vol. 46, No. 4, 2011, pp. 197–221.

West, Martin R., Matthew A. Kraft, Amy S. Finn, Rebecca E. Martin, Angela L. Duckworth, Christopher F. O. Gabrieli, and John D. E. Gabrieli, "Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling," *Educational Evaluation and Policy Analysis*, Vol. 38, No. 1, 2016, pp. 148–170.

APPENDIX A. Implementation Analysis Methods and Limitations

To explore what PL looks like, we drew on interviews with teachers and administrators, focus groups with students, and classroom observations conducted during in-person site visits to nine NGLC schools to present brief vignettes that highlight exemplar implementation of the four key PL strategies.

To explore how PL practice in NGLC schools differs from practice nationally, we compared teacher and student survey responses from the NGLC schools with those from surveys with nearly identical questions administered to national samples of teachers and students.[1] To facilitate this comparison, we first weighted the national survey results to more closely reflect the NGLC sample in terms of geographic locale (e.g., urban), grade level, subject taught (by teachers), and gender (of students). We lacked the necessary data to include family income in the student survey weighting process, and the national sample appears to be somewhat more affluent than the NGLC sample. Moreover, the NGLC surveys were conducted in the spring, and the national surveys were conducted in the summer; responses may have been affected by differences in how clearly respondents could remember details about the practices and experiences we inquired about. We also present additional evidence of PL implementation from the NGLC administrator interviews, site-visit interviews, and teacher logs.

We observed a trend toward larger PL treatment effects in charter schools than in district schools. We therefore examined our implementation data separately for charter and district schools, where feasible, to explore questions such as whether levels of implementation differed in these two contexts or whether factors that hindered or facilitated implementation differed.

Throughout this report, we used a holistic approach when deciding what information to present. We focused on presenting meaningful evidence of differences (or similarities) between implementation of PL and more-traditional practice. Where we were able to perform tests of statistical significance we used those results to guide our decisions about what material to present.

In some cases, we describe differences that were not statistically significant but that were large in magnitude and qualitatively meaningful in that they shed some light on substantive questions about implementation. Student and teacher survey results can be found in the online addendum to this report.

When interpreting the implementation data, it is important to keep in mind the limitations of the data sources, which rely on the self-reports of stakeholders who voluntarily participated. We had no independent means of verifying the accuracy of their responses. Where response rates were lower, particularly for the teacher survey and logs in some schools, responses may not accurately represent the perceptions of the whole stakeholder group, limiting generalizability. Survey responses likely vary across several factors, such as grade-level configuration (e.g., elementary versus secondary schools), but we avoided breaking down the data by these features because of the small numbers of respondents in some categories. Additionally, the self-reported nature of the surveys may limit their ability to accurately measure differences across schools. As just one example, West and colleagues (2016) have documented a phenomenon known as *reference bias*, where responses can be influenced by the respondent's frame of reference or social context. To illustrate, a student might answer a question about being "given opportunities to demonstrate my strengths and weaknesses," with the response option "mostly true." The actual amount of such opportunities necessary to meet the threshold of "mostly true" can vary from student to student, influenced by their own experiences as well as the norms of the school or the attitudes of their peers. Thus, two respondents who responded "mostly true" might actually be experiencing different levels of these opportunities. This can reduce the validity of comparisons of responses between groups (such as students in NGLC schools versus a national sample, or teacher versus student reports of a PL implementation feature).

[1] The national survey was administered by Grunwald Associates. The items were identical to those administered in the NGLC schools except for tense; the NGLC items were in present tense because the survey was administered in the spring, whereas the national survey items were in past tense, referring to the previous school year, to reflect the fact that students and teachers completed it during the summer.

Although we weighted the national student and teacher surveys to make the respondent profiles more similar to the PL samples, data limitations prevented us from doing so with respect to family income, limiting the comparability of the student survey samples. We opted not to include years of teaching experience when weighting the teacher survey to allow for the possibility that hiring less-experienced teachers was something some NGLC schools did intentionally (in other words, reliance on newer teachers might be considered part of the NGLC approach to operating schools, rather than an extraneous factor that we would want to control for). We observed differences in the mean years of teaching experience in each sample (3.9 years of experience for teachers in the national sample; 2.7 years for teachers in the NGLC sample) and this difference could affect responses in ways that are not related to the implementation of PL practices. We also have no information about the extent to which the schools in the national sample were implementing PL strategies, and therefore comparisons between the national sample and the PL sample should be interpreted with these limitations in mind.

The small number of district schools in the implementation sample (nine) limits our ability to make reliable comparisons between the district and charter schools in the NGLC sample. Moreover, the teacher workforce appears to differ between district and charter schools, which could affect responses through reasons other than the implementation of PL practices. Overall, charter teachers were less experienced than district teachers: 22 percent of charter teachers reported having one year of experience or less, compared with 5 percent of district teachers. Charter teachers were also more likely to have received their certification through a nontraditional program (27 percent, compared with 14 percent of district teachers). For those reasons, district–charter comparison results should be interpreted with caution.

Site Visits

We conducted one-day site visits at nine schools in spring 2015. The visits included a one-hour interview with the principal, 45-minute individual interviews with three instructional staff, a one-hour focus group with six to eight instructional staff, a one-hour focus group with six to eight students, and 10- to 15-minute observations of at least two classrooms, one mathematics and one English language arts (ELA). The purpose of the site visits was to gather in-depth information about implementation of the school model and instructional practices and to solicit student perspectives.

Site visit schools were selected based on fall 2014 administrator interviews and documentation. We purposefully selected schools that varied on several dimensions: the extent to which the school was implementing competency-based progression, extent to which the school was implementing technology-based PL, grade configuration, and organizational structure (e.g., a school that was part of a charter management organization versus one administered by a traditional district). Teachers were randomly selected for the interviews and focus groups so that there was some variation across grade level taught, subject taught, and years of teaching experience. Students were selected for the focus group by a school administrator so that the group would include students with a mix of ages and learning levels, as well as students from both genders.

Administrator Interviews

We interviewed an administrator by telephone at each school, district, or charter management organization in the fall of the 2014–15 school year. We conducted a second set of telephone interviews in the spring with an administrator at the school level, usually the principal or assistant principal. At site-visit schools, the spring administrator interviews were conducted in person. The interviews helped gather other information about instructional practices, including what types of technology the school was implementing, whether the school used standards-based grading, and whether there were opportunities for learning outside of school. The interviews lasted one hour.

Teacher Logs

Teachers of mathematics and ELA were asked to complete logs, which were brief, online surveys that included questions about daily instructional practice and the factors that influenced their teaching on a particular day. We administered the logs over two 10-day periods in 2014–15, once in the fall and once in the spring, for a total of 20 logs per teacher. In the fall, the logs were distributed to a sample of 331 teachers, and 228 teachers completed at least one log in which they indicated that they had provided instruction that day, for a response rate of 69 percent. In the spring, the logs were distributed to a sample of 330 teachers, and 189 teachers completed at least one log in which they indicated that they provided instruction that day, for a response rate of 57 percent.

The number of logs completed varied by teacher; missing logs were due either to a response of "I did not provide instruction today" or to noncompletion. Each day, teachers answered a series of questions while focusing on their interactions with one student during the first 45 minutes of mathematics or ELA instruction. Teachers were asked to focus on a different student for each day that they completed the log. The rationale for asking teachers to focus on a single student rather than the entire class is that the instruction offered, and the nature of the student-teacher interactions, can vary across students. This variability is particularly likely to occur in PL environments.

Teacher Surveys

Teachers of mathematics and ELA were also asked to provide their perceptions about various aspects of the models, including professional training and support, access to resources, the quality of instructional and curricular 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 survey was distributed to a sample of 330 teachers and the response rate was 74 percent. The teacher surveys were administered online in spring 2015. Although most of the survey items were developed specifically for this study, a few were adapted from other RAND surveys or from surveys developed by the University of Chicago Consortium on Chicago School Research (CCSR).

Student Surveys

Students were asked to describe their study habits, attitudes toward learning, perceptions about their school, level of access to technology, and other topics. The student surveys were administered online in the fall and spring of the 2014–15 school year to students in 36 schools with enrolled students who met the age criteria: grades 6 and above or age 11 and older if the school did not use traditional grade levels. The fall survey focused on study habits and attitudes toward learning; the spring survey supplemented these with the remaining topics. Student responses to items that appeared on both surveys were similar, so this report focuses on the spring results that cover a broader range of topics. We distributed the fall survey to 9,294 students and the spring survey to 9,058 students. Response rates were 71 percent and 69 percent, respectively.

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 CCSR surveys; the High School Survey of Student Engagement, developed by the Center for Evaluation and Education Policy at Indiana University; and the Tripod survey, developed by Harvard University's Ronald Ferguson, to measure student opinions of teacher quality.

National Surveys

To provide comparative data for our teacher and student surveys, the Bill & Melinda Gates Foundation engaged Grunwald Associates to administer the surveys to a national sample. Those surveys were administered during the summer after the 2014–15 school year. The questions on the survey were nearly identical to those on our surveys, although the language was adapted to refer in the past tense to the 2014–15 school year.

Analysis of Interview and Focus-Group Data

The analysis of the interview and focus-group data proceeded in several steps. First, interview notes were compared to the audio recording and cleaned to serve as a near-transcript of the conversation. The cleaned interview notes were then loaded into the qualitative analysis software package NVivo 10 and auto-coded by interview question (so that responses to specific interview questions were easily accessible) as well as coded using a thematic codebook developed by the evaluation team. 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 competency-based progression?). In this stage, we used an inductive coding process (i.e., codes were derived from the data rather than a structured codebook) to develop responses to the questions of interest.

APPENDIX B. Achievement Analysis Methods and Limitations

This study is designed to use the most rigorous method that can be applied to the situation. In particular, given the portfolio of NGLC schools, it was not possible to create randomly assigned treatment and control groups; nor did we have access to data from neighboring schools that might have matched the NGLC schools. Moreover, as new schools, they lacked a history of data from before they began implementing PL, which would have enabled other analytic methods for determining achievement effects. With these limitations, we determined that a matched comparison group design is the best available quasi-experimental method for estimating the effect of NGLC schools on student outcomes. If the NGLC students can be matched to comparison students who are equivalent at baseline, this method can produce unbiased estimates of the NGLC effect.

To create the matched comparison group, NWEA drew on its large national database of testing data to identify VCGs—comparison groups of students who had starting performance similar to the PL students and who were attending schools serving similar populations. Details about the matching method and the statistical models we used to estimate results are described below. This process enables us to make "apples-to-apples" comparisons of learning growth between the students in the PL schools and a similar population of students attending other schools.

Limitations of Achievement Analysis

However, there are limitations to this method. Although, as detailed below, we find that the observable characteristics of the comparison students are well matched to those of NGLC students in the study, the comparison students could possess other unidentified or unobserved differences from the NGLC students. Those differences could confound efforts to measure the impact of PL. For example, parents of NGLC students might have greater interest in nontraditional schooling environments and this could be related to how well their children do, independently of the NGLC schools' PL treatment. Differences like this are a type of selection bias that could affect estimates of treatment effects, in either a positive or a negative direction. The VCG approach also assumes that the students in the comparison group are attending more-traditional schools that are not using PL practices, but there is no way to verify this assumption. If this assumption is not true—if any of the comparison schools were indeed using PL practices—estimates comparing NGLC students to VCG students could underestimate the magnitude of the effect. Because of these limitations, achievement results should be interpreted with some caution.

While the basic empirical strategy remains the same between this report and Pane et al. (2015), there were some refinements that arose from obtaining an updated and richer data set from NWEA. The first change related to the matching algorithm performed by NWEA, and other changes were to the analytic methods described below.

Matching Method for Virtual Comparison Group

Previously, when searching for matches to create VCGs for students in one PL school, NWEA would require all matches to come from schools outside the same governing organization (generally speaking, a school district or charter management organization). This restriction prevented contamination of the control group with other PL students in the same governing organization, but conceptually could enable PL students from one governing organization in the study to be included in the VCG for a different governing organization. At our request, NWEA updated their VCG-matching algorithms to exclude students from any governing organization in the study. As such, the updated matching algorithm was as follows: For each NGLC student, NWEA created a VCG of up to 51 students from its database. Separate comparison groups were created for the mathematics and reading tests and for each time span examined. The analysis uses fall scores as pretests and spring scores as posttests (from the same academic year for one-year analyses, and from

the following academic year for two-year analyses). The following student and school-matching criteria were applied to create the VCG.[1]

REQUIREMENTS FOR ALL VCG MATCHES

- Students have valid scores for the pretest and the posttest.
- Students are not in any of the governance organizations containing schools in the PL sample.
- Schools have the same locale classification (e.g., urban, suburban, rural, etc., according to the National Center for Educational Statistics Public School Universe Survey).
- Students are the same gender and in the same grade as the treatment-group students to whom they are matched.

APPROXIMATE MATCHING CRITERIA

- Schools differ by no more than 15 percentage points on the portion of students participating in the FRL program.
- Students scored similarly on the pretest MAP assessment. Preference is given for students with the same pretest score, but this can be expanded to within five points on NWEA's RIT scale if necessary to find matches.[2]
- Number of days elapsed between the pretest and posttest testing differs by no more than 18 days.

Refinements to the Statistical Estimation Strategy

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 PL student and we could account for this duplication in our analysis. To do so, we now use a type of Coarsened Exact Matching (CEM) estimator (Iacus et al., 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 similarities in observables across several dimensions together, instead of collapsing the matching space into a univariate distance metric, as is done with propensity 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 1 shows the definition of these weights.

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

where i indexes students, j indexes each VCG group student i appears in, and $|VCG_j|$ is the number of VCG students in that group; T_i is a treatment indicator equal to one for PL students and zero 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, the data set is reduced to having one observation per student test score, instead of retaining multiple records, as was done with the within-estimator used in Pane et al. (2015). The weights are then applied in a weighted linear regression, as described below. The CEM estimator used here departs slightly from that of Iacus et al. (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 approaches 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 a

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

[2] NWEA's RIT (Rasch Unit) 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.

standard deviation of one within each grade level, and posttest scores reflect the standardized growth. Because of small samples and volatility of scores in the highest grades, we classified grades 11 and 12 into a single "late high school" group for the grade-level indicators. We then divided the standardized growth by the number of days elapsed between pretest and posttest, to account for variation in the time elapsed, to obtain a standardized measure of growth in achievement per day. We regressed the standardized growth in achievement per day on treatment status and the following covariates: an indicator of whether the school is district-operated, 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 are included in the regression, but are implicitly controlled for.

In a second change to the analysis, we now use a clustering algorithm and degrees of freedom estimators that are more robust when there are small numbers of clusters (Pustejovsky and Tipton, 2016). We cluster at the district level and we use both the treatment and VCG clusters instead of clustering on treated schools, as was done previously.

Finally, we updated our analysis to make use of more up-to-date norms published in Thum and Hauser (2015).

Number of Schools and Students in Achievement Analysis Samples

Table B.1 displays the number of schools and students entering into the overall analysis of mathematics and reading for the 32 NGLC schools in the 2014–15 analysis, and the 16 NGLC schools in the 2013–15 analysis. Students had to remain in one of the NGLC schools in our sample to be included in the analysis. The table indicates the students' grade level at the start of the relevant time span.

Table B.2 displays the number of schools and students entering into the comparison of charter-operated and district-operated NGLC schools for the 2014–15 academic year. In this analysis there were 24 charter schools covering all grade spans, and eight district schools covering the middle and high-school grades.

Table B.1. Number of schools and students in aggregate analyses

		Group	Number of Schools	Number of Students by Grade at Start of Time Span												
				K	1	2	3	4	5	6	7	8	9	10	11	12
2014–15	Reading	NGLC	32	253	251	159	160	89	259	1,133	955	529	1,149	454	82	1
		VCG	5,040	5,807	6,724	5,341	5,797	3,062	4,918	32,082	22,634	15,415	23,402	12,503	1,643	51
	Math	NGLC	32	250	255	153	158	84	258	1,235	928	528	1,159	410	77	44
		VCG	4,837	5,122	6,995	5,264	5,682	2,876	4,467	31,187	23,324	15,058	24,515	11,754	1,893	1,496
2013–15	Reading	NGLC	16	78	52	56	76	55	100	554	309	69	394	65	1	
		VCG	2,723	1,443	1,919	1,986	2,913	2,219	3,488	11,982	7,004	1,055	9,753	1,040	51	
	Math	NGLC	16	91	43	52	70	52	102	555	304	111	395	66	36	
		VCG	2,745	2,304	1,843	1,634	2,680	1,802	3,217	11,268	6,859	1,376	9,822	1,593	1,172	

Table B.2. Number of NGLC schools and students in the 2014–15 charter-district analysis

	Group	Number of Schools	Number of Students by Grade											
			K	1	2	3	4	5	6	7	8	9	10	11	12
Reading	Charter	24	253	251	159	160	89	259	886	612	232	621	85	75	
	District	8							247	343	297	528	369	7	1
Math	Charter	24	250	255	153	158	84	258	873	603	235	565	86	68	38
	District	8							362	325	293	594	324	9	6

Assessment of Balance Between the Treatment Group and the VCG

The VCG is intended to be very similar to the study group in terms of students' observable characteristics prior to treatment. This is true by construction for the criteria that were matched exactly (namely, the grade level of the student and the urbanicity of their school). For the approximate matching criteria, we examined whether the groups appear to be the same. Table B.3 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 VCG means, wherein we weight the 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). We find standardized differences that are always smaller than 0.25 in the unweighted comparison, and particularly close for baseline scores. Restricting the sample to one observation per VCG student caused some of this minor imbalance relative to the data initially received from NWEA. The standardized differences are even smaller after weighting, reflecting the role that the weights play in helping to restore balance.

Sensitivity Analyses

To help evaluate the robustness of the main findings discussed above, we performed a variety of sensitivity analyses. The extent to which these alternative analyses produced similar or different estimates than our main analyses could help validate the treatment estimates or place likely bounds on true treatment effects.

Analyses Based on Conditional Growth Norms

First, we used an alternative method for estimating treatment effects using conditional expected growth estimates based on norms (CGN) calculated by NWEA. CGN uses students' starting scores and elapsed time to predict a typical posttest score based on normative data from a national sample (for more on the CGN methodology, see Thum and Hauser, 2015, p. 38). The CGN method does not consider other factors that are part of the VCG matching, such as student gender, schoolwide measures of poverty (e.g., FRL), and geographic locale. For each relevant subgroup (school, grade span, or overall), we estimated the average difference between the treated students' realized growth and their CGNs, under the assumption that national norms generally represent typical growth in schools that are not NGLC schools.

Restriction to the Same School Type

In a second sensitivity analysis, we set additional constraints for the VCG matching. Many of the NGLC schools are charter schools, which may tend to enroll a select group of students. As one example, families make an affirmative decision to enroll their children. Family involvement in education might influence student

Table B.3. Balance between NGLC and VCG groups on variables not exactly matched

			NGLC	Unweighted VCG			Weighted VCG		
	Subject	Variable	Mean	Mean	Difference	Std. Diff.	Mean	Difference	Std. Diff.
2014–15	Reading	Start RIT	204.77	202.93	1.84	0.07	204.71	0.07	0.00
		FRL	74.94	74.86	0.09	0.00	74.38	0.56	0.03
		Elapsed days	243.56	238.03	5.54	0.20	241.40	2.17	0.08
	Math	Start RIT	211.63	210.13	1.50	0.05	211.58	0.05	0.00
		FRL	74.95	74.36	0.58	0.03	74.42	0.52	0.02
		Elapsed days	242.40	237.63	4.77	0.17	240.11	2.30	0.08
2013–15	Reading	Start RIT	203.81	201.38	2.43	0.09	203.77	0.04	0.00
		FRL	79.25	75.81	3.44	0.16	77.24	2.01	0.09
		Elapsed days	591.56	593.02	–1.46	–0.04	589.31	2.25	0.07
	Math	Start RIT	212.05	208.12	3.93	0.14	211.99	0.06	0.00
		FRL	78.84	75.57	3.27	0.15	76.66	2.17	0.10
		Elapsed days	592.09	594.48	–2.40	–0.07	589.89	2.20	0.06

Note: the unweighted VCG columns show sample characteristics after restricting to one observation per VCG student, per subject, per year.

achievement in positive ways unrelated to the schools' influence on achievement. To the extent VCGs are drawn from schools that are not charter schools, there is the potential that a difference in family involvement, or in other factors that might influence students to enroll in schools of choice, could bias the results. We investigated this concern by attempting to make the treatment and control groups more similar on such factors. We only kept VCG students that had the same school type (district or school of choice) as their treated NGLC student. We compared the treatment effect estimate using the matched school-type VCG with that from the standard VCG matching criteria that ignore choice. The concern about unmeasured differences between choice and nonchoice schools would gain credence if the schools-of-choice VCG produces meaningfully lower treatment effect estimates than the standard VCG analysis.

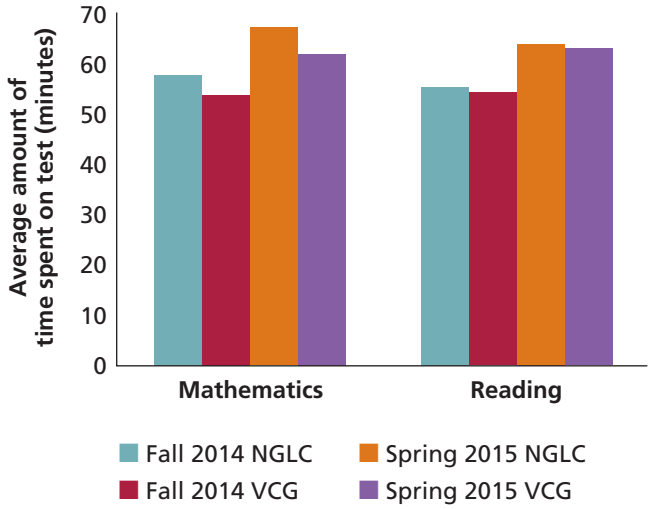
Filtering and Alternative Time Span Analyses

Finally, we discovered concerning patterns in test duration (the amount of time students spend taking the test) among students and schools in the study. Briefly, some student-test events had very long durations or large changes in test duration between the fall pretests and spring posttests. This raised concerns that differences in duration, or testing conditions that drive changes in duration, might influence estimates of the treatment effect of attending an NGLC school.

To that end, we performed a set of sensitivity analyses related to duration to gain a better understanding of how anomalies in test duration might be affecting treatment effect estimates. We applied filters to remove students with anomalous test durations or anomalous changes in test duration between pretest and posttest. We also applied filters at the school level based on aggregate patterns of test durations of the participating students. Finally, we examined the use of a different time span in spring-to-spring because this pretest–posttest pair tends to have less discrepancy in test duration. The difference in duration change is less dissimilar between NGLC and VCG students than was observed between the PL and VCG students in Pane et al. (2015). Figure B.1 shows that NGLC students generally spent more time on the tests than their VCG counterparts. However, in both subjects, the fall-to-spring time increases were about the same for NGLC and VCG students, at 16 percent.

Although the average durations presented in Figure B.1 do not suggest concerns about duration, we applied the same filtering methods as in Pane et al. (2015) to further

Figure B.1. Test durations for NGLC and VCG students



assess the risk that anomalous test duration growth for some students might influence the estimated treatment effects. First, we began by filtering out outlying test durations both among the treated students and the VCGs. We used the following filters:

- **Filter 1:** Drop if fall or spring test durations are below 5th percentile or above 95th percentile for grade and subject (national duration, provided in personal communication by NWEA).
- **Filter 2:** Drop if the change in test duration from fall to spring exceeds the national 90th percentile of change in test duration for grade and subject.
- **Filter 3:** Drop if the durations meet the criteria of both filters 1 and 2.

If an NGLC student met a filter's criteria, all of the VCG records for that student were also filtered out. However, if a VCG student was filtered we did not drop the corresponding NGLC student, or other VCG records that did not meet filter criteria. Table B.4 presents the percentages of NGLC and VCG student records that were filtered out. In every case, more VCG records are filtered out than NGLC records. Filter 3, by construction, filters out the smallest fraction of students.

Table B.4. Percentages of NGLC and VCG students dropped by filters

	Filter 1		Filter 2		Filter 3	
Subject	NGLC	VCG	NGLC	VCG	NGLC	VCG
Mathematics	30%	54%	21%	45%	14%	35%
Reading	32%	56%	22%	47%	14%	35%

Although some changes in test duration could reflect inappropriate test administration conditions, in some cases these changes might be due to factors that could legitimately be attributed to treatment effects, such as academic growth that results in more-difficult (and more time-consuming) items being administered in the spring, or increases in students' willingness to persist through challenging test content. Where this is the case, it would be incorrect to filter such students out, and the treatment effect would be biased if part of the treatment were increasing student human capital in ways that would appear to result in anomalous test duration or change in duration. To that end, we additionally evaluated the overall treatment effect, where instead of filtering individual students out, we only filtered out anomalous schools. We used two methods to filter out schools:

- calculate average durations by subject and grade for all students in the school and filter out the school if filter criteria are met
- filter out a school if over 40 percent of students in that school meet filter criteria.

Using these filtered data sets, we applied the same statistical models used previously to estimate treatment effects overall and for each school, for each subject, and time span.

As an alternative to filtering, we can use multi-year data to estimate treatment effects using time spans other than fall to spring. An alternative to the fall-2014-to-spring-2015 span is the spring-2014-to-spring-2015 span. The purpose for this is that the differences in test duration are generally between fall and spring, with fall durations typically shorter than spring durations. Therefore, using spring-to-spring time spans alleviates the issue. This also reflects a common span used in other educational achievement analysis, where only spring achievement tests are available.

However, there are potential problems with this alternative. First, the new span includes summer, and researchers have found evidence that students experience test-score declines over the summer. If summer declines are an outgrowth of differences in testing conditions and not related to actual learning, then including summer may result in a more accurate measure of learning during the school year because the pretest and posttest are administered under more-similar conditions. However, it may be that some of this summer loss is true loss of the achievement that accrued the prior school year, which should be attributed to the schools and their practices, in which case, time spans that include summer are more problematic. Moreover, if we believe that the fall or spring test durations are so short or so long as to result in invalid scores, these alternative durations may also suffer from the same problem.

An additional problem is that, if most of the treatment effect happens in the first year of exposure to the school or to NGLC, then this will be missed by not starting from a baseline fall score.[3]

Although these alternate time spans use two-year data to create additional estimates of one-year effects, they differ in important ways from estimates made from one-year data. In addition to the differences already noted, the data have differences both in the treatment students included (students need to have been present in the NGLC schools for both years and tested at least three times, as opposed to needing the students present just for the two tests in the same year for the one-year span) as well as having a potentially different set of VCGs. For these reasons, we considered the comparison of the different spans with each other, but did not directly compare them to the filtered treatment effect estimates. By using the 2013–15 span to get the needed data, we are restricted to using the 16 schools available.

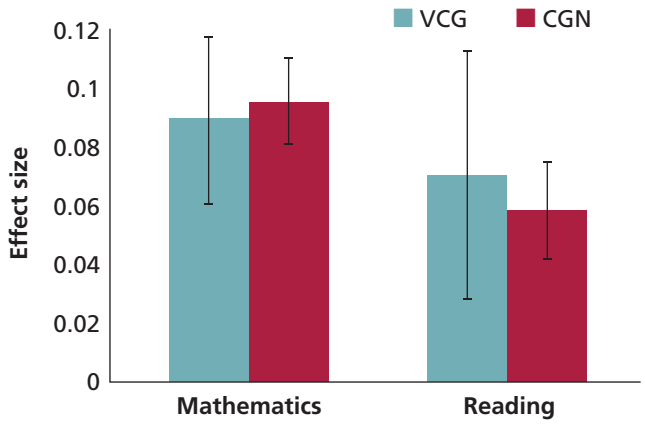
Results of Sensitivity Analyses

CONDITIONAL GROWTH NORMS

First, we estimated treatment effects using CGNs, shown in Figure B.2. To interpret these results, we focus on the fact that the CGN analysis estimates similar positive effects as the main analysis. We interpret this as helping to validate the main results. Although the CGN analyses vary slightly from the VCG method, we focus less on the magnitudes of the CGN estimates because the VCG method is more rigorous in carefully developing a matched comparison group, as opposed to benchmarking against national norms as is done in the CGN method. The VCG estimates in this chart differ from the main analysis because CGN estimates were not provided for some students (e.g., 12th graders), and those students were dropped from both analyses for the sake of using a consistent sample for this sensitivity test.

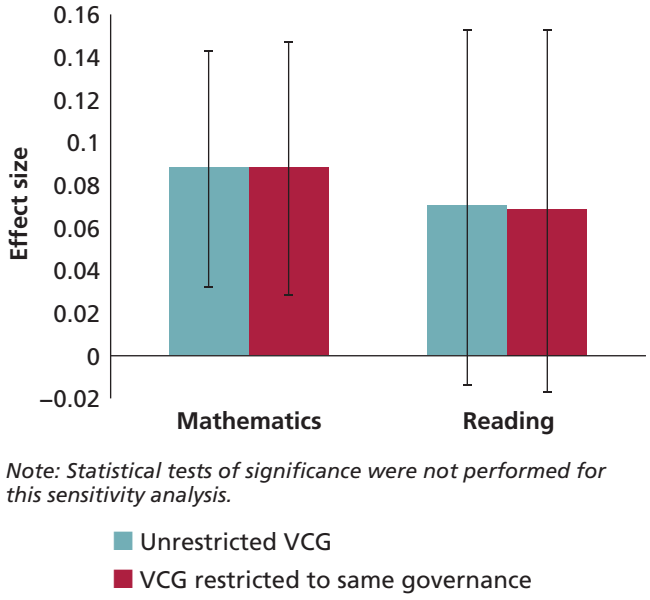
[3] Also, on a more technical note for our current data, when we use spring pretests, the students are not matched to their VCGs on this pseudo-baseline. To account for this, we also evaluate a treatment effect where we drop all VCGs not within three points on the RIT scale (approximately 95 percent of VCGs are within plus-or-minus three points of the PL student's score on the interim spring test, while an even higher proportion of VCGs are within plus-or-minus three for the true baselines on which they were matched).

Figure B.2. Comparison of VCG and CGN methods
2014–15



Note: Statistical tests of significance were not performed for this sensitivity analysis.

Figure B.3. Comparison of VCG and same-governance VCG analyses
2014–15



Note: Statistical tests of significance were not performed for this sensitivity analysis.

Unrestricted VCG
VCG restricted to same governance

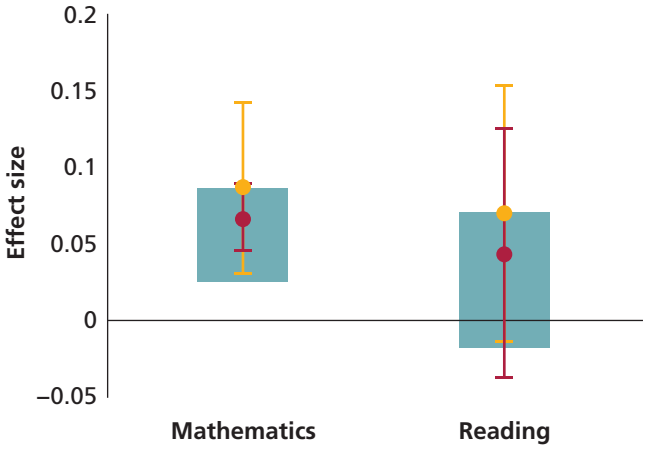
RESTRICTING THE COMPARISON TO SCHOOLS OF THE SAME GOVERNANCE

Next, we examined treatment effects using a VCG composed of students only from the same governance structure (district or charter) as the corresponding NGLC school. Figure B.3 presents these results. The results are virtually the same in both subjects. We conclude that these results help to affirm the treatment effects estimated by the standard VCGs.

DURATION ANALYSIS

We applied a variety of student-level and school-level filters to remove anomalous test durations from the analysis. Applying the filters at the student- and school levels yields a range of estimates. Figure B.4 focuses on the main analytic sample and displays the unfiltered estimate and confidence interval in yellow, and the median filtered estimate and its confidence interval in red. The blue bars show the range of the filtered estimates (but not their confidence intervals). In both subjects, the median filtered estimate is smaller than the unfiltered estimate. For mathematics, the median filtered estimate is positive and statistically significant, and none of the filtered estimates are negative. For reading, the median filtered estimate is positive but significantly indistinguishable from zero, with two of the nine filters producing negative estimates. The decrease in the treatment effect from unfiltered to the median filtered estimate is 23 percent for mathematics and 39 percent for reading.

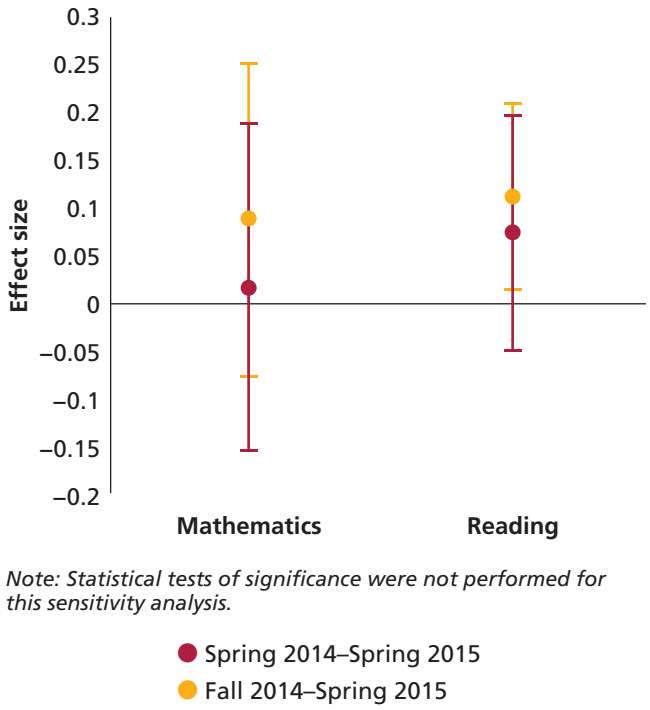
Figure B.4. Analyses with Test Duration Filters



Note: Statistical tests of significance were not performed for this sensitivity analysis.

Range of filtered estimates
Unfiltered estimate
Median filtered estimate

Figure B.5. Alternative time span comparison



Note: Statistical tests of significance were not performed for this sensitivity analysis.

Finally, we looked at alternative spans for the subset of NGLC schools that had been operating for at least two years. Figure B.5 presents these results. Results from the main fall-2014-to-spring-2015 analysis are shown in blue, with the alternative span of spring 2014 to spring 2015 shown in red. The spring-to-spring analysis produces smaller treatment estimates, particularly in mathematics. The fall-to-spring results differ from the estimates for the whole sample because only a subset of students had the requisite set of scores to participate in this sensitivity test.