Skip all front matter: Jump to Page 1

COMPETENCY-BASED EDUCATION IN THREE PILOT PROGRAMS

Examining Implementation and Outcomes

Jennifer L. Steele, Matthew W. Lewis, Lucrecia Santibañez, Susannah Faxon-Mills, Mollie Rudnick, Brian M. Stecher, Laura S. Hamilton

RAND CORPORATION

RAND EDUCATION

COMPETENCY-BASED EDUCATION IN THREE PILOT PROGRAMS

Examining Implementation and Outcomes

Jennifer L. Steele, Matthew W. Lewis, Lucrecia Santibañez, Susannah Faxon-Mills, Mollie Rudnick, Brian M. Stecher, Laura S. Hamilton

Prepared for the Bill & Melinda Gates Foundation

The research described in this report was sponsored by the Bill & Melinda Gates Foundation and was produced within RAND Education, a division of the RAND Corporation.

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-0-8330-8725-6

Cover images: top left photo, michaeljung/Shutterstock; bottom right photo, Goodluz/Shutterstock; bottom left chart courtesy EffectiveSC.org

RAND OFFICES

SANTA MONICA, CA • WASHINGTON, DC

PITTSBURGH, PA • NEW ORLEANS, LA • JACKSON, MS • BOSTON, MA

CAMBRIDGE, UK • BRUSSELS, BE

www.rand.org

Preface

In 2011, the Bill & Melinda Gates Foundation created the Project Mastery grant program to support competency-based education initiatives in large school systems that serve a high proportion of disadvantaged youth. Competency-based education is education that meets students where they are academically, provides students with opportunities for choice, and awards credit for evidence of learning, not for the time students spend studying a subject. The grants supported the development and implementation of technology-enabled curricula, online learning management systems, and teacher professional development during the 2011–2012 and 2012–2013 academic years. The three recipient organizations—which included two large school districts and one intermediary organization—carried out their pilot programs in a total of 12 public secondary schools distributed across five school districts in four states. The Foundation asked RAND to evaluate these efforts in terms of implementation, students' experiences, and student performance. This report presents final results from that evaluation. It provides an overview of competency-based education and the Project Mastery grant projects and describes the implementation of competency-based educational features under each project. It also reports on student survey data from each of the projects. Finally, it describes the academic performance of students exposed to the sites' competency-based models relative to similar students or schools. The report concludes with six lessons for policy, partnerships, and practice.

The results should be of interest to educational policymakers and practitioners interested in competency-based education models as an approach to K–12 education reform. This study was undertaken by RAND Education, a business unit of the RAND Corporation, on behalf of the Bill & Melinda Gates Foundation.

Contents

Figures

Tables

Summary

In 2011, the Bill & Melinda Gates Foundation extended grants to three educational organizations working to develop or enhance competency-based approaches to secondary-school education. Competency-based education is education that meets students where they are academically, provides students with opportunities for choice, and awards credit for evidence of learning, not for the time students spend studying a subject. The grant initiative, called Project Mastery, funded the development of technology-enhanced tools, including curriculum materials and online learning management systems, as well as their implementation in secondary school classrooms. The tools were designed to facilitate personalized educational experiences by addressing students' individual needs and interests, and by holding students accountable for showing that they had met their state's academic standards. The Project Mastery grantees included Adams County District 50 (Adams 50), a suburban school district in Colorado that in the first year of our study (2011–2012) served about 10,000 students in 19 schools; the Asia Society, a New York–based nonprofit organization that advocates global competence and partners with 34 schools across the country through its International Studies Schools Network (ISSN); and the School District of Philadelphia (Philadelphia), a large, urban district in Pennsylvania that in 2011–2012 served roughly 154,000 students in 266 schools.

Project Mastery Initiatives and Evaluation Design

Adams 50 carried out the Project Mastery grant in one high school and three middle schools, focusing on development and implementation of mathematics games for grades eight and nine, teacher-created instructional videos, and upgrades to its competency-focused learning management system. As an intermediary organization rather than a school district, Asia Society carried out the grant in partnership with four secondary schools across the country, including three ISSN schools of choice in Denver and Houston (Denver Center for International Studies, Denver Center for International Studies–Montbello, and Sharpstown International School) and one traditional high school in Newfound, New Hampshire (Newfound Regional High School). The grant effort focused on creation of performance outcomes and rubrics at the eighth- and tenth-grade levels, sample curriculum modules, and professional development modules for teachers, as well as the adoption of a new, online learning management system. Philadelphia carried out the grant in six high schools within the district, focusing on the development of a new ninth-grade English curriculum and adoption of a new learning management system, as well as an optional afterschool program focused on digital filmmaking. The Project Mastery grants designated the 2011–2012 school year (Year 1) for materials development and 2012–2013 (Year 2) for implemen-

tation, so the sites focused largely on creating the new materials in Year 1, and on rolling them out to classrooms in Year 2. However, because the development process was iterative in many respects, the activities of the two years overlapped to some extent.

Near the end of 2011, the Foundation asked the RAND Corporation to evaluate the Project Mastery grant initiatives, including an analysis of implementation, student experiences, and student performance. RAND commenced the study in early 2012 and visited the sites in the spring of 2012, the fall of 2012, and the spring of 2013. The number and timing of the visits differed by grantees; every school and grantee office was visited between one and three times. During the spring of the 2012–2013 school year, we also surveyed students in Project Mastery classrooms, obtaining a 65.3 percent overall response rate.

During primary data collection, we found that the Project Mastery sites varied considerably from one another in terms of the scope of the tools they were developing and the number of students who would experience the tools during the pilot period. This information had implications for our analysis of student outcomes under the grant. For the implementation analysis (which included site visits and student surveys), we focused on students and classrooms identified as targets of the Project Mastery interventions. For the student outcomes analysis, we adjusted the sample and analysis based on (1) whether the intervention was designed for a class, school, or district; (2) the extent of students' exposure to Project Mastery tools during the study period; and (3) whether the sample was also part of a similar district or state initiative.

Defining Features of Competency-Based Education

Based on a related but separate set of interviews with experts in the field of competency-based education (Lewis et al., 2013), our study sets forth three defining features of competency-based models: instruction meets students where they are, which results in flexible pacing; students have choices about their learning; and students are evaluated based on their demonstrations of proficiency in the academic standards for their education level and subject area. Using this framework, we identified three distinct typologies among the Project Mastery sites (treating each Asia Society school as its own site) in terms of their broader approaches to competency-based education. Adams 50 emphasized flexible pacing and evaluation for proficiency, with secondary focus on student choice of materials and products. Two of the Asia Society schools—Newfound Regional High School (Newfound) and the Denver Center for International Studies–Montbello (DCIS-M)—focused on student choice and evaluation for proficiency, with secondary focus on flexible pacing (see Table S.1). Meanwhile, two other Asia Society schools—the Denver Center for International Studies (DCIS) and Sharpstown International School (Sharpstown) in Houston—and the Philadelphia initiative focused primarily on student choice in the context of project-based learning, with secondary focus on both flexible pacing and evaluation for proficiency. This typology is based on how staff at the sites described their areas of emphasis and implementation, but it is important to note that all sites incorporated each of the three features to some extent.

Table S.1
Relationship of Pilot Designs to the Defining Features of Competency-Based Education

		Asia Society				
Feature	**Adams 50**	**Newfound**	**Sharpstown**	**DCIS**	**DCIS-M**	**Philadelphia**
Instruction meets students where they are (flexible pacing)	Key component (multi-age levels based on proficiency)	Some evidence (emphasis on revision)	Some evidence (graduation portfolios)	Some evidence (graduation portfolios)	Some evidence (graduation portfolios)	Some evidence (teacher-developed scaffolding)
Students have choices to personalize learning	Some evidence (choice of online resources)	Key component (project choice; Extended Learning Opportunities)	Key component (choice of project topics and execution)	Key component (choice of project topics and execution)	Key component (choice of project topics and execution)	Key component (choice of project topics and execution)
Students are evaluated on evidence of proficiency	Key component (Scantron-based; mastery grading)	Key component (rubric-based; mastery grading)	Some evidence (rubric-based)	Some evidence (rubric-based)	Key component (rubric-based; mastery grading)	Some evidence (rubric-based)

NOTE: Shading is provided to make patterns easier to see at a glance. Green shading represents a key component for that site; yellow shading represents some evidence that the feature was used at that site. This table appears as Table 3.2 in Chapter Three of the report.

Insights from Implementation

When staff were asked about implementation of competency-based programs, several tensions became clear. The first tension concerned how to provide credit for out-of-school learning in which students demonstrated proficiency in required standards. Teachers varied in their agreement that certain demonstrations of proficiency were equivalent to class assignments or to courses, and district policies varied in the extent to which travel-based and experiential learning experiences could supplant course-based credits. A related issue was whether nonteachers would be allowed to evaluate students' out-of-school work for credit. None of the sites offered this option due to reliability concerns in some cases and to student safety and privacy concerns in others. The second tension concerned the challenge of holding all students to a common definition of proficiency. Some teachers reported struggling with the idea that students should be evaluated strictly on what they learned and not on the effort and practice they put forth in learning. A third tension concerned how to ensure sustainability of the models. One strand of this tension focused on the technical challenges facing each of the learning management systems in use—all were in beta or preliminary versions that did not interact well with extant student data infrastructure. Another strand focused on financial and logistical sustainability. Some sites lacked the computer hardware to give students access to online instructional tools. Developers partnering with other sites placed limits on use of the new curriculum that made scaling up cost-prohibitive. The final key tension concerned equity challenges in implementing competency-based education. In some sites—especially those that were emphasizing flexible pacing and evaluation based on competency—some teachers reported disengagement among students who struggled academically. Echoing a sizable body of evidence that more-challenging graduation requirements may yield lower rates of high school completion (Dee &

Jacob, 2006; Holme, Richards, Jimerson, & Cohen, 2010; Papay, Murnane, & Willett, 2008, 2011; Warren, Jenkins, & Kulick, 2006), this finding suggests the importance of instruction that inculcates persistence and engagement (Holme et al., 2010) and considers these to be attributes of competence (Heckman & Rubinstein, 2001). It also suggests the importance of monitoring student performance gaps under competency-based systems and directing additional academic support to students who appear to struggle under these systems.

Student Experiences and Performance

Our surveys showed that despite considerable heterogeneity in the features of each program, students' reported experiences across the sites were more uniform than we anticipated. This relative uniformity pertained to their belief in the value of school, their reported engagement in school, and in their experiences of flexible pacing and academic choice. The one site that stood out as having the highest reports of student engagement, flexible pacing, and choice was DCIS, in which students were engaged in yearlong, self-directed projects that applied academic content to real-world contexts. We also found modest racial/ethnic differences in students' perceptions of the Project Mastery instructional tools and approaches. This finding suggests the importance of monitoring differential student responses to competency-based approaches.

To examine student achievement, we applied diverse analytic methods due to differences in the available data from site to site. Given extant design constraints, these methods also varied in their ability to support causal interpretations. We found that student performance varied considerably among the sites. For Adams 50, we found that the conversion to a competency-based model was associated with lower mathematics performance on the state accountability test in the five years since the transition (by about 0.22 of a student standard deviation in 2013)—a difference that was significant at the 10-percent level. For Newfound, we found that conversion to a competency-based model was associated with increased performance of about 0.1 of a student standard deviation in reading. This difference was not statistically significant, although the analysis had low power that made it very difficult to distinguish effects of this size from zero. Among the other Asia Society sites, we found that DCIS and Sharpstown students markedly outperformed demographically similar peers in their respective states, though both were schools of choice, which made it difficult to attribute these differences strictly to the competency-based model. In the other school of choice in the Asia Society sample, DCIS-M, students performed similarly to their demographically comparable peers. In Philadelphia, Project Mastery students performed below their peers on the fall language arts assessment (after three months of exposure) and about the same as their peers on the winter assessment (after five months of exposure) and in rates of promotion to tenth grade the following year. Yet for the school year, their attendance rates were 1.4 percentage points higher than that of their peers—a difference that was significant at the 5-percent level. For Project Mastery students who regularly took part in an optional, after-school filmmaking badge program, both attendance rates and promotion rates were higher than those of non–Project Mastery students by statistically significant margins of 3.0 and 6.6 percentage points, respectively.

We also examined associations between programs' competency-education typologies and student performance under competency-based education in each site. For this analysis, we found that effects appeared more positive in the programs that made student choice their primary emphasis (the Asia Society schools of DCIS and Sharpstown, and Philadelphia), though it was mixed in the latter case, and though our research design was weakest in the former two cases. The site that emphasized flexible pacing and evaluation for proficiency (Adams 50) showed the most negative results, though the performance was possibly consistent with an implementation dip in that the district appeared to be improving relative to average state performance after an initial drop. The sites that emphasized student choice and evaluation for proficiency (the Asia Society schools of Newfound and DCIS-M) showed mixed results, none of which were statistically significant. The challenge in terms of extrapolating from these findings is that we cannot disentangle the effects of our research designs (which were least robust for the schools with no precompetency data) from the distinctive approaches to competency-based education. In other words, we can describe the patterns observed, but our observational research design does not permit us to conclude that differences in the competency-based approaches at each site were entirely responsible for differences in student outcomes in each site.

Lessons for Policy, Partnerships, and Practice

Our report concludes with six lessons for policy, partnerships, and practice; the first two lessons concern policy. First, competency-based education programs ideally should be assessed on a variety of near-term and longer-term outcomes. These include nearer-term measures, such as the test score outcomes examined in this report, as well as high school attendance and persistence rates where available. They also include longer-term measures, such as students' high school completion rates and their college enrollment, remediation, and persistence rates. The reasons are that student performance after a new reform effort may decline before it increases, and that better preparation for college and careers is an important goal of competency-based education. Second, policymakers may wish to allow for flexible timing of accountability tests rather than testing all students at the same time each year. In this way, students would be tested for proficiency only after they had been exposed to the tested content and had a chance to master it.

The next two lessons pertain to partnerships. We suggest that partnerships between districts and technology developers include licensing agreements that let districts affordably scale the technology innovations they help to develop and pilot. In addition, we suggest that collaborations between funders and districts take into account the existing infrastructure of the districts. This would help ensure, for instance, that adequate hardware is available for software use.

The final two lessons concern educational practice. The first is that increasing student autonomy calls for skillful teaching that generates student engagement. Rather than reducing the importance of quality classroom instruction, the student autonomy demands of competency-based education make creative and supportive instruction especially important. The final lesson is that competency-based systems must attend to equity concerns. Because competency-based education emphasizes academic skill over effort and increases the importance of student self-direction, systems adopting these approaches must be alert for widening achievement gaps and provide extra support to students who appear to fall behind.

Acknowledgments

We are indebted to many people without whose help this evaluation would not have been possible. Most prominently, these include the Project Mastery coordinators, school principals, instructional coaches, and teachers at each project site who agreed to meet with our research team, in some cases repeatedly and at length, during the course of the study, and who generously hosted our visits. We are particularly indebted to Kevin Byers, Jenny Gotto, Grace Cannon, Fran Newberg, Tonya Wolford, Adrienne Reitano, Gregory Benjamin, Alexis Menten, Jeremy Opitz, James LeBaron, Heidi Hirsch, Jennifer Portillo, Dan Lutz, Marjorie Larner, Jade Adams, Thuy Le-Thai, and Meredith Wedin. In addition, we benefited greatly from insights provided by the Bill & Melinda Gates Foundation's consultants to the Project Mastery sites, JoEllen Lynch and Melissa Giraud.

We are particularly indebted to the Bill & Melinda Gates Foundation for supporting this evaluation, especially Senior Program Officers Adam Tucker and Robert Torres and Senior Research Officer Ed Dieterle. Each provided invaluable suggestions and feedback throughout the project.

This report has been strengthened by an internal quality assurance review from Cathy Stasz at RAND and by a formal peer review by Erika Stump at the University of Southern Maine, Rebecca Herman at RAND, and John Pane at RAND, as well as by feedback from representatives of the project sites. The project has also benefited from administrative support by Donna White, Jamie Greenberg, and Patricia Clevenger. Finally, we are grateful for the editorial support provided by Arwen Bicknell and Steve Oshiro in RAND's publications department.

This document would not have been possible without the pathbreaking work of educators in each of the Project Mastery sites. Our hope is that through this report, their hard work will benefit not only their own students and communities but also other educators seeking new ways to meet students' individual learning needs in our nation's public schools.

Abbreviations

Adams 50	Adams County School District 50 (Westminster, Colo.)
ATT	average effect of treatment on the treated
CCSS	Common Core State Standards
DCIS	Denver Center for International Studies (Denver, Colo.)*
DCIS-M	Denver Center for International Studies–Montbello (Denver, Colo.)*
ELA	English language arts
ELO	Extended Learning Opportunities
Foundation	Bill & Melinda Gates Foundation
GPS	Graduation Performance System (Asia Society)
ISSN	International Studies Schools Network (Asia Society)
K–12	kindergarten through 12th grade
LMS	learning management system
Newfound	Newfound Regional High School (Bristol, N.H.)*
PMI	Progressive Mathematics Initiative
Philadelphia	The School District of Philadelphia (Philadelphia, Pa.)
PhilWP	Philadelphia Writing Project
RISC	Re-Inventing Schools Coalition
SCG	synthetic comparison group
SD	standard deviation
Sharpstown	Sharpstown International School (Houston, Texas)*

*Indicates an Asia Society partner site

CHAPTER ONE

Introduction

The Resurgence of Competency-Based Education

The movement toward proficiency-based or competency-based education is gaining momentum in both kindergarten through 12th grade (K–12) and postsecondary education settings, spurred in part by advances in digital learning technologies (Priest, Rudenstine, Weisstein, and Gerwin, 2012; Soares, 2012). These approaches allow students to progress at their own pace through a sequence of personalized learning experiences. In a competency-based system, students receive credit not as a function of how much time they spend studying a subject but based on demonstrations and assessments of their learning. Instruction is tailored to students' current level of knowledge and skills, and students are not constrained to progress at the same rate as their peers. The approach allows for accelerated learning among students who master academic material quickly, and it provides additional time and—ideally—additional support for students who need it. The theory is that a larger share of students will ultimately reach proficiency in a given content area if they are given the freedom to advance at their own pace and if their learning experiences are tailored to their needs and interests (Bloom, 1976; Lewis et al., 2013; Priest et al., 2012; Sturgis & Patrick, 2010).

Spurred by Benjamin Bloom's research in the 1970s on one-to-one learning models, the movement toward competency-based or "mastery learning" generated promising research well into the 1980s before tapering off in the 1990s. The movement shares commonalities with the standards-based accountability movement of the 1990s that gave rise to No Child Left Behind, in that both emphasized assessing students for their mastery of well-defined academic standards. Yet mastery learning's focus on student-centered instruction and pacing was likely a poor fit for standards-based accountability's emphasis on school accountability through annual standardized testing. Whatever the reason, the notion of competency- or mastery-based education fell onto the periphery of education reform for about two decades.

In recent years, however, competency-based education models have begun to garner new attention (e.g., Hanford & Smith, 2013), and policymakers have been responsive to the growth of the competency-based movement. According to a March 2012 report, 36 states had authorized waiver mechanisms or other alternatives to their seat-time requirements for high school graduation (Grossman & Shipton, 2012), and New Hampshire has maintained a competency-based credit system since 2009 (New Hampshire Department of Education, 2012). In a separate analysis, the Carnegie Foundation for the Advancement of Teaching (2013) found that 29 states allowed districts to choose competency-based crediting if they wished, while another ten states had policies that granted seat-time waivers under certain circumstances. It, too, noted that only New Hampshire had a policy requiring that districts award credits based on competency rather than seat time.

Despite taking on considerable momentum in the field, competency-based systems have not been extensively researched. Recent studies have described the experiences of educators working to undertake competency-based reforms or have highlighted promising models, but these studies have not systematically examined the effects of these models on student learning or persistence (Priest et al., 2012; Silvernail, Stump, Duina, & Gunn, 2013; Silvernail, Stump, McCafferty, & Hawes, 2014; Sturgis & Patrick, 2010). Nevertheless, a few promising models have emerged. In 2001, the Chugach School District, which served 214 students in a remote part of Alaska, won a Malcolm Baldrige National Quality Award for organizational improvement. After adopting a competency-based approach, the district reportedly raised students' reading, English language arts (ELA), and math scores by an average of 41 percentile points over a four-year period (National Institute of Standards and Technology, 2002). In 2009, the Lindsay Unified School District, a small northern California district of about 4,000 students—roughly half of whom are English language learners and three-quarters of whom are economically disadvantaged—began implementing a competency-based model. By 2012, the district was showing marked declines in gang membership and safety-related incidents (Lindsay Unified School District, 2014). In fact, competency-based models have been embraced by a number of schools nationally. They have proven especially popular in the New England region, where many of the practitioners are either rural schools or alternative schools, like the Boston Day and Evening Academy and the Diploma Plus program, which serve "overage and undercredited" students. These are students who are older than their grade-level peers and lack enough credits to graduate on time (Priest et al., 2012; Silvernail & Stump, 2012). The popularity of competency-based education in very small and alternative school settings seems natural, since the heterogeneity of student needs and skill levels may call for a differentiated, flexible approach. The fact that these models seem to flourish in rural and alternative environments nevertheless raises the question of whether they are equally feasible in populous urban districts, where student performance and high school graduation rates are especially low (Swanson, 2009).

A few recent exemplars suggest that the answer may be yes. Since 2007, a competency-based Yuma, Ariz., charter school called Carpe Diem has demonstrated faster outcome improvement than the state as a whole, especially in mathematics, despite enrolling a socioeconomically disadvantaged population (NBC News & the Hechinger Report, 2013). Similarly, the Young Women's Leadership Charter School of Chicago, which also serves disadvantaged students and employs a competency-based model, has posted graduation rates that are 35 percentage points higher than the Chicago Public Schools as a whole (Chicago Public Schools, 2013; Young Women's Leadership Charter School of Chicago, 2013). In each of these cases, there are no studies that would allow us to attribute outperformance to the competency-based education systems alone, but the data suggest that competency-based models may hold promise even in more urbanized settings.

The Project Mastery Initiative

The current report adds to the research base on competency-based education by describing implementation and outcomes from three competency-based education pilot programs carried out in five school districts during the 2011–2012 and 2012–2013 academic years. The school districts were geographically dispersed, and their approaches to competency-based education were highly diverse, though they shared a few important commonalities. First, they generally served urban and disadvantaged populations. Specifically, four of the five districts were large urban or suburban districts that served a large proportion of low-income or minority students. Second, the programs shared an emphasis on leveraging technology to facilitate competency-based learning. All of the pilot initiatives involved the development of technology-enriched curricula, which means curriculum materials that incorporate computers or online resources, *and* the development or expansion of online learning management systems (LMSs), which are online tools for facilitating student learning or for managing student data. These LMSs were intended to deliver curriculum to students in and outside of school settings, or to track students' progress across teachers, courses, and school years.

The commonalities among the pilots were intentional, since all were funded as part of the Project Mastery grant initiative undertaken by the Bill & Melinda Gates Foundation (the Foundation). This initiative was designed to support technology-rich, competency-based learning models in large school systems that served many disadvantaged students. The Foundation aimed to support programs whose deployment of technology offered strong potential for scalability in urban and disadvantaged schools. The Project Mastery grants were given to the following awardees:

- *Adams County District 50* (Adams 50), a suburban district of about 10,100 students in Westminster, Colorado, which had embarked upon a district-wide competency-based reform during the 2008–2009 academic year. As part of its Project Mastery plan, Adams 50 developed and piloted four online mathematics games, undertook a blended-learning pilot in which teachers created their own videos and used other online instructional content, and upgraded the district's competency-based LMS, called Educate.
- *Asia Society,* an intermediary organization that provides its partner schools with a set of learning outcomes and rubrics focused on global competence, along with professional development for teachers on how to develop and implement performance assessments. Under its Project Mastery grant, Asia Society developed new performance outcomes and rubrics, new curriculum modules, and new professional development modules. It also hosted a new, online learning management system called ShowEvidence, which was created by a technology development company of the same name. Finally, it worked with four of its partner schools to pilot these tools. The Asia Society partner schools in the grant were the Denver Center for International Studies (DCIS) and the Denver Center for International Studies–Montbello (DCIS-M) in Denver, Colorado; Newfound Regional High School in Bristol, N.H. (Newfound); and Sharpstown International School (Sharpstown) in Houston, Texas.
- *School District of Philadelphia* (Philadelphia), a large, urban school district in Philadelphia, Pennsylvania, serving more than 150,000 students. Under Project Mastery, Philadelphia purchased and piloted a technology-enriched, project-based, ninth-grade English curriculum and a new, online LMS. The district also collaborated with other local organizations to create a voluntary afterschool program for students involved in the ninth-grade language arts pilot in their English classes. Students in the afterschool program earned "badges"—that is, certifications of competency for skills in digital filmmaking—some of which could be applied toward credit in their high school language arts classes.

The Project Mastery grant programs began in the 2011–2012 school year (Year 1), which was designated primarily for development and planning. The grants continued through the 2012–2013 school year (Year 2), which was devoted largely to the implementation and roll-out of materials developed in Year 1. The pilot grants concluded at the end of the 2012–2013 school year. In follow-up conversations in 2014, the sites reported that they were continuing to use and build upon the materials they developed during the pilot years.

Purpose and Organization of this Report

In February 2012, the Foundation commissioned RAND to collect data on Project Mastery implementation and to provide a descriptive (noncausal) examination of student outcomes following exposure to the interventions. Despite their commonalities, the funded interventions differed considerably from one another in terms of their features, implementation scope, and even their implementation timelines. Our analysis of student outcomes, therefore, varied by program to accommodate programmatic variability. We collected implementation data on the programs between the spring of 2012 and the spring of 2013, so our primary data collection spanned the final one-and-a-half years of the pilot period. We collected data about student outcomes through the fall of 2013.

This report summarizes findings about both implementation and outcomes from RAND's study of the pilot programs. It is intended to provide lessons for the Foundation and the field about implementation of technology-rich, competency-based approaches in a set of mostly diverse, urban, public schools. It is also intended to illuminate for funders, educators, and policymakers what we do and do not know about the promise of competency-based education models. Because it is one of the first studies we are aware of since the late 1980s that has attempted to estimate the impact of competency-based models on students' academic outcomes, it may be of interest to educators and policymakers at the local, state, and federal levels who are seeking empirical analyses of student performance in competency-based systems. However, it is important for readers to bear in mind two key caveats. First, the models described in this report are idiosyncratic. Each embodies particular attributes of competency-based models, but no single program should be viewed as an archetype of competency-based education. Second, none of the students in this study were randomly assigned to competency-based education, nor were any of the educators or schools that implemented competency-based models assigned at random to those programs. That means that our estimates of the impact of competency-based education are likely to be confounded by *student-level selection* into these programs (especially in schools of choice, which include most of the Asia Society schools in the pilot), by *teacher selection* into these programs (especially when teachers were allowed to volunteer to be part of the pilot, as in Philadelphia), and by *school or district selection* into competency-based models (as in Newfound and Adams 50).[1] We attempt to disentangle the effects of competency-based models from these confounds as much as possible by controlling for observable differences between treated and comparison units in each analysis. We do this using not only regression adjustments, but also a synthetic comparison group design for the Newfound and Adams 50

[1] Asia Society schools' focus on global competence may be a more salient feature to parents choosing these schools than the focus on competency-based or project-based learning models. Methodologically, the concern is that parents who choose these schools may differ from those who do not in unmeasured ways that are related to their children's performance.

analysis and a propensity-score weighting in the Philadelphia analysis (see Chapter Six for a detailed description of these methods). Yet even with these adjustments, we cannot fully control for all possible, relevant confounds. Therefore, our estimates provide an indication of how students perform in each of these grant programs relative to similar students who are not in the programs, but we cannot fully attribute all of these performance differences to the competency-based models themselves. It is possible that the students, teachers, schools, or districts choosing these models are distinctive in other ways (for example, motivation, effort, etc.) that affect the outcomes in question.

Our report is organized as follows: The remainder of this chapter discusses the rationale behind the competency-based education movement and the evidence that competency-based approaches raise student achievement. Chapter Two summarizes the Project Mastery interventions in greater detail and frames our research questions and methods. The third chapter describes the implementation of the Project Mastery interventions in terms of the materials the sites developed or purchased and what they were able to implement during the pilot period. Chapter Four describes the implementation process, including the advantages and tensions that educators reported in each site. The fifth chapter summarizes students' experiences in the pilot as reported on surveys. Chapter Six describes our approach to analyzing student performance outcomes in each site and presents these results for each site. Chapter Seven provides a conclusion, including considerations for policy and practice.

Defining Features of Competency-Based Education

In 2013, RAND prepared a report for the Bill & Melinda Gates Foundation that provided a framework for defining and studying competency-based education in K–12 settings (Lewis et al., 2013). The report was based on informational interviews with more than 20 educators, competency-based education advocates, and federal, state, and local policymakers regarding the key features of competency-based systems. It was also informed by interviews with educators at the Project Mastery sites and by a literature review of recent developments in the competency-based education movement. Based on those sources, the report articulated three defining features of competency-based education, which it described as follows:

Instruction meets students where they are (flexible pacing). Competency-based education takes as its starting point students' current level of competency in learning progressions (sets of accepted standards that define competency), not their placement in an age-based cohort.

Students have choices to personalize learning. Competency-based education takes into account the fact that children learn differently, have different skill levels, and encounter different opportunities to learn outside of school. As students mature, competency-based education provides opportunities to make more choices regarding how to acquire skills and knowledge, as well as how to provide evidence of proficiencies. This personalization can lead to increased engagement.

Students demonstrate proficiency—and earn credit—by applying knowledge and skills. In a competency-based system, academic credit is awarded based on evidence that students are proficient in specified academic standards, and not based on the time they have spent studying content or the effort they have exerted. Such standards may reflect age-appropriate developmental trajectories (as in the Common Core State Standards [CCSS], which are specified by grade level and subject area), but students may not acquire proficiency at the same rates of speed.

In our descriptions of the sites' Project Mastery initiatives in Chapter Three of this report, we refer back to these defining features in order to discuss the central—and diverse—ways in which the sites operationalized the notion of competency-based education. It is nevertheless important to bear in mind that neither the Foundation nor the evaluation team expected a given site to demonstrate all three of these features. This is because their pilots were diverse by design, and also because the pilots were not defined with these particular criteria in mind. Rather, as part of Project Mastery, the sites were expected to address three tenets of competency-based education: *defined progressions toward mastery* (meaning that learning progressions are clearly articulated), *anytime/anywhere learning* (meaning access to out-of-school learning opportunities), and *credit for mastery* (meaning that academic credit is based on evidence of learning and not only completion of a certain number of days or hours of instruction in a course). While not identical, these tenets are closely aligned with the three defining features we describe above. Specifically, we interpreted defined progressions toward mastery and credit for mastery as two components of *students demonstrating competency by applying knowledge and skills.* In addition, we determined anytime/anywhere learning opportunities as one facet of *student choice*—that is, of where and how students learn. Though not explicitly stated, the notion of *flexible pacing* was implicit in the Project Mastery efforts even for sites that featured a mostly synchronous (i.e., group-paced) learning model. Therefore, the current definition simply helps to solidify the importance of this component to the notion of competency-based education in general.

Rationale for Competency-Based Education

The competency-based education movement has roots in several strands of progressive education. The emerging literature distinguishes "competencies" from content-specific standards by noting that competencies are broadly applicable skills that students should master before completing high school, whereas content standards are more granular and narrow (Sturgis, 2012). For example, the New Hampshire College and Career Readiness Competencies in ELA specify that graduating students will "demonstrate the ability to comprehend, analyze, and critique a variety of increasingly complex print and nonprint informational texts" (New Hampshire Department of Education, 2013). This is the second of nine competencies that students are expected to master before completing high school. The CCSS in reading that the state links to this particular competency goal include reading for key ideas and details, reading for craft and structure, and reading for integration of knowledge and ideas. The state also links this competency to CCSS standards for knowledge of language, vocabulary acquisition and use, research to build and present knowledge, and comprehension and collaboration (New Hampshire Department of Education, 2013).

This emphasis on teaching content within the context of broader analytic strategies is reminiscent of the five "habits of mind" espoused by the Coalition of Essential Schools (Meier, 1995, 2002; Sizer, 2004), and the literature defining competencies acknowledges this heritage (Sturgis, 2012).[2] Similarly, the movement's emphasis on graduation portfolios, rubric-based feedback on student work, and "authentic" performance assessments (in which students apply

---

2 The five habits involve considering the *significance* of a problem or issue, the *perspective* from which it is being viewed, the *evidence* that exists about the issue, the *connection* or applicability of the issue, and the *supposition* that it might be otherwise (Meier, 1995, 2002; Sizer, 2004).

academic skills to real-world problems) has roots in the progressive tradition, which emphasizes a student-centered approach to education that is driven by experience and discovery rather than lecture and memorization (Meier, 1995). This tradition is largely rooted in John Dewey's (1902) call to allow students to construct their own learning inductively, through real-world engagement, in lieu of a deductive approach that emphasizes the application of algorithms and facts. Perhaps the key thought leader in the competency-based movement, however, was the educational psychologist Benjamin Bloom, who introduced the concept of "mastery learning." Based on his research with students in Chicago, Bloom argued that tailoring instruction to the skills and learning rates of individual students would eventually reduce performance gaps between students of lower and higher natural aptitude, and would allow virtually all students to master the academic material taught in K–12 settings (Bloom, 1976). He later worked to corroborate this theory with a study in which he showed that individualized, one-on-one tutoring produced learning gains that were two standard deviations above those produced by whole-group learning (Bloom, 1984)—a difference that dwarfs effect sizes in much of the education literature as a whole (Hill, Bloom, Black, & Lipsey, 2007).

Like many education reform initiatives, the competency-based movement seeks to reduce inequities in students' outcomes and shrink achievement gaps. Its underlying premise is that students fall behind in a traditional system because they are forced to move on to new concepts or skills before they have mastered the current skills, resulting in weak academic foundations and in knowledge gaps that are never filled (Priest et al., 2012). Advocates argue that if students are allowed to advance at their own pace and are exposed to appropriate learning experiences or tasks, a larger share of students will master core content rather than moving forward on shaky foundations (Bloom, 1976; Priest et al., 2012). This, in turn, will result in a larger share of students meeting academic content standards, even if not all students achieve those objectives at the same points in time. In the parlance of the movement, it is time, rather than learning, that becomes the variable in students' educational experiences (Priest et al., 2012; Slavin, 1984). The movement's emphasis on portfolio assessments and rubric-based evaluation measures also follows this logic, since evaluation against well-designed rubrics provides students not only with an indication of how well they performed on a task, but also with qualitative indications of what they would need to do differently to master it (Sturgis, 2012). In other words, rather than just indicating how well students have learned a particular skill, assessments in a competency-based framework are intended to offer students a clear pathway to improvement—indicating where they are on the correct path, and what they must do to achieve mastery. Even more critically, students working in a competency-based system are given time to redo assignments and revisit material until they have mastered it. In other words, they do not move on until they have met the academic standards or subsidiary standards in question (Priest et al., 2012).

With its emphasis on demonstrated mastery of well-defined content objectives, the competency-based movement builds on the success of the standards-based education movement of the 1980s and 1990s. The standards movement led many states to adopt well-specified standards in core content areas (e.g., mathematics, reading, writing, science, and social studies) for what students should know and be able to do at each grade level, as well as developing tests designed to measure students' mastery of those knowledge and skills. (For an overview of the standards movement, see Hamilton, Stecher, & Yuan, 2009.) However, proponents of competency-based models have shown a philosophical preference for performance-based and portfolio-based assessments over standardized tests. They portray the latter as sacrificing cognitive complexity and real-world authenticity for reliability and low-cost efficiency (Priest et al., 2012; Sturgis, 2012).

Though many of the ideas underpinning the competency-based education movement are not new, the movement has been catalyzed in recent years by several forces. First, since the federal No Child Left Behind Act took effect in 2002, schools in all 50 states have faced accountability pressure to ensure that all students attain proficiency in core academic content and that low-income and minority students are not lagging academically behind their peers. As schools have struggled to close longstanding achievement gaps, the widespread use of "social promotion"—that is, promotion of students to the next grade who have not yet attained proficiency in grade-level content—has been identified as one source of the problem (e.g., Jacob & Lefgren, 2004; McCombs, Kirby, & Mariano, 2009). At the same time, the body of literature on social promotion suggests that simply putting students through the same grade twice does not necessarily help them catch up. Instead, retention appears to work when it is accompanied by additional support and intervention (McCombs et al., 2009; Nagaoka & Roderick, 2004). Competency-based education builds on this idea; its logic dictates that if students are not permitted to move on without mastering core material, they will not accrue the gaps in learning that undermine their ability to succeed at each new content level. Moreover, a competency-based education system is intended to be more flexible than a traditional system that simply decides whether students should progress each year. Rather, the vision is that each student progresses through content at his own pace, and potentially at different paces in different content areas, so there is no dichotomous end-of-year decision about whether a student is promoted. Ideally, the student is always receiving instruction that matches his level of skill, and new skills are introduced only after he has demonstrated proficiency in the underlying ones.

Also catalyzing the competency-based movement are new advances in educational technology that have made it easier for students to progress academically at their own pace (Pennsylvania Department of Education, 2013). For example, the online mathematics video repository Khan Academy has won accolades for providing high-quality mathematics instruction videos that span a wide array of topics and difficulty levels. Teachers in some schools are using these videos to "flip" their classrooms so that students watch lectures on their own time, tailored to their own level of need, and spend their class time getting help practicing the methods they are learning online. The effects of this "flipped classroom" approach, though not yet clear, are currently being studied in the Los Altos Unified School District in California (Edutopia, 2011). There is also some evidence that instructional systems that incorporate both online and teacher-led instruction can lead to higher student performance than teacher-led instruction alone. For example, several randomized trials of Scholastic's READ 180 curriculum, which combines both teacher-led and flexibly paced, computer-based instruction, have shown it to be more effective than comparison curricula at raising students' reading achievement (Loadman et al., 2011; What Works Clearinghouse, 2009). And a seven-state randomized trial of Cognitive Tutor Algebra I, a curriculum that supplements teacher-led instruction with a flexibly paced, computer-based tutoring system, found that students exposed to the curriculum outperformed those exposed to traditional, textbook-based curricula by about eight percentile points, though the benefits did not emerge until the second year of implementation (Pane, Griffin, McCaffrey, & Karam, 2014). Such advances may have helped elevate the visibility of the competency-based education movement, raising the question of why students in a classroom should still be expected to advance at the same pace, when skilled deployment of technology can yield highly tailored learning experiences.

Extant Evidence Supporting the Approach

Within K–12 education, one context in which competency-based education has a well-established history is among alternative schools and credit-recovery programs (Pennsylvania Department of Education, 2013). These programs often serve students who are "overage," meaning they are older than their same-grade peers, and "undercredited," meaning that they have not accumulated the credits they need to graduate on time. For these students, earning a high school diploma depends on demonstrating within a compressed amount of time that they have met the academic standards required for graduation. For example, the Boston Day and Evening Academy, which serves 355 students across its daytime, evening, and online programs, has been using a competency-based model to help students recover credits and obtain high school diplomas for 17 years. Diploma Plus, a network of 27 alternative schooling programs across the country, also has 17 years of experience using a competency-based model that meets students where they are in terms of learning experiences and allows them to progress once they have demonstrated proficiency at their current level (Nellie Mae Education Foundation, 2011). In this model, students can accelerate their learning if they are able to do so, and they can take more than the standard amount of time to learn a set of material if they need to.

A few research syntheses in the 1970s and 1980s examined the effects of two mastery-based curricula—Learning for Mastery and Personalized System of Instruction—each of which incorporated formative assessment and tailored instruction. Learning for Mastery provided a Response to Intervention-type of approach in that students who did not demonstrate mastery received extra support in K–12 settings. Personalized System of Instruction provided instruction that was self-paced and administered in postsecondary settings (Kulik, Kulik, & Bangert-Drowns, 1990; Slavin, 1984). In comparison to the default curricula in each study, effects of the mastery-based curriculum ranged from slightly to considerably positive, with effect sizes ranging from 0.04 to about 0.78 of a standard deviation, depending on factors like the length of the intervention and the extent to which the outcomes were evaluator-developed or standardized achievement tests (Guskey & Gates, 1985, 1986; Kulik, Kulik, & Cohen, 1979; Slavin, 1984). In a follow-up meta-analysis of 108 studies, Kulik, Kulik, and Bangert-Drowns (1990) found effects of 0.49 for the 17 studies that were conducted in K–12 settings; estimates were more conservative (0.1 to 0.27) for studies that used standardized rather than locally developed achievement measures. The studies employed either randomized or quasi-experimental designs, and the meta-analysis found no meaningful difference in average effects between the two design categories. The studies in the K–12 subgroup used the Learning for Mastery curriculum, which emphasized group-level rather than self-paced instruction but provided extra support for students who were not mastering a given set of content. Across both K–12 and postsecondary studies, the authors found positive effects (with an average postsecondary effect of 0.53 of a standard deviation), and the median instructional time received by the mastery learning groups was only 4 percent greater than in the comparison groups, which suggests that the approaches were cost-effective from the perspective of instructional time invested. The authors found more-positive effects overall for group-based models as compared to self-paced models, for models that provided more rather than less feedback to students based on their quiz performance, and for models that held students to a very high standard of mastery on formative assessments (e.g., 91–100 percent) before they could progress. Finally, the authors reported stronger benefits for low-achieving students than for higher-achieving students, a finding that echoed Guskey and Gates' (1986) earlier synthesis and that bode well for improving equity of outcomes.

Despite the flurry of encouraging evidence from the 1970s and 1980s, we found only one study that had more-recently examined student outcomes associated with competency-based reforms. Haystead (2010) found student proficiency rates on state accountability tests to be 37 to 55 percentage points higher in seven districts and/or schools using Re-Inventing Schools Coalition (RISC) competency-based reforms—based on the Chugach, Alaska model—than in the eight districts and/or schools chosen as comparisons units. The RISC model, which was developed by the author's organization, was described as promoting transparency of curricular expectations, flexible student pacing, students' ownership of their learning processes, and high standards. However, the study's procedures for choosing the one-to-one comparison units were not given (other than noting that the comparison units were similar in urbanicity, ethnicity, and size), nor did the study adjust for baseline student performance prior to the RISC reforms. In addition, the study likely overstated the statistical significance of its hypothesis tests by conducting a student-level analysis without adjustments for clustering within the district- and school-level units of reform. Given these limitations, the study unfortunately sheds little light on the effectiveness of competency-based education.

In part, the lack of recent research on competency-based education may be due to variability around the concept of competency-based education itself (Silvernail et al., 2014). As the aforementioned studies suggest, the definition is a moving target in the research literature. Even in the Kulik, Kulik, and Bangert-Drowns (1990) meta-analysis, the Learning for Mastery model emphasized group-based instruction with feedback and extra support for struggling learners, rather than the kind of individually paced approach advocated by Bloom (1984). In addition, there is a real question of how best to evaluate competency-based models. If students are progressing at their own rates, then examining student outcomes at a fixed time of year for students in largely age-based grade levels may mean that students are being tested on material they have not seen yet, while others are being tested on information they just learned or learned some time ago. An approach that is more flexible in terms of the timing of assessments relative to students' exposure to content might provide a better measure of students' progress. Still, there may be a benefit in evaluating whether some schools or systems help students advance more quickly than others. Such comparisons might be established by assessing what fraction of students reach a given level of progress at a given age, adjusting for students' baseline risk factors.

Some of the impetus behind a competency-based approach is informed by Bloom's finding that one-on-one tutoring yields substantially faster learning than whole-group instruction (1984). In the days before online technology made personalized learning more scalable, whole-class instruction was designed to allow many students to move through the education system in sync, thus vastly reducing the cost of a one-on-one, face-to-face model, which would have proven untenable at scale. Yet the limitation of whole-class instruction is that it forces teachers to teach to the modal student in the class, so that the instruction may be too fast for some and too slow for others. Intelligent computer-based tutoring systems, which adapt instruction based on students' responses to problem solving tasks, provide a potentially scalable alternative. One recent synthesis of 29 studies on these systems found them to be as effective as human tutors, raising student achievement by 0.76 of a standard deviation in comparison to no tutoring at all. The author of the synthesis, VanLehn (2011), estimated the corresponding effect for one-to-one human tutors to be nearly identical, at 0.79 of a standard deviation. (His estimate for human tutors was substantially lower than the two-standard-deviation effect reported by Bloom in 1984; it accounted for nine other studies in addition to that of Bloom.) If VanLehn is correct that individualized computer tutoring can be as effective as individualized human tutoring, then it may have potential to improve upon the whole-class model of instruction that currently prevails in the United States. Moreover, one might expect that the promise of technology-enabled personalized instructional models will only improve as online educational tools are enhanced and refined. Still, the best method for reforming public education in a way that maximizes the potential of new technologies remains unresolved. The Project Mastery initiatives examined in this study provide a few sample cases for considering this question.

CHAPTER TWO

Evaluation Settings and Methodological Approach

Settings for the Study

This report describes the implementation of the three Project Mastery pilot programs undertaken in five school districts during the 2011–2012 and 2012–2013 academic years. It also describes the student performance outcomes associated with the inception of competency-based education in each of the sites, though some sites had adopted competency-based models that predated the Project Mastery grants themselves. As noted in Chapter One, each pilot program was multifaceted and distinctive. In addition, each incorporated technology in ways designed to enhance student engagement and learning.

This chapter discusses the implementation contexts in each site, including the size and demographic characteristics of the districts and schools in which Project Mastery was carried out. Table 2.1 summarizes this information for each site, and we discuss the settings and samples in the subsections below. Following that discussion, we describe our data collection pro-

Table 2.1
Project Mastery District and School Characteristics in 2011–2012 (Year 1)

Site	Setting	Enrollment in 2011–12	Minority (%)	Subsidized Meal Eligible (%)	Student-Teacher Ratio	Schools in District	Ltd. English Proficient in District (%)
Adams 50	*Suburban school district*	*10,124 in district*	*76.0*	—	*18.7*	*19*	35.0
Asia Society	*Intermediary organization*	*2,325 in the four partner schools*	*65.0*	*50.4*	*NA*	*NA*	*NA*
	• Newfound	428	1.9	32.9	11.4	6	0.6
	• Sharpstown	972	89.8	54.7	16.5	309	26.8
	• DCIS	695	59.6	42.4	23.2	166	31.4
	• DCIS-M	230	94.3	88.3	16.7	166	31.4
Philadelphia	*Urban school district*	*154,262 in district*	*77.3*	*66.2*	*16.6*	*266*	*8.0*
	• Six Project Mastery high schools	4,231	71.5	60.6	--	266	8.0

NOTES: Data come from the U.S. Department of Education's Common Core of Data and pertain to the 2011–2012 school year. Minority definition excludes students classified as white or Asian/Pacific Islander. In Adams 50, about 551 students in Mathematics Level 11 took part in the Project Mastery pilot in 2012–2013. In Philadelphia, about 528 ninth-grade students took part in the Project Mastery pilot in 2012–2013. NA: Not applicable.

cedures and our overall analytic methods for the project. In Chapters Five and Six, we provide additional methodological details relevant to the survey and outcome analyses, respectively.

Adams 50

Adams County District 50 is a public school district located in a suburb of Denver, Colo. In Year 1, its 19 schools collectively served 10,124 students in grades prekindergarten through 12. About 72 percent of the district's students were Hispanic or Latino, 19 percent were white, 5 percent were Asian, and 2 percent were African-American or American Indian. Seventy-four percent were eligible for free or reduced-price lunch, and about 35 percent had limited English proficiency. In Adams 50, the entire district had migrated from a traditional education model to a competency-based model starting in the 2008–2009 academic year. This meant that in elementary through high school, students were assigned to academic levels based on their skills and knowledge, not their age-based grade levels. Students were grouped into classes based on their levels in a given subject, and classrooms often included students working at different levels. Teachers provided flexibly paced learning materials that allowed each student to master state standards at his or her own pace. Moreover, students could be reassigned to different classrooms during the school year as they advanced from one level to the next.

The Project Mastery pilot in Adams 50 focused specifically on Mathematics Level 11, which was the performance level corresponding to grades eight and nine in the CCSS. Approximately six teachers of Mathematics Level 11 took part in the pilot—four in the district's only comprehensive high school (Westminster High) and one teacher each from two of the district's three middle schools. The district estimated that the total number of students in the Mathematics Level 11 pilot classes in Year 2 was 551.

Asia Society

Asia Society is a nonprofit educational organization that promotes global understanding and partnership between the United States and Asia (Asia Society, 2014a). Through its International Studies Schools Network (ISSN), it partners with 34 schools across the United States to promote global competence and academic preparedness. Its partner schools receive curriculum and assessment materials and professional development as part of their membership in the ISSN community. For its Project Mastery initiative, the Asia Society joined with three of its ISSN schools (all schools of choice in large, urban districts) and one small-town high school that was not an ISSN member but had a strong focus on competency-based education. The ISSN schools included DCIS and DCIS-M, both located in the Denver Public School District in Colorado, and Sharpstown in the Houston Independent School District in Texas. The non-ISSN school participating in Project Mastery was Newfound. As shown in Table 2.1, the Asia Society's Project Mastery sites collectively served about 2,325 students in Year 1, about 65 percent of whom were minority (again, defined here as not white or Asian) and about 50 percent qualified for free or reduced-price lunches. However, these proportions varied among the sites, from a 2 percent minority population in Newfound to a 94 percent minority population at DCIS-M. DCIS, DCIS-M, and Sharpstown were all schools of choice within their districts, meaning that students were required to choose them rather than being zoned into them by default. Also, their admissions policies differed considerably from each other. Because DCIS was oversubscribed, it took students' prior academic backgrounds into account when admitting students. Sharpstown had reportedly begun to administer selective admissions practices in recent years, but the students who were enrolled in the school through Year 2 of the study had reportedly been admitted during a period when the school was undersubscribed and thus accepted all applicants. The school served a large share of low-income and minority students, most of whom reportedly came from the local neighborhood. DCIS-M was available to students from Denver's economically disadvantaged, far northeast neighborhoods. Under district policy, these students could choose among schools in that part of the city. As was true for DCIS and Sharpstown, students were not assigned to DCIS-M by default but had to choose it. However, since its opening, it had not reportedly been oversubscribed and thus had not needed to turn applicants away.

Table 2.1 also presents characteristics of the districts in which the Asia Society's schools were located. In 2011–2012, Denver and Houston were large, urban districts with 166 and 309 schools, respectively. The Newfound Area School District was markedly smaller, with only six schools. Newfound Regional High School was its only high school.

Philadelphia

The School District of Philadelphia is the largest school district in Pennsylvania. In the 2011–2012 academic year, its 266 schools served about 154,262 students, of which about 77 percent were minority (that is, not white or Asian) and 66 percent qualified for free or reduced-price meals. About 8 percent had limited English proficiency.

To keep the pilot initiative manageable in scale, the Project Mastery grant in Philadelphia focused only on six of the district's high schools, and only on students of the eight ninth-grade ELA teachers who volunteered to take part in the project. Philadelphia began implementing the Project Mastery initiative with a few ninth-grade classes in the final weeks of Year 1, but our study focuses on Year 2, in which about 528 ninth-graders received yearlong exposure to the materials funded by the initiative.[1] Of the schools in the pilot, four were comprehensive high schools with enrollment based on neighborhood zoning. The other two were special admission schools that based admission decisions in part on evidence of prior academic performance. As shown in Table 2.1, the proportion of minority students in the Project Mastery schools was 72 percent, and the proportion of students qualifying for subsidized meals was 61 percent, both of which were only slightly lower than the districtwide averages.

Approach to the Evaluation

As the setting descriptions demonstrate, the Project Mastery grants were awarded to sites that generally served a large share of low-income and minority students. In all of the sites, at least a third of students qualified for subsidized meals. Also, all of the sites other than Newfound were situated in medium-to-large urban or suburban districts and served schools in which more than half of the students were minorities. Given that many of the recent examples of competency-based education have occurred in small or rural settings, the question of how well competency-based education models fare in urbanized settings was of particular interest to the study. It is with this context in mind that we turn to our research questions. In part, these questions focused on the implementation of the Project Mastery grants and on students' educational experiences within classrooms most closely associated with the grant initiatives. In addition, we were interested in the

---

[1] By one estimate, this figure was 643, but some of those students had only partial-year implementation due to a teacher's late entry to the project, so we focused on 528 with full-year exposure.

tensions and insights that emerged from implementation of the sites' various competency-based education approaches, as these may help other sites anticipate and plan for likely hurdles. Finally, we sought to understand how students performed academically under the diverse competency-based education models that each site put into place, even when these models predated and were more expansive than the Project Mastery initiatives themselves.

Research Questions

Our particular research questions were as follows:

1. What materials did the Project Mastery sites develop and implement during the pilot period, and to what extent did their efforts align with the defining features of competency-based education?
2. What were the key tensions or challenges that emerged in the implementation of competency-based approaches?
3. Among students exposed to the Project Mastery pilot programs, how did their self-reported learning experiences vary within and across programs?
4. To what extent did students' exposure to competency-based education models predict their academic performance in mathematics, reading, attendance, or promotion to the next grade?

Research Question 1 is largely the focus of Chapter Three of this report, in which we discuss implementation of the Project Mastery initiatives in each site. Research Question 2 is the focus of Chapter Four, in which we highlight several tensions that arose in the sites with regard to competency-based education. Chapter Five focuses on Research Question 3 by discussing the results of the student surveys from each site. Chapter Six addresses Research Question 4 by focusing on evidence of student performance outcomes in each site, including students' performance on mathematics or reading accountability tests and, where possible, their promotion and attendance rates.

Data Collection Procedures

Site Visit Data

To understand what the sites developed and implemented under Project Mastery and the challenges that arose during implementation, we visited the research sites up to three times each, interviewing the pilot program leaders and curriculum developers for each program, as well as the professional development leaders and teachers involved in the Project Mastery pilots. Interviews focused on characterizing the curricula, LMSs, professional development offerings, district and school support for the pilot program, and alignment of the programs with key features of competency-based education. The final round of interviews also focused on lessons learned and future directions.

Data were collected by two-person teams, consisting of one PhD-level RAND researcher and one junior team member (either a RAND research assistant or a Pardee RAND Graduate School doctoral student). At each site visit, the junior team member typed transcripts. Interviews were audio recorded to allow for accuracy checking and cleaning of typed notes. After the site visits, the junior team member also produced a short summary of each interview, and the lead researcher in each pair typed a set of notes summarizing the site visit overall. Both sets of notes were shared with other team members and discussed among the research team after each site visit. Site visits also included observations of classrooms that were piloting the materials. The observations typically lasted about 30 minutes each. During the observations, we took handwritten notes about the activities under way. The classrooms we visited were chosen by the site leaders to be representative of classrooms focusing on Project Mastery efforts.

We conducted site visits at the end of pilot Year 1 (May or June of 2012), in the fall of Year 2 (November 2012), and in the spring of Year 2 (May 2013). We visited all grantee organizations between one and three times each. Adams 50 and Philadelphia, each of which is a single district, were visited three times each. Because the schools affiliated with the Asia Society were geographically separated, we visited them fewer times. Specifically, we visited the Asia Society headquarters in the late spring of Year 1, and we visited two of the four schools—DCIS and DCIS-M—in both fall and spring of Year 2. The other two Asia Society sites—Newfound and Sharpstown—were visited one time each.

Table 2.2 summarizes the number of individuals interviewed and classes observed, by site and across all sites. The “Total” column aggregates the number of interviews or observations across all visits, though some individuals were interviewed more than once. The “Unique Individuals/Classrooms” column therefore represents just the number of unique individuals interviewed and unique classes observed throughout the study. Finally, the right-hand column represents just the number of unique interviewees in each site who were classroom teachers.

Student Surveys

Students in Project Mastery classrooms were surveyed in May 2013, near the end of Year 2. Questions addressed students’ experiences with the grant-funded curriculum materials, as well as students’ study habits, engagement in school, and postsecondary plans. The survey items were adapted from existing student surveys published by the Chicago Consortium on School Research (2012) and the Tripod Project student survey used in the Measures of Effective Teaching project (Bill & Melinda Gates Foundation, 2013). They were augmented with items specific to the competency-based interventions at each site. The survey was pilot-tested with four middle school students to ensure that questions were comprehensible to students on a sixth- to eighth-grade reading level and could be completed in no more than ten minutes.

Surveys were then administered with the help of each program site. Adams 50 used paper-based surveys; the other sites used an online version. Parents were notified ahead of time about the survey via take-home letters and given a choice regarding their children’s participation, though none opted out. Students provided active informed consent when entering the survey; about 3.8 percent of students who linked to the survey assent information chose not to proceed further. Taking into account these opt-outs as well as students in target classes who never logged in (due to absence from school or lack of time provided in the classroom), overall student survey participation rates were 61.7 percent for Adams 50, 60.6 percent for Asia Society, and 78.4 percent for Philadelphia. The lower panel of Table 2.2 presents the number of students in each site who responded to the survey. Because students were surveyed at only one point in time (May 2013), the survey numbers in columns 4 through 6 (May, Total, and Unique) are identical.

Student Performance Data Sets

We measured students’ academic performance using longitudinal data about students’ academic performance on state accountability tests. These data were aggregated to the district level in the Adams 50 analysis and to the school level in the Asia Society analyses, and were

Table 2.2
Summary of Primary Data Collection Efforts Throughout the Study

Site	Individuals			Total (Individuals/ Instances)	Unique Individuals/ Classrooms	Unique Teachers Interviewed
	May 2012	November 2012	May 2013			
Interviews						
Adams 50	12	8	4	24	17	6
Asia Society—Central	8	1	1	10	8	NA
Asia Society—Newfound		6		6	6	2
Asia Society—Sharpstown			10	10	10	4
Asia Society—DCIS		10	11	21	15	9
Asia Society—DCIS-M		9	6	15	9	7
Philadelphia	6	13	11	30	20	5
Cross-site total	**26**	**47**	**43**	**116**	**85**	**33**
Classroom Observations						
Adams 50	1	2	2	5	4	
Asia Society—Central				NA	NA	
Asia Society—Newfound		4		4	4	
Asia Society—Sharpstown			5	5	5	
Asia Society—DCIS		1	2	3	3	
Asia Society—DCIS-M			4	4	4	
Philadelphia	1	4	2	7	6	
Cross-site total	**2**	**11**	**15**	**28**	**26**	
Student Surveys Received						
Adams 50			340	340	340	
Asia Society—Central			NA	NA	NA	
Asia Society—Newfound			146	146	146	
Asia Society—Sharpstown			288	288	288	
Asia Society—DCIS			105	105	105	
Asia Society—DCIS-M			106	106	106	
Philadelphia			414	414	414	
Cross-site total		**0**	**1,399**	**1,399**	**1,399**	

NOTE: The Asia Society headquarters does not include a school site, so no surveys or observations were conducted there.

publicly available from the state websites in each case. We tracked student performance from the earliest publicly available year (2005–2006 or later) through the 2011–2013 academic year where possible.

Demographic data about the schools came from multiple sources. For the analyses of Newfound, DCIS, and DCIS-M (all Asia Society schools), school-level demographics were drawn from the National Center for Education Statistics' Common Core of Data, rather than from the state data sets, since state data provided demographic information at the district rather than school levels. For Adams 50, we used district-level demographics from the state website, and for the other Asia Society school, Sharpstown, school-level demographics were available from the state website.

In analyzing the effects of Project Mastery in Philadelphia, our focus was on student-level rather than district- or state-level data. We received de-identified student-level data from the district for all of its ninth-graders in Year 2. These data included students' indicators of Project Mastery enrollment, as well as baseline (seventh and eighth grade) scores on state tests in ELA, mathematics, and science. They also included student-level demographics, ninth-grade attendance, and ninth-grade indicators for on-time graduation to grade ten. Finally, they included students' scores on benchmark tests in ELA administered in the fall and winter of students' ninth-grade year.

Analytic Methods

Research Questions 1 and 2

To address the first two research questions, which focused on what the sites developed and implemented (question 1) and on the challenges that arose during implementation (question 2), we analyzed the detailed notes from each interview and site visit, using a matrix for each program, in which rows for each interview were crossed with themes collectively established by the research team. The team chose to focus on themes that, in our early debriefs about the site visits, appeared salient across sites. The thematic categories that we employed included

1. key accomplishments under the grant (in terms of curriculum, technology, and professional development)
2. incorporation of the original Project Mastery tenets (defined progressions toward mastery, anytime/anywhere learning, and credit for mastery)
3. insights from the implementation process (challenges, enablers, stakeholder engagement, and lessons learned)
4. policy context of the sites
5. next steps envisioned by the site.

The matrix allowed us to identify commonalities and discrepancies across sites. We referred back to the raw notes for context and direct quotations as needed.

Research Question 3

For the third research question, involving students' self-reported learning experiences in pilot classrooms and schools, we summarized the proportion of students agreeing or strongly agreeing with survey items that concerned their acquisition of valuable skills, their enjoyment of learning, their ability to choose what and how they learned, their ability to learn at a flexible pace, their use of technology to track their own learning, and their ability to learn anytime/anywhere. Responses to key items were disaggregated by program site, as well as by students' gender and race/ethnicity.

Research Question 4

To examine trends in student performance across sites, we focused on students' exposure to competency-based education models rather than on their exposure to the Project Mastery–funded materials, per se. The reason is that students' exposure to the materials would have been unlikely to yield a notable effect on student achievement or attendance in most cases. In Adams 50, for example, the Project Mastery materials (described in Chapter Three in greater detail) were designed to supplement existing teaching materials but not to constitute a major part of the curriculum. In the Asia Society, the curriculum materials were designed as exemplars for teachers, and the performance and outcomes and rubrics were designed to guide teachers' lessons and assessment practices, but neither constituted a major departure from the existing teaching approaches in these sites. It was instead the Adams 50 and Asia Society contexts, above and beyond their Project Mastery–funded endeavors, that seemed particularly relevant for understanding students' performance under competency-based education models. For this reason, we chose to focus our analysis in the Adams 50 and Asia Society sites on the performance of their students relative to similar students in their respective states. Our intent was to estimate the effect of students' exposure to competency-based education as compared to the typical education approach in a given state. However, our ability to draw causal conclusions about the effects of competency-based models is constrained by design limitations, as described in Chapter Six.

It was only in Philadelphia that the Project Mastery materials replaced the curriculum in a given content area (ELA) for a substantial amount of time (one school year), and for a well-defined subset of students (those in Project Mastery classrooms). Moreover, the context in Philadelphia outside of the Project Mastery classrooms was that of a traditional urban school system that had not established a competency-based educational approach. These facts allowed us to take a different analytical approach in examining student performance outcomes in Philadelphia. There, we compared the performance of ninth-graders in the district who were and were not enrolled in Project Mastery classrooms during the implementation year, 2012–2013.

The analyses used for each program are summarized in Table 2.3 and described below. In each site, we focused on the available academic outcomes—generally mathematics or ELA scores—that were most salient to each site's emphasis in its Project Mastery initiatives. For Adams 50, this was mathematics, so our analysis focused on mathematics test scores. In Asia Society schools, humanities courses were described as placing greater emphasis on competency-based approaches, so we focused primarily on students' performance in ELA. In Philadelphia, the initiative focused specifically on ELA arts classes, so we focused on students' performance in that content area. We also were able to examine students' attendance and on-time promotion to the next grade with the Philadelphia data.

In Adams 50, because competency-based reform occurred at the district level in 2008–2009, we compared academic performance in the district to similar districts in Colorado using a district-level synthetic comparison group approach (Abadie, Diamond, & Hainmueller, 2007). In this method, the other districts in the state were weighted according to their similarity to Adams 50—the "treatment" district—in the years prior to implementation of the intervention. To assess the competency-based education effect, we compared student performance in the treatment and weighted comparison districts in the years following Adams 50's implementation of competency-based education.

Table 2.3
Summary of Analytic Approaches by Site

Program	Analytic Approach
Adams 50	District-level synthetic comparison group
Asia Society	School-level synthetic comparison group for Newfound; covariate adjustment similar schools analysis for DCIS, DCIS-M, and Sharpstown
Philadelphia	Matched student-level benchmark data and attendance within school, grade, and year

As noted above, the one traditional high school among Asia Society schools (Newfound) was also located within a state (New Hampshire) that had legislated a competency-based high school graduation policy to take effect in the 2009–2010 academic year. Though the policy shift occurred statewide, Newfound was reported to have been the first school in New Hampshire to make the conversion when the policy took effect. Our conversations with stakeholders in the competency-based education movement, many of whom were based in the New England region (Lewis et al., 2013), corroborated this understanding. Consequently, our analysis of student performance in Newfound used a synthetic comparison group approach dating back to the inception of the state policy, 2009–2010, which was also the year that Newfound had transitioned to a competency-based model. The question of interest is how well Newfound, which was on the vanguard of the transition to competency-based education, performed relative to similar schools in the state after the policy took effect.

Among the schools in the Asia Society Project Mastery partnership, three were schools of choice, meaning that not all attendees were zoned to attend the schools but instead had to choose them. This choice occurred through a competitive application process at DCIS and through an open admission process at DCIS-M and Sharpstown for the students enrolled during the study years.[2] These three Asia Society schools have had a competency-based orientation since their inception, with an emphasis on project-based learning, using rubrics to evaluate assignments, and requiring students to complete high school graduation portfolios. This fact made it difficult to define pre- and post-intervention periods in terms of each school's academic performance. Instead, we examined the performance of these schools relative to their states' average performances *and* relative to schools that had similar demographic characteristics. To do so, we regressed the performances of all schools in the state on a school-level treatment status indicator, as well as an array of time-varying, school-level demographic variables.

For Philadelphia, we used student-level data to compare the ELA performance of ninth-graders on the district formative assessment to the performance of their peers exposed to the standard curriculum, adjusting for students' baseline demographics and performance. We use student-level propensity score weighting in regression models that also control for student-level and school-level characteristics. The models use school random effects and/or fixed effects (in some specifications) to adjust for the nesting of students within schools. Using the same analytic strategy, we also examined students' yearlong attendance rates and on-time promotion rates from grade nine to ten. In Chapter Six, we provide additional details about the analytic methods for all of the site.

---

2 Sharpstown has since moved to a more selective admissions process due to oversubscription, but the students who were enrolled in the school during the years of the study had reportedly enrolled at a time when admission was still open to those who applied. At both DCIS-M and Sharpstown, a majority of students were reported to come from the schools' economically disadvantaged local neighborhoods. This is reflected in the demographic characteristics of the schools, as shown in Table 2.1.

CHAPTER THREE

Intervention Development and Implementation

As part of its Project Mastery proposal, each grantee committed to build or purchase curriculum materials and instructional tools that would facilitate competency-based instruction. In the timeline of the project, the 2011–2012 school year was designated primarily for intervention development, and 2012–2013 was designated for implementation of the newly developed tools. In practice, the activities of the two project years were somewhat less discrete, in that some of the sites began implementing their new tools during 2011–2012, and some of the product development and refinement continued into 2012–2013.

Materials Developed and Implemented in Each Site

In the first part of this chapter, we summarize the key features of each grantee's Project Mastery program. Each discussion begins with a subsection on *development*, in which we describe what the initiative developed or purchased under the grant. Following this discussion, we provide a subsection on *implementation*, in which we describe what the site had implemented in its schools by the end of Year 2, based on our interviews with site leaders and teachers and based on what we observed during site visits. The purpose of these subsections is to familiarize readers with the Project Mastery initiatives and the ways in which they complemented the instructional context in each of the sites. By illuminating the features of the grant projects themselves, we provide context for the implementation lessons we describe in Chapter Four, for the student survey findings described in Chapter Five, and for the student outcomes we present in Chapter Six. Although Chapter Six focuses mainly on students' exposure to competency-based education writ large rather than on exposure to the Project Mastery initiatives, our description of the initiatives in this chapter helps to justify that analytic choice.

In the second part of this chapter, we discuss the alignment of instructional practices in each Project Mastery site—including the individual Asia Society partner schools—with the defining features of competency-based education that we described in Chapter One: flexible pacing, student choice, and evaluation based on evidence of proficiency. We do this not as a measure of the sites' fidelity to a prescribed model, since the defining features we articulate were identified in the course of this study and not at its outset, and because the sites' approaches were never intended to be homogenous. Rather, we discuss the sites' alignment with these features in order to illuminate the distinctive elements of competency-based education that were emphasized in each site. This discussion, too, sets the stage for our qualitative and quantitative analyses in the chapters that follow.

We turn first to Table 3.1, which presents an overview of what each site developed and implemented, and to our discussion of the materials developed and implemented as part of each grantee's Project Mastery initiative.

Table 3.1
Summary of Project Mastery Pilot Interventions, Implementation, and Contexts

Issues	Adams 50	Asia Society	Philadelphia
What the sites created	• Four Operation Space Wolf games focused on geometry, exponents, and rational numbers • "Flipped classroom" teacher-made videos • Upgrades to Educate competency-based LMS • LevelUp LMS in development to marry student-info systems to competency-based curricula	• Performance outcomes, and rubrics in grades eight and ten, aligned with extant versions for grade 12 • 12 curriculum modules (one- to four-week units) spanning ELA, math, science, world languages, art, and history • Eight one-day teacher professional development modules • ShowEvidence LMS	• 5 Educurious units for grade 9 ELA • 1 Philadelphia Writing Project (PhilWP) unit for grade 10 ELA • Afterschool badges course in digital filmmaking • Pearson LMS and Pathbrite portfolio system
Scope of pilot (schools, grades, teachers, students)	• Seven teachers in math level 11 (grades eight and nine) blended-learning pilot and about 551 students in one high school and three middle schools	• About 1,064 students in four secondary schools	• Eight teachers and about 528 pilot students in six high schools
Implementation depth (frequency/duration)	• Teachers created several flipped-classroom video lessons • Four games were rolled out as a curriculum supplement • Educate LMS was upgraded for ease of use	• As designed, performance outcomes and rubrics would cover whole school year; modules would cover one to eight weeks but were designed as exemplars; teachers also created two modules per year • Grade 10 performance outcomes and rubrics were available online during 2012–2013 school year • Grade eight performance outcomes and rubrics were introduced in summer 2013 • Pilot teachers were exposed to Show-Evidence LMS	• Educurious units supplanted the standard ninth-grade curriculum in Project Mastery classes throughout 2012–2013 school year after a mini-pilot at the end of the 2011–2012 year • About 50–60 students took part in badges program for approximately seven months
Context	• Districtwide competency-based curriculum with flexible pacing, mastery grading (based on extent of mastery—not letter based)	• Three schools of choice that emphasized project-based learning and portfolio-based assessment • One traditional high school sharing these emphases	• Traditional, large, urban school district

NOTE: Summaries cover the two study years, 2011–2012 and 2012–2013.

Adams 50

The Project Mastery in Adams 50 focused on a "blended learning" pilot program, where learning takes place both in physical classrooms and in virtual, online settings. Because the entire district maintained a competency-based education program, teachers needed to be able to provide a diverse array of instructional options that would allow students to progress at different rates. To this end, the pilot involved the development of four online mathematics games, as shown in Table 3.1. It also included training to help teachers create and post their own instructional videos on the district's internal website. Because student learning was flexibly paced and involved a diverse set of curriculum materials, teachers needed a way to record evidence of students' learning across data from many sources, including classroom assignments and computer-based assessments. The Project Mastery initiative involved upgrades to the LMS, Educate, that teachers used to track students' learning, and it involved collaboration with developers of another tool, LevelUp, which was meant to further streamline the data management process.

Operation Space Wolf Mathematics Games

Development

A core element of the Project Mastery initiative in Adams 50 focused on the development of four online mathematics games, collectively known as Operation Space Wolf. The games were created for Adams 50 by Intific, a software development company based in Texas. The Intific team reportedly started game development by building a library of engaging game mechanics, and then applying the appropriate learning targets to those mechanics. Four games were developed in total, all intended for math students at level 11, defined by Adams 50 as corresponding to the CCSS in mathematics for grades eight and nine. The four games were called Quadrant Defender, Space Base–10, Rational Blaster, and Flexigons. All games were designed to be iPad-friendly and accessible from school or from home. Math concepts and learning targets addressed through these games included distinguishing rational from irrational expressions and undertaking geometric transformations. According to the developers, the goals of the games were twofold: to provide students with the opportunity to practice specific math skills, and to act as a formative assessment that highlighted the skills on which students needed additional support. Developers described the games as one tool in a teacher's toolbox, not as products intended to replace the teachers themselves. Throughout the development process, the Intific team worked closely with Adams 50's blended learning pilot teachers, administrators, and consultants. Intific staff also conducted student focus groups to gain student input and opinion about the games.

Implementation

The Operation Space Wolf games were rolled out during the 2012–2013 school year, with final release in the spring of 2013. The games were reported to have been well received by both teachers and students. One teacher voiced appreciation for the fact that even though the games were academically challenging, students continued to work with them. We also heard reports of students logging into the games and playing them at home. However, in-school use of the games seemed to vary across classrooms. Some teachers noted that the games were used primarily as "add-ons"—as rewards or content-area reviews for students—rather than being more deeply incorporated into course content. There was also a report, though, of one case in which two teachers teamed up to build a unit around one of the games. District leaders said they anticipated seeing more of this type of collaboration in the future, particularly as teachers heard success stories from their colleagues.

The timing of the game release may have hindered implementation of the game in some cases, as it occurred near the end of Year 2, just prior to spring testing. At the time of our final site visit to Adams 50 in May 2013, district staff reported that efforts were underway to encourage use of the games. Games were being advertised on television monitors in school hallways, for example, and there had been a contest for students to see who could earn the most badges—or certifications of proficiency—within the games. However, limited hardware resources in the form of computers and tablets were reported to be an ongoing challenge faced by teachers at Adams 50 for implementation of the Operation Space Wolf games and other blended learning tools.

The data features that provide teachers and administration with information about student use of, and success with, the games were also rolled out during the 2012–2013 school year. However, teachers described some limitations of these features. For example, data were only recorded if a student logged in, and clock time wasn't recorded, so a teacher would not know if a student had played a game for any length of time but had never earned a badge.

Development of the games was funded by the Project Mastery grant. As a partner in that development and testing, Adams 50 retained rights to use the Operation Space Wolf games going forward, but Intific, as the developer, retained license and intellectual rights. District staff said they viewed the games as a nice supplemental resource, but one that they could not have afforded without the Project Mastery grant, so they had no plans to continue developing games like Operation Space Wolf in the future. They planned instead to make use of free, online instructional resources to the extent possible.

Flipped-Classroom Teacher-Made Videos and Other Blended Learning Resources

Development

Given that instruction in Adams 50 happened at different paces for every student, teachers described preparing file cabinets with leveled worksheets to allow students to proceed through content at a flexible rate. However, administrators and teachers said they were seeking instructional alternatives that would provide students with a richer array of flexibly paced learning experiences. District staff reported that all teachers were offered two three-day professional development workshops in the 2011–2012 school year on how to create instructional videos to flip their classrooms—that is, to let students watch lectures on their own time and complete assignments, ideally with teacher guidance and support, during class time. This training was well received by interviewed teachers, though it was described as giving teachers the "basics," leaving them to feel their own way in actual video development.

As part of its Project Mastery funding, the district also purchased two online mathematics instructional units from the Florida Virtual Schools. These were intended to provide additional digital resources for students in the blended learning pilot program.

Implementation

Development and use of flipped classroom videos reportedly varied from teacher to teacher in Adams 50. According to district staff, a few teachers had created online libraries of videos for students to access, while others had attended the flipped classroom training but had not yet created their own videos, and some were accessing videos that their peers had developed. Overall, the teachers we interviewed described themselves as enthusiastic about flipped videos as a resource, though they noted that time and technology resources were limitations to development and implementation.

District staff noted that teachers also sought out and began using other online instructional resources for their students, including a free, competency-based, online resource called IXL, which provided leveled mathematics problem sets; the Progressive Math Initiative's (PMI's) online instructional resources and assessments developed by the New Jersey Center for Teaching and Learning; Raz-Kids, which reportedly provided downloadable instructional slides; and the instructional videos developed by Khan Academy. Some teachers mentioned using Everyday Math, a McGraw-Hill product that provided online instructional resources and assessments by subscription.

In one classroom, we saw an example of how teachers in the blended learning pilot program were incorporating online resources. The teacher had created worksheets for different math learning targets. Each worksheet included practice questions, vocabulary, and notes, as well as Quick Response Codes that linked students directly to a Khan Academy video, flipped classroom video, or another online resource.

District staff stated that their focus would continue to shift toward open-sourced digital curricula, because they viewed the ability to mix and match these resources as being well aligned with the competency-based education model. They noted that the Florida Virtual Schools unit had much lower takeup than the free resources that teachers located themselves because the Florida units were not designed around a model of highly flexible student pacing. District staff noted, for instance, that teachers had found the free, online PMI tools so useful that the district planned to adopt PMI as an official math curriculum during the 2013–2014 academic year.

Upgrades to the Educate LMS Platform

Development

Educate was a LMS adopted by Adams 50 to help teachers track students' mastery of required skills within the district's competency-based system. District staff said they were drawn to Educate in particular because, rather than linking students' learning outcomes to particular courses, it allowed for continuous tracking of students' mastery of academic standards throughout their academic careers. Though Educate was adopted by the district prior to the Project Mastery grant, part of the Project Mastery resources were focused on working with the developer to improve the system. The district worked directly with Educate's creator to make the software as useful as possible within the Adams 50 context. Efforts were under way to make Educate more of a "robust" LMS rather than simply a gradebook for teachers. Part of this involved developing application programming interfaces that would allow Educate to communicate with other district software, including the Infinite Campus system used by the district to manage other student data, such as attendance and contact information. District staff said their goal in providing the Educate system for teachers was to "push" resources out to teachers, students, and parents that made it possible to track students' mastery over time of the skills required for high school completion.

Implementation

During the years of the Project Mastery initiative, the Educate platform was reportedly being used by teachers across the district, from elementary school to high school. However, district staff and teachers reported that the district's initial adoption of Educate, which occurred well before the inception of Project Mastery, had been a bumpy process. Educate's initial interface was not viewed as user-friendly, and teachers said they experienced some technical challenges in using the system itself. Furthermore, Educate was not the only data management system employed by teachers. As noted, Infinite Campus was also used to record administrative data such as attendance and contact information. Several teachers said they found Educate unwieldy and that they also needed to track student progress through Excel spreadsheets that they maintained separately. They explained that they had lost unsaved data easily in Educate and that Educate allowed them to record evidence of mastery but not other important student performance data, such as missing assignments. Manual tracking of students' progress in Excel spreadsheets was cited by some teachers as necessary due to reliability and flexibility limitations in the Educate software.

Improvements to Educate were ongoing during the Project Mastery implementation period, and we heard reports from district staff and some teachers that the system was improving, including features such as its interface. However, as of our last site visit to Adams 50 in May 2013, the high school had decided to discontinue use of Educate in the following school year and to use the gradebook feature of Infinite Campus instead. We heard mixed messages as to whether this decision was motivated by limitations in the software's functionality or limitations in its interoperability with the district's Infinite Campus administrative data system.

Creation of the LevelUp Platform

Development

LevelUp is an open-source platform being developed for Adams 50 at no charge by EffectiveSC, a nonprofit technology development firm based in Denver, Colo. District officials said their vision in partnering with EffectiveSC was that LevelUp would eventually act as the "middleware" between the district's student information systems (e.g., Infinite Campus and Educate) and the online instructional resources in which students completed learning tasks and demonstrated evidence of content mastery (e.g., watching Khan Academy videos, earning badges within Intific games, or completing online problem sets through IXL). In other words, the idea was that LevelUp would integrate evidence about students' learning from a variety of sources, tabulate the data, and track evidence of students' mastery longitudinally so that teachers themselves did not have to spend so much time on data entry and performance tracking. EffectiveSC and district staff reported a strong collaboration between the district and the company in developing LevelUp. This collaboration reportedly included a committee of teachers who met on a monthly basis to provide feedback on the system.

During an early site visit to Adams 50, district officials referred to LevelUp as the "key to success" for the district's competency-based education system, and this sentiment was echoed throughout many of the interviews. There appeared to be high hopes across the board that LevelUp would be the tool needed to make competency-based education in Adams 50 manageable. The district described LevelUp as integral to its blended learning strategy for incorporating online curricula and assessments into students' learning experiences. However, EffectiveSC did not receive resources under the Project Mastery grants and was developing LevelUp without external payment, in the hope of creating a tool that might prove useful to competency-based systems.

Implementation

The development of LevelUp reportedly slowed over the course of the Project Mastery initiative. Because EffectiveSC was developing LevelUp pro bono, the district had little leverage to push on the development timeline. Near the end of the grant period, one teacher noted that the LevelUp committee had not met in several months. At the time of our final site visit, LevelUp remained in an alpha stage, with a fuller beta release anticipated for summer 2013. As it was still in development, teachers were not yet using the LevelUp platform in their classrooms.

Asia Society

Asia Society's Project Mastery grant focused on building the curriculum and assessment resources it provided through its International Studies Schools Network. Under the grant, Asia Society created performance outcomes (i.e., sets of expected competencies) and corresponding assessment rubrics for grades eight and ten in several content areas. The organization also developed ten sample curriculum units that teachers could use in their classrooms and deploy as models for their own curriculum development. The grant also funded the creation of four professional development modules for teachers, and it supported Asia Society's partnership with an LMS developer called ShowEvidence. ShowEvidence was creating an eponymously named LMS through which students could upload assignments and receive online feedback and rubric-based evaluations from their teachers. ShowEvidence was also designed to serve as a two-way platform allowing teachers to share and procure curriculum modules throughout the ISSN.

Performance Outcomes and Rubrics

Development

Asia Society's Graduation Performance System (GPS) was intended to articulate clear learning progressions for its students and teachers. Prior to the Project Mastery grant period, Asia Society had partnered with the Stanford Center for Assessment, Learning, and Equity at Stanford University and with teams of teachers from across the ISSN to create performance outcomes for students at the 12th-grade level across different disciplines—including global leadership, ELA, mathematics, science, history/social studies, world languages, and the arts. This partnership also created a rubric with four levels (emerging, developing, proficient, and advanced) for each of the performance outcomes. As part of its Project Mastery initiative, Asia Society was working on "backmapping" these 12th-grade performance outcomes and rubrics to grades eight and ten. Asia Society's headquarters was undertaking the backmapping process so that the outcomes and rubrics for grades eight and ten led naturally to the final 12th-grade outcomes. In other words, the lower levels of the 12th-grade rubrics became the upper levels of the tenth-grade rubrics, and so on.

Implementation

The grade-ten performance outcomes and rubrics were made publicly available on the Asia Society's ISSN website during the summer of 2012, and the grade-eight performance outcomes and rubrics were released during the summer of 2013. Use of the GPS framework and these performance outcomes and rubrics appeared to vary across sites.

At DCIS, the GPS was considered the framework for all of the school's Project Mastery components and a guiding force for teachers' curricular choices and development. The GPS performance outcomes and rubrics were reported to be actively adapted and used in classes, particularly in the 11th-grade "Passages" class and the 12th-grade "Portfolio" class, both of which emphasized independent, project-based learning carried out at a flexible pace over the course of the school year. Similarly, teachers at Sharpstown said they used the GPS performance outcomes and rubrics to guide their project-based learning and the creation of the units within their courses. At DCIS-M, while individual teachers said they knew about the GPS framework and sometimes used it as a model for their planning, teachers reported less collective focus on using the performance outcomes and rubrics than did the teachers in the other two ISSN sites.

Newfound was the only Asia Society partner in Project Mastery that was not actually part of Asia Society's ISSN, so its teachers described less familiarity with the GPS performance outcomes and rubrics than teachers in the other Asia Society sites. However, Newfound's International Club, which was an afterschool, credit-bearing Extended Learning Opportunity (ELO), was reportedly organized around the Asia Society's GPS framework. Staff reportedly used the Asia Society performance outcomes and rubrics when evaluating students' out-of-school learning products for academic credit.

Sample Curriculum Modules

Development

Asia Society headquarters created 12 sample curriculum modules with embedded performance assessment tasks. The modules were curriculum units intended to last approximately one to four weeks each. As samples, they were intended for teachers not only to use with their students but also to emulate in building their own modules and performance assessment tasks.

Most of the new modules focused on mathematics, ELA, or science, but Asia Society also developed modules for world languages, arts, and history. The modules included instructional resources and references, formative and summative assessments of student learning, scoring guides, and examples of what student work would look like at various levels of proficiency. They were intended to be disseminated to Asia Society teachers through ShowEvidence, an online LMS that Asia Society had adopted as part of the Project Mastery initiative.[1] ShowEvidence is described in greater detail later on.

The modules that were developed under the grant were aligned with Asia Society's GPS outcomes at the grade nine-to-ten level or the 11-to-12 level. As modules were developed, they were intended to be piloted across Asia Society's ISSN and then refined and improved based on feedback. Using these modules as examples, ISSN teachers were also responsible for developing their own curriculum modules using the GPS framework—at least two per year—and sharing them with Asia Society headquarters for review as part of an external evaluation of the GPS (unrelated to this study). Asia Society staff reported that two curriculum modules were developed in Year 1, and that ten additional modules were completed by the end of Year 2.

---

[1] Project Mastery funded the customization of ShowEvidence to meet the needs of the Asia Society sites. Funding for schools' licenses to use ShowEvidence was secured through a separate grant from the Hewlett Foundation.

Implementation

The two modules that were developed during Year 1 were introduced to instructional coaches during their summer professional development workshops of 2012. An additional four modules were introduced in October 2012, near the start of Year 2. The final six were introduced during the summer of 2013.

After receiving the modules, instructional coaches were able to share them with teachers in the four professional development sessions they led for teachers each school year. The professional development sessions aimed at supporting teachers in creating their own modules and performance assessment tasks. The sample modules were also available to teachers through the ShowEvidence platform—though, as discussed below, many teachers reported that they were not regularly using the platform. Teachers across the Asia Society sites reported that they received professional development for designing modules and performance assessment tasks and were able to collaborate with their colleagues in designing their own modules.

Though teachers reported limited use of the sample curriculum modules in their classrooms, they described extensive development and use of their own modules. At DCIS in particular, teachers talked at length about creating their own modules, with performance tasks that were designed to be applicable to real-world challenges. One teacher described the modules she had created in previous years as being part of the "toolkit" she was able to use in her current classes. At Sharpstown, teachers also described creating their own modules and expressed the desire to be able to share modules with other teachers at other Asia Society schools, which was not something they were able to do at the time.

Newfound had received less exposure to the Asia Society curriculum modules than the other sites because it was not part of the ISSN network. Nevertheless, teachers there did describe developing their own curriculum modules with performance-based assessment tasks, and our review of these materials suggested that they were similar in design to the Asia Society modules. Newfound teachers said they also used Asia Society principles, such as student choice and global awareness, to inform the competencies they expected all students to achieve.

Teacher Professional Development Modules

Development

Asia Society headquarters developed a series of professional development modules designed to take teachers through a sequence of learning about implementing the GPS and developing their own performance assessments for students. Each module represented one full day of professional development and was housed on the ShowEvidence platform. Asia Society had chosen to develop four professional development modules per year to reflect the amount of professional development teachers realistically had time to receive within a school year. To that end, they developed four in Year 1 of the grant and an additional four in Year 2, bringing the total number of new professional development units to eight during the grant period.

Implementation

Professional development modules were administered to teachers through a two-step process. First, ISSN instructional coaches received training each summer on the new, daylong professional development modules and on the sample curriculum modules that Asia Society had developed for teachers to emulate—as well as on the process of using ShowEvidence to access the modules. Coaches then returned to their schools to deliver the new professional development—including the sample curriculum modules and how to operate ShowEvidence—

at periodic intervals during the school year. The focus of these professional development modules, as already noted, was on guiding teachers to create their own curriculum modules and performance assessment tasks for students. Though coaches could use their discretion about the timing and emphasis of their professional development efforts, teachers at the ISSN sites did report receiving professional development on the creation of their own curriculum modules, as well as guidance on how to use the ShowEvidence platform.

ShowEvidence LMS Platform

Development

ShowEvidence is an online platform that houses multimedia curricular and professional development modules for students and teachers, and enables teachers to deliver evidence-based feedback directly onto uploaded student work. The platform is online, so it requires no alterations to school or district hardware and is designed for integration with school districts' data systems. Asia Society had been partnering with the developers at ShowEvidence prior to receiving the Project Mastery grant but used the award to help accelerate the platform's progress. Nevertheless, refinements to the platform came more slowly than Asia Society had anticipated. As of May 2013, ShowEvidence was operational in some of its functions, but not all of them. The platform allowed teachers to author and upload modules, tasks, and rubrics, and to upload and score student work. However, data reporting and student portfolio components were not yet ready for implementation. An updated version of ShowEvidence with improved functionality was reportedly introduced to coaches at Asia Society's summer professional development workshops in 2013.

Implementation

Delays in ShowEvidence development yielded delays in school-level implementation as well. Teachers and administrators in the Project Mastery sites reported that they had heard of ShowEvidence, and many said they had received some professional development on the system. However, most of the teachers we spoke with said they did not anticipate ShowEvidence being ready for regular use during Year 2, and described mixed feelings about incorporating ShowEvidence into their practice. Some said they felt like it was an exciting new tool that would help make planning easier, while others expressed concerns that it would actually demand more work. As of May 2013, near the end of Year 2, a few teachers at Sharpstown said they were testing ShowEvidence and providing feedback to the developers. Others at the school said they would prefer to focus on creation of their own GPS curriculum modules and that they did not have enough time to learn to use a new, unfamiliar technology platform. At Newfound, teachers were unsure whether they would ultimately use the ShowEvidence platform, given that they were not officially part of Asia Society's ISSN and were also concerned about the technological burden. In the interim, teachers in the Asia Society sites commonly reported that they were using Google Documents, a free resource, to host students' online portfolios.

Philadelphia

Philadelphia's Project Mastery initiative focused on the development of five ninth-grade ELA curriculum units that emphasized project-based learning and assessment with rubrics. Created by a curriculum development company called Educurious, these units were designed around the CCSS for ELA and were used to replace the traditional, textbook-based ninth-grade language arts curriculum in the pilot classrooms. In addition, the district created its own Educurious-style curriculum unit for tenth-graders through a partnership with the Philadelphia Writing Project (PhilWP), a professional development group comprising local language arts teachers. The district also partnered with three local organizations—the Philadelphia Youth Network, the Big Picture Alliance, and the public radio station WHYY—to create an after-school program focused on digital filmmaking that was designed to meet several of the CCSS for ELA in the areas of both writing and speaking. Finally, Philadelphia purchased access to an online LMS and a digital portfolio management system intended to improve teachers' and students' ability to communicate online about students' work.

Ninth-Grade Educurious Language Arts Units, and a Locally Developed Tenth-Grade Unit

Development

As part of its Project Mastery initiative, Philadelphia purchased five new ELA units from Educurious. These were intentionally designed around the CCSS in ELA. Each unit was focused on a core piece of literature deemed suitable for a ninth-grade reading level, with reference to related documents, including historical and other nonfiction writing. Each unit also featured several performance tasks of increasing complexity, building toward a capstone task that included not only a traditional writing assignment, but also a multimedia component that reinforced the ELA standards for effective, audience-specific communication. Although the units were designed primarily for synchronous learning rather than flexible pacing, they did incorporate elements of student choice. Specifically, students had choices of supplemental reading content (i.e., books aligned to the unit theme that students chose for at-home reading), of the topics on which their projects would focus and of the order in which to complete particular unit tasks.

Implementation

Two of the pilot teachers began implementing the Educurious curriculum in the final six weeks of Project Mastery Year 1, even before the planned implementation year. At that point, the teachers prepiloted the ELA unit on dystopian fiction. The unit called upon students to read and write about canonical literature, such as Kurt Vonnegut's short story "Harrison Bergeron," as well as more-contemporary fare, including the popular teen novel *The Hunger Games*. It also prompted them to create their own dystopian narratives and create short films about them, designed to develop storytelling and media skills required in the CCSS. The teachers who took part in the prepilot reported high student engagement and higher-than-usual end-of-year attendance, which our research team also observed during a June 2012 visit.

In the summer of 2012, all of the Project Mastery pilot teachers received a training workshop from Educurious on the new materials. That summer and throughout the 2012–2013 school year, teachers communicated regularly with Educurious developers to offer feedback and suggestions on the curriculum.

During Year 2, Educurious completed the other four planned units, and the eight Project Mastery teachers implemented four of the five units. Though they had planned to implement all five during the pilot year, project leaders said teachers realized they had enough content to spend one nine-week marking period on each unit of the four units. This left one unit available for teaching in tenth grade; two of the Project Mastery teachers planned to transition from ninth to tenth grade and used the remaining Educurious unit with their tenth-grade classes. In interviews, teachers reported that the materials reinforced the way they had always wanted to teach, in that the units were very project-based and the literature engaged the students. However, teachers noted—and we observed—that they sometimes had to scaffold the content with supplemental material they developed in order to make the units more accessible to struggling readers and writers. For instance, they would add short quizzes about the reading material to hold students accountable or would create vocabulary lists to explain unfamiliar concepts.

One challenge that the district reported with the Educurious materials was obtaining the right to use them. Through Philadelphia's agreement with Educurious, the district had a right to use the materials only in the Project Mastery classes and for only two years, since licensing fees had not yet been established when the contract was put in place. The 2013–2014 academic year was the final year of that agreement. Since the materials were not yet on the market when the district purchased them, the contract was a research and development agreement, and Philadelphia was a partner in helping Educurious test, refine, and improve the materials as they were created. The district was working with Educurious in 2014 to negotiate usage terms that would allow them to scale the materials beyond the original teachers in the pilot.

In addition, the district had partnered with the PhilWP to build its own Educurious-like units for tenth grade. By the terms of that agreement, the district would own the units and be able to use them as broadly as it wished. During the 2012–2013 school year, PhilWP developed a tenth-grade unit focused on Philadelphia history. This unit was reportedly used in Year 2 by the two Project Mastery teachers who moved to tenth-grade teaching assignments.

Digital Filmmaking Badges Curriculum

Development

In spring through fall of 2012, Philadelphia's collaboration with the Big Picture Alliance and with the Philadelphia Youth Network yielded a roughly seven-month curriculum for an afterschool program that awarded students competency-based badges in digital filmmaking. Badges were given for evidence of proficiency in areas such as storytelling, collaboration, technical skills, and even attendance. The local public broadcasting station, WHYY, joined in the curriculum development process to ensure that the skills taught were relevant to entry-level workplace demands in broadcasting. The partners explicitly designed the curriculum around the CCSS for ninth-grade language arts and recruited three Project Mastery teachers to lead the badges programs after school in their respective schools. The lessons were structured so that students could meet language arts standards in both narrative and persuasive communication through their creation of short films, and so that the earning of a filmmaking badge would certify students' proficiency in skills that would help them find employment.

Implementation

The filmmaking program was implemented from about November 2012 through May 2013 with 50 to 60 students, about 31 of whom attended at least 50 percent of the sessions. The program was implemented in three of the Project Mastery schools, with facilitation by Big Picture Alliance staff, sometimes accompanied by the Project Mastery classroom teacher. During the semester-long program, the students worked in groups to create several short films, including a fictional narrative and a nonfiction documentary. They honed skills in script development, editing, blocking, acting, sound, lighting, procuring locations (mostly through on-campus permissions), digital video recording, and editing. They obtained peer feedback by sharing rough cuts in classroom workshops, and they even made trailers for their short films. Although the curriculum was designed to meet numerous Common Core language arts standards, the three teachers involved in the project differed in their willingness to award classroom language arts credit for students' work in the badges program. One teacher was described as enthusiastic about crediting students' language arts work as demonstrated through filmmaking. However, another teacher with whom we spoke was concerned that demonstrating proficiency in persuasive writing and persuasive documentary filmmaking, for example, were distinct skills that complemented but did not supplant one another. Although the badges curriculum developers pointed out that students had to write scripts (even for documentary scene planning) in order to create their films, the teacher remained reluctant to award in-school credit for out-of-school tasks.

Program leaders reported that students were more engaged by the filmmaking skills than by the earning of digital badges, per se, though the program leaders speculated that this might change if badges began to have greater cache in the larger culture and workforce.

Pearson LMS Platform

Development

Working with the education publishing company Pearson PLC, which also provided the district's SchoolNet student information system, Philadelphia commissioned a LMS that would allow teachers to upload assignments and make them available online to students outside of school, anytime and anywhere. The LMS included functionality so that students could upload their assignments and track which deliverables they had (and had not) completed for a unit. It also included discussion board functionality to allow students to discuss assignments and content with one another.

Implementation

Teachers in the pilot reported that they were using the LMS to post assignments, and that students were using it to upload their work and track assignments they still needed to complete. Indeed, we witnessed teachers linking to assignments in the LMS and projecting their computer screens on their interactive whiteboards, and we saw students logging into the system and tracking their assignments in some classrooms.

Still, implementation was not without challenges. District officials reported that the main challenge was technical. Teachers found that uploading course materials to the LMS was more time-consuming than they anticipated because of particularities of the platform specification. Although teachers reported that they still used the tool, district officials were disappointed that the upload process was less streamlined than they had hoped. The second challenge was also technical. Though the developer had reportedly planned to integrate the LMS with SchoolNet for ease of data management, Pearson had not achieved full integration of the two systems by the end of the Project Mastery pilot period. The result was that students could post assignments and receive teacher feedback within the LMS, but that their official grades were maintained in the district's SchoolNet online gradebook. Consequently, they had to log into two different systems to see the full picture of their performance in a Project Mastery classroom (assignment completion data and feedback in the LMS, and gradebook data in SchoolNet). Teachers, too, had to interact with each system separately.

At the end of the Project Mastery grant, the district elected not to scale the LMS more broadly until the technical challenges could be worked out further. Philadelphia was in negotiation with Pearson for access to another tool in the company's LMS suite that would better meet the technical specifications and instructional needs of the district.

Pathbrite Online Portfolios

Development

As part of its contract with Pearson, Philadelphia also received access to an online portfolio management tool called Pathbrite. The tool was designed to allow students to create portfolios in which they could store, display, and tag exemplary work throughout their high school careers. Unlike work students posted to the LMS, which was always linked to a particular course, work products posted in Pathbrite would exist outside of any given course and could demonstrate students' growing proficiency in core content areas throughout high school. The district viewed Pathbrite as a tool that supported student choice—in that students could post exemplary work of their choice and do so anytime/anywhere—and evaluation for proficiency, since the portfolios provided evidence of student learning that was independent of any given course or length of study. The district envisioned that in time, students might post exemplary work in Pathbrite that would be accessible from each of their course-specific records in SchoolNet.

Implementation

Pathbrite was rolled out in Project Mastery classrooms in the spring of Year 2 but did not catch on strongly with the Project Mastery pilot teachers. District officials said that, despite the aesthetic appeal of the Pathbrite platform, the rate at which teachers adopted it was much slower than their adoption of the LMS. Officials attributed this to the fact that Pathbrite was a late-term addition to the Project Mastery effort. They noted that there was still a need to establish digital portfolios of exemplary work that would follow students throughout their high school years.

Alignment with Defining Features of Competency-Based Education

Beyond summarizing the features of the Project Mastery grants, we were interested in the ways in which each grantee's educational context aligned with the defining features of competency-based education that we described in Chapter One. These contexts are important not only to our understanding of implementation of the Project Mastery grant efforts but also to our analysis of student outcomes under competency-based education models. In this part of the chapter, we disaggregate our discussion of Asia Society sites by partner school because the instructional contexts were somewhat varied across sites.

Table 3.2 summarizes the extent to which each Project Mastery site's design appeared to align with the defining features of competency-based education. These summaries are based on the site visit data we collected during the course of the study, and also on the Project Mastery grant applications and related documentation prepared by the grantees.

With regard to the *flexible pacing* feature of competency-based education, this was a central component of the Adams 50 competency-based model, in that students progressed and were evaluated at their own pace and were typically arranged in multigrade classrooms. This was not only reported by project leaders, but also was observed in our classroom visits, where we saw students working on different content in the same classrooms. In all of the sites but Adams 50, students generally were taught in age-based grade-level cohorts and worked on

Table 3.2
Relationship of Pilot Designs to the Defining Features of Competency-Based Education

Feature	Adams 50	Asia Society: Newfound	Asia Society: Sharpstown	Asia Society: DCIS	Asia Society: DCIS-M	Philadelphia
Instruction meets students where they are (flexible pacing)	Key component (multi-age levels based on proficiency)	Some evidence (emphasis on revision)	Some evidence (graduation portfolios)	Some evidence (graduation portfolios)	Some evidence (graduation portfolios)	Some evidence (teacher-developed scaffolding)
Students have choices to personalize learning	Some evidence (choice of online resources)	Key component (project choice; ELOs)	Key component (choice of project topics and execution)	Key component (choice of project topics and execution)	Key component (choice of project topics and execution)	Key component (choice of project topics and execution)
Students are evaluated on evidence of proficiency	Key component (Scantron-based; mastery grading)	Key component (rubric-based; mastery grading)	Some evidence (rubric-based)	Some evidence (rubric-based)	Key component (rubric-based; mastery grading)	Some evidence (rubric-based)

NOTE: Shading is provided to make patterns easier to see at a glance. Green shading represents a key component for that site; yellow shading represents some evidence that the feature was used at that site.

common assignments. Nevertheless, we found some evidence of flexible student pacing in Newfound, in that teachers spoke of requiring students to redo and revise work until it met the standards for proficiency. Similarly, in the Asia Society ISSN sites (Sharpstown, DCIS, and DCIS-M), teachers described requiring students to redo work many times until it showed proficiency, especially in terms of work for students' graduation portfolios. In Philadelphia, flexible pacing was not designed as a core part of the Project Mastery pilot, but our classroom observations revealed evidence of teachers adapting the Educurious units, which they saw as quite demanding, to meet the individual needs of students. They appeared to do this by creating supplemental materials (e.g., vocabulary sheets and checks for understanding) and through one-on-one conferencing with struggling students about their work, while other students moved ahead with an assignment.

Regarding the second defining feature, *student choice,* all sites reported allowing students to choose some projects, reading materials, or anytime/anywhere activities, though students still had to demonstrate mastery of particular content via specified standards. In the three Asia Society ISSN schools (Sharpstown, DCIS, and DCIS-M), teachers and project leaders described student choice as a key component of their model, in that "Student Choice" was actually the first principle in their SAGE framework for project-based learning.[2] Site leaders referred to this principle as meaning that students had choices about project themes, topics, and execution.[3] For example, students in the ISSN schools chose senior projects with real-world applications, which they had to propose, complete, and defend in the manner of a

[2] SAGE stands for student choice, authentic work, global significance, and exhibition to a real audience.

[3] The other principles in the framework were Authenticity, meaning that a project was similar to a real-world task an adult might have to do in a job; Global Significance, meaning that a project focused on global issues and topics; and Exhibition, which means that students were asked to present to an audience other than their teachers, such as classmates or external audiences (Asia Society, 2014b).

graduate school thesis. However, even in the ISSN sites, we heard about idiosyncrasies in the awarding of credit for out-of-school activities. For instance, DCIS-M described awarding academic credit for a white-water rafting trip that addressed biology standards. At DCIS, the site's attempt to give full-course credit for a class trip to Peru that addressed myriad standards was stymied by district regulations, although the site eventually found a way to award the credit. Project-based learning with elements of choice was also central to the Educurious units developed in Philadelphia, in that students had a number of project-based deliverables for each thematic unit—e.g., essays, films, speeches—for which they could choose topics, the order in which they prepared the deliverables, and the supplemental reading materials they used as part of the unit reading requirements. In Adams 50, project leaders noted that students had opportunities for anytime/anywhere learning, in that students had the option of applying proficiency evidence from nonschool sources (e.g., a Boy Scout badge, or a speech given at church) for in-school credit. Project leaders noted that whether a teacher decided to award credit for these experiences was currently "the luck of the draw," but they said the district would like to formalize that decision process with clearer guidelines. Adams 50 students could choose from among a number of online resources—including, but not limited to, the Operation Space Wolf games—to demonstrate proficiency in a content area. In Newfound, students were also engaging in project-based learning that included elements of student choice, even though the school reported that it had not fully adopted the Asia Society's SAGE framework. In fact, the teacher-developed performance tasks (e.g., curriculum units) that Newfound provided included explanations of how the tasks incorporated choice. For instance, one 40-day physics task focused on building a car (an authentic assessment) and gave students choices about not only the look of the car but also key design elements, including the type of drive train to incorporate. Many of the sites, including Newfound, DCIS, and Philadelphia, had developed ELOs to give students additional opportunities for project-based, credit-earning learning in an afterschool environment. In Newfound, which had been developing credit-bearing, competency-based ELOs for the previous five years, the ELOs were a particularly important mechanism for choice, since they offered students opportunities to design their own electives and solicit professional, external mentoring in topics that suited their interests.

Evaluation for proficiency appeared to be most prevalent in Adams 50, Newfound, and DCIS-M, each of which had made a concerted effort to define clear learning progressions and to evaluate students based strictly on evidence of meeting associated criteria. In Adams 50, this was accomplished in at least two ways. First, students were evaluated with formative assessments for each task completed. Second, teachers' formative evaluations of students' mastery of content was confirmed periodically by standardized Scantron assessments that, if passed at the "proficient" level or better (3.0 on a 4.0 scale), allowed students to move on to the next level in the learning progression (Adams County District 50, 2011). In Newfound, a common set of proficiency expectations was provided by reasonably focused, state-supported efforts to reach a common definition and understanding of the definition of competency. In other words, the pursuit of a common competency definition was a key feature of the New Hampshire effort, and this also led to some evidence in Newfound of flexible pacing, in that interviewed teachers reported asking students to redo work that did not meet the standard, rather than moving on. Moreover, both Adams 50 and Newfound had created competency-based report cards. In Newfound, the report card included both traditional letter grades and four-point competency-based scores for each defined competency, as well as a translation key for equating letter grade and competency scores. DCIS-M project leaders also emphasized their focus on moving teachers toward mastery-based grading (on a 4.0 scale rather than using letter grades) and on evaluating students based on proficiency-criteria rather than effort. In other sites, a common set of rubrics developed by the Asia Society or by Educurious in the Philadelphia pilot did provide common criteria against which each key deliverable was intended to be evaluated. However, teachers still appeared to vary in their approaches to student evaluation; one teacher at DCIS described requiring students to redo assignments five to seven times until they reached proficiency, while another discussed the importance of grading for effort and work habits as well as proficiency. We heard similar sentiments from one teacher in Philadelphia—that is, the perspective that students' evaluations should take their baseline skill levels and needs into account rather than holding all students to a common definition of mastery on any given task. Meanwhile, one teacher at Sharpstown described incorporating students' self-evaluations of how well they had met the standards. Each of these perspectives—including the importance of student self-evaluation—has currency in the modern education landscape, and committed educators might plausibly endorse any one of them. However, the variety of perspectives we heard in several of the sites—particularly DCIS, Sharpstown, and Philadelphia—regarding how students' competency should be evaluated suggested that a common definition of competency was not being uniformly applied in all cases. Among the sites, Adams 50 provided perhaps the clearest safeguard of consistency by using Scantron assessments at key milestones. Because these assessments did not rely on teachers' individual judgments of students' knowledge and skills, they likely increased assessment reliability beyond what individual teachers' judgments might have provided.[4]

Summary

One pattern that emerges is that the Asia Society and Philadelphia sites, despite their geographic dispersion and varying school contexts, seem to have adopted approaches that were more similar to each other than to the Adams 50 approach. The unifying difference is that the Asia Society and Philadelphia approaches put greater emphasis on student choice and authentic assessment through project-based learning and less emphasis on flexible pacing and evaluation for proficiency. They did not disregard the latter two components, but they appeared reasonably consistent in their primary emphasis on project-based educational methods. Adams 50, in an effort to accommodate heterogeneous student levels and needs, placed greater emphasis on allowing students to move at their own pace through the curriculum. This required teachers to devote greater resources to tracking students' proficiency and less to traditional instructional models, since students were essentially finding instruction and demonstrating mastery at their own pace. This, in turn, appeared to leave less time for developing structured projects that could engage students at many levels and offer them choices about topics or execution. To the extent that students had choices, they were largely choices of instructional resources to use (e.g., teacher-created videos, IXL, Raz Kids, and Khan Academy instructional tools) rather than how to execute tasks connected to real-world scenarios. This does not necessarily mean that Adams 50 teachers discounted the need for project-based learning. One teacher we interviewed, for instance, said she required students to complete a project if they wanted to obtain

[4] For a discussion of reliability limitations in teachers' assessments of students' work, including student portfolios, see, for instance, Koretz, Klein, McCaffrey, & Stecher (1994).

credit for *mastery* as opposed to just *proficiency*. But the extent to which teachers were able to create cohesive projects was limited, given the sheer quantity of student performance data and instructional resources they were expected to track across students working on diverse tasks. One teacher reported that he spent 40 percent of his time on data entry alone. Two teachers noted that the work of providing diverse learning experiences for each student each day had left little time for other instructional endeavors, like creating flipped-classroom videos, which both said they had hoped to do. Both the flexible-pacing and project-based models embraced key features of competency-based education, but we saw little evidence among the Project Mastery pilot sites of schools that were able to smoothly integrate all of the features into a single approach.

CHAPTER FOUR

Tensions in the Implementation of Competency-Based Models

This chapter focuses on key tensions and questions that arose across sites as competency-based education was being implemented. We offer a brief description of each of these tensions and examples of how they were manifested in the Project Mastery sites. The tensions converged around the kinds of student work that could be counted for credit and who judges that evidence; how to hold students to a common definition of proficiency; how to make personalized, choice-enabled learning experiences sustainable at scale; and how to promote equitable outcomes within highly personalized systems.

Equating Evidence from Anytime/Anywhere Learning

One key tension that emerged in the study revolved around how to equate evidence of proficiency from different sources. This tension was especially apparent with respect to students' out-of-school learning activities, or what the Project Mastery tenets termed anytime/anywhere learning, in all three of the Project Mastery initiatives. In Adams 50, as discussed in the previous chapter, project leaders noted that the decision about awarding academic credit for out-of-school endeavors was left to the teachers' discretion. However, they expressed a desire to formalize the decision rule so that students could anticipate what kind of credit they would earn for completing a certain task outside of school. In DCIS, project leaders described the importance of travel-based learning experiences in helping students acquire global competence. Yet they said teachers found it difficult to get district approval to turn school-led international travel experiences into course credits for students. Because the district would not allow the school to create new courses without a formal vetting process, teachers had learned to repurpose outdated course numbers in order to credit their students for work performed overseas, such as a blogging project students had undertaken on a summer trip to Peru. This created an extra administrative burden for teachers and required them to find creative ways to award credit for students' out-of-school learning that was not associated with formal classes.

In Philadelphia, as described in Chapter Three, tension arose over whether English teachers should offer their students language arts assignment credit for standards-aligned filmmaking tasks the students completed in the afterschool badges course. As noted, the ELA teachers involved in the badges program varied in their willingness to allow students' assignments from the afterschool program to supplant course assignment expectations. Project leaders said they found teachers' reluctance surprising because the developers of the digital filmmaking badges program had identified the CCSS for ELA that each assignment fulfilled and had included these in the curriculum materials. They included, for example, standards related to crafting and revising pieces of narrative and persuasive writing, which students had to do as part of their script development. However, when we talked with one of the teachers about his reluctance, his rationale was clear: He said he did not believe that the afterschool assignments were suitable equivalents of the projects he assigned in his class. He was not convinced, for instance, that the persuasive writing skills appropriate for script development of a documentary film were the equivalent of what he expected students to achieve in a persuasive essay, especially in terms of issues like essay structure and sentence structure. He believed that preparing a short nonfiction film and preparing a persuasive essay tapped different skills, both applicable to the real world, but not interchangeable.

The teacher's position seems reasonable; the CCSS (and many other state standards) often leave details open for interpretation. For instance, the CCSS ELA/literacy standard W.9-10.1 specifies that ninth- and tenth-graders should be able to "write arguments to support claims in an analysis of substantive topics or texts, using valid reasoning and relevant and sufficient evidence" (CCSS Initiative, 2014). Both a film script and an essay could satisfy this standard, though a related, secondary standard (W.9-10.1C) specifies that students should be able to "use words, phrases, and clauses to link the major sections of the text, create cohesion, and clarify the relationships between claim(s) and reasons, between reasons and evidence, and between claim(s) and counterclaims." The extent to which a film script accomplished this secondary standard would likely depend on how much of the exposition was provided through narration/voiceover as opposed to the juxtaposition of scenes. Still, most of the five secondary standards related to W.9-10.1 focus on the structured presentation of ideas and would therefore transcend the medium of delivery. A teacher's willingness to help students see how the standards apply to various media—including but not limited to essays—may be a determining factor in what receives credit. In other words, there was not common agreement about whether work done outside of a course could count as evidence that students had met some of the objectives of the course.

The site that appeared to have established the most systematic approach to crediting out-of-school learning experiences was Newfound. Staff there reported that they had established ELOs five years earlier as a way for students to obtain elective credit for their involvement in afterschool clubs. In fact, it was this effort, one teacher said, that had put them on the path of evaluating student performance in terms of desired competencies rather than the amount of time they spent in a course. The process of developing the ELOs reportedly helped push the teachers to collectively establish which competencies students were expected to meet in order to graduate. Teachers then worked with students to develop individual or group ELOs that were linked to specific competencies and could be undertaken after school or during the summer. The staff reported that about 30 students took part in individualized ELOs each year, and as many as 200 (out of 440 students in the school) took part in group ELOs. The purpose, the staff said, was to augment the core curriculum and give students an opportunity to pursue topics that were not offered—or were offered in oversubscribed courses—during the main school day.

The Newfound approach offers a promising model for the other sites to consider, since it involves a collective process of early consensus on competencies that students must achieve to graduate, and formally linking ELOs to particular competencies. Despite its promise, this approach was not resource neutral. Newfound staff reported that they used external grants to support the time teachers spent establishing the competencies. (Note that Adams 50 had undertaken a similar, grant-funded approach to compensate its teachers' development of its vertical achievement levels.) Newfound also used a combination of external grants and a district afterschool fund to compensate teachers for leading ELOs during the summer and after school, though some teachers reportedly led some individualized ELOs without additional compensation.

Determining Who Can Authorize Credit

Regarding out-of-school learning, a related question that sites grappled with concerned the evaluation of student work by individuals other than teachers. At issue was whether experts in the field could make binding decisions about the quality of student work and its suitability for academic credit. Several of the sites had cultivated partnerships with businesses and professionals to provide out-of-school learning experiences. Sharpstown, for instance, required its students to take part in internships as part of their graduation requirements. Newfound's ELOs often involved working with community partners, which included local retail businesses, the fire department, a local university, a newspaper, and even a Skype-based partnership with a NASA scientist. In Philadelphia, the Educurious units were initially designed to require that students receive online feedback on their work from experts in the field, but the district had restricted this functionality due to concerns about student safety and privacy.[1] Adams 50 and the other Asia Society sites also created community partnerships to provide out-of-school learning opportunities for students. However, none of the sites had elected to make the community partners formal arbiters of credit. The issue received the most attention in Philadelphia, where the tension between authentic feedback and student privacy had led to the restriction of the expert-feedback component, and where teachers disagreed about granting credit for the digital filmmaking products. Despite a lack of consensus around the latter issue, the Big Picture Alliance, which delivered the afterschool filmmaking classes, did make final decisions about the quality of students' work and whether it met the proficiency levels required for earning badges. For this reason, Philadelphia offered the most salient example of a site in which an outside, community-based partner was delivering instruction and making decisions about students' proficiency on CCSS-aligned tasks, even if these decisions did not always translate to credit in students' regular classrooms.

Maintaining a Common Definition of Proficiency

A second tension that was prominent across sites involved holding all students to a common definition of proficiency. This tension focused, in part, on teachers' hesitancy to disentangle students' effort and prior learning experiences from statements about their current proficiency. Some teachers noted a concern that their preferred grading practices took students' effort and growth into account in ways that might result in a struggling student and a highly proficient student getting the same grade for work of different quality. On one hand, this type of grade-inflation to reward effort or to compensate for poor baseline skills was largely the problem that competency-based models set out to correct. As one Adams 50 staff member explained, "in every classroom, an 'A' was defined differently" under the old system, and a staff member in DCIS-M said that under a traditional, letter-based system, they would have "students with all A's and B's who couldn't read."

It was not clear that simply moving from a letter-based grading model to a model based on rubric assessments and mastery-grading, in which students received proficiency ratings rather

[1] All adults working with students in Philadelphia schools, whether face-to-face or online, were required to undergo federal and state background checks.

than letter grades, was entirely changing this practice. Proficiency rubrics are commonly used in many school districts, so they may not have been particularly new for many teachers (Hamilton, Halverson, et al., 2009). We asked two of the language arts teachers in Philadelphia, for instance, whether they had used proficiency rubrics before they began using Educurious units, and both said yes. One of these teachers also reported taking students' skill levels into account when assigning grades within the Educurious materials. This teacher was concerned that struggling students may have to exert intense and sustained effort to close gaps with their higher-achieving peers. The teacher believed that assessments of student performance should account for that effort and for students' baseline skills.

Moreover, in a high-poverty school system where most students are performing below grade-level standards, it is possible that teachers would lose sight of what evidence of proficiency should look like. This is where norming—the process of helping teachers calibrate their grading to a common set of expectations—may be especially important. The Project Mastery site that appeared to have placed the greatest emphasis on the norming of teachers' evaluation standards was DCIS-M. The school required evaluation for proficiency—sometimes termed "mastery grading"—in all classrooms, and the principal had conducted a "reorientation" workshop for teachers during the 2012–2013 school year, in which teachers were asked to respond to a number of grading scenarios to try to achieve consensus. Teachers also met weekly in grade-level teams to review their grades together. Numeric and letter grades were then determined post-hoc for report cards by the proportion of activities in a course on which students demonstrated proficiency.

Even in a school that strongly emphasized common grading expectations, teachers noted the challenges inherent in grading strictly for evidence of proficiency. One DCIS-M teacher explained that 35 percent of her students failed the first semester once she began grading in this manner. She also pointed to the challenges of holding a common academic expectation, citing the difficulty of evaluating a student who knows the material but has not done the work, as compared to a student who "has been working and working but still can't pass the test." The teacher wanted to be able to credit students for their participation and work ethic, because some of her course objectives were linked to speaking skills, wherein daily participation was critical for eventually achieving mastery.

Though it placed especially strong emphasis on mastery grading, DCIS-M was not the only site to deliver professional development on maintaining common expectations. Sharpstown reportedly also provided a workshop to teachers on using rubrics consistently, and Newfound teachers received a workshop on rubrics that one teacher described as "excellent." Yet another Sharpstown teacher observed that grading approaches were "personal" to the individual teachers, and a Newfound teacher observed that "teachers have different ideas about what proficiency looks like in this building."

Adams 50 was different in that much of the evaluation involved crediting students' demonstrations of proficiency based on successful completion of worksheets and online problem sets. These demonstrations of mastery were then validated by Scantron tests that students completed quarterly. In other words, competency-based education in Adams 50 relied more heavily than in the other sites on standardized and machine-scored assessments as evidence that students had met the standards. This approach relied less on subjective evaluations of students' proficiency levels, and may therefore have yielded better adherence to shared standards of proficiency. However, a tension existed around the percentage of standards that students had to reach to be considered proficient in a given level—80 percent versus 100 percent. Eighty percent was the working figure in the 2012–2013 school year.

Adams 50 staff also acknowledged the tension inherent in holding common expectations, noting that a mastery grading system made it more difficult for a student to achieve top marks:

> A kid with a 3.0 who has made it through all those [proficiency] levels is more prepared for college than some kid who got a 4.0 for "doing school," but colleges don't see that. A 3.0 won't even open the door . . . For kids who need [scholarship] funds, it's a real challenge. They might go to a [high] school where [a 4.0] is easier to accomplish.

At the same time, the staff noted that parents of struggling students thanked them for making their students' areas of struggle more transparent, so that the school could better address and remediate those students. Staff members said that in a system where 18 percent of graduates enrolled in college and 86 percent of those needed remediation upon arrival, it was imperative to be more transparent with students, their parents, and their future colleges and employers about students' actual performance and skills.

An educator in DCIS-M, which was also emphasizing mastery grading, made a similar point: "If you pull our proficiency data, it mirrors our state assessment data. People ask what we're doing wrong, and we say, 'Well, actually, we're just being honest with our assessments.'"

Building a Sustainable Model

The third tension revolved around the question of how to make competency-based education sustainable for teachers and districts, in both a logistical and a financial sense. The sustainability question arose in all of the sites, though the nature of the sustainability challenges varied among sites.

To help schools capitalize on technology-based efficiencies, the Project Mastery initiative deliberately funded programs that included technological components, including LMSs as well as technology-rich curricula. In Adams 50, the efficiencies were intended to come from increasing the breadth of online instructional resources available to teachers and students, through the development of the Operation Spacewolf games and the purchase of Florida Virtual Schools units. The district also pursued new record-keeping efficiencies by providing upgrades to the Educate LMS, which allowed teachers to track students' learning across courses and years. To a similar end, the district was working with EffectiveSC on the LevelUp tool, which was intended to integrate student performance data from multiple sources rather than requiring that teachers input the data by hand.

In Asia Society sites, the efficiencies were intended to come from the ShowEvidence online platform, which was designed in part so that teachers could upload, search for, and download performance-based curriculum modules, including those developed by Asia Society, as well as those built by teachers across Asia Society's ISSN. In addition, it was designed so that students could upload their work and teachers could respond to students' work in terms of rubric-based criteria. In short, it was intended to create efficiencies both in curriculum development (by increasing ISSN teachers' access to quality curriculum) and by increasing the ease with which teachers tracked and evaluated student work.

In Philadelphia, the efficiencies were intended to come from the provision of a fully developed, project-based curriculum for ninth grade, which would enable teachers to deliver CCSS-aligned, project-based instruction without having to build their own units. They were also intended to come from the new LMS by Pearson, which functioned as a course-management system in which teachers could post assignments; students could upload and discuss their work, and teachers could provide feedback.

Technical Barriers to Efficiency

Across sites, staff described barriers in terms of how well the efficiencies materialized. Some of the challenges sites encountered were technical and pertained mainly to the LMS tools. In Adams 50, several staff members described the competency-based Educate LMS as unreliable—in that it did not auto-save and sometimes caused them to lose data—and inefficient—in that it did not have a way to identify missing assignments, to capture percentages instead of categorical proficiency scores, or to link assignments to a clear timeline of due dates. Staff noted that these problems with Educate predated the grant, and that the interface had improved during the two years of the pilot, which had paid for upgrades. Still, several teachers we interviewed reported that they kept their own Excel gradebooks in addition to the Educate records so they could maintain a more complete and reliable record. In fact, teachers' lobbying for an alternative to Educate led the high school to adopt Infinite Campus—the student information system that Adams 50 teachers already used for attendance data—rather than Educate as the default gradebook system for the 2013–2014 academic year. The key difference, project leaders explained, is that Infinite Campus evaluates students with respect to courses, and Educate tracks their accomplishment of proficiencies across courses and time. So project leaders interpreted the reversion to Infinite Campus as a step backward in the transition from traditional to competency-based education. With regard to the LevelUp tool, which would have combined student learning data from multiple online tools into a single interface, it was not funded as part of the grant effort and remained in development at the end of the pilot period.

Teachers in the Asia Society pilot sites received ShowEvidence professional development during the 2012–2013 academic year, and many said they had tried using the platform, but none said they had begun to use it extensively. The chief reason they gave was that developers were still refining it and fixing various bugs in the system. For instance, one teacher noted in a 2013 interview that not all of the rubrics had been uploaded when the teacher tried to use it in 2012, so a rubric from a different content area had to be used instead. Another teacher mentioned that the Common Core and Texas standards had been slow to be integrated, which made it difficult to credit students' work for meeting particular standards. Yet another teacher noted that the rubrics could not be customized in ShowEvidence, which made them impractical for many assignments. Another teacher stated that the time required to provide feedback on ShowEvidence was impractical for evaluating a stack of more than 30 long essays or projects. In the absence of a well-established platform, staff in the Asia Society sites generally reported that they were using Google Docs or EdModo (another online platform that a teacher described as "like Facebook") as repositories for student portfolios. Despite these concerns, some staff members expressed hope for the ShowEvidence platform, noting its attractive aesthetics and the possible benefits of having a student-learning repository that could follow students across courses. Newfound staff, though they reported having used ShowEvidence the least, said they were still enthusiastic about its ability to incorporate examples of student work from a variety of out-of-school learning experiences.

In the Philadelphia pilot, a key technical challenge lay in creating interoperability between the district's student information system, SchoolNet, which maintained students' longitudinal records and test scores, and the Project Mastery LMS, which was a course-management tool but allowed teachers to respond to student work and allowed students to track their own performance. Both products were developed by Pearson, which indicated that the two products could be made interoperable within the first year, though true interoperability was not attained during the study period. During the subsequent contract renewal process for SchoolNet, Pearson offered to provide a different LMS that would have greater interoperability with SchoolNet.[2]

Financial and Logistical Barriers to Efficiency

Other barriers to the anticipated efficiencies were financial and/or logistical. They included such issues as access to and ownership and development of materials. In Adams 50, the main barrier to improved efficiency through improved online course offerings appeared to be severe limitations in students' access to computers. At the outset of the study, project directors explained that the infusion of new online tools was supposed to give teachers options so that they did not have to just "point [students] to the file cabinet" for the next worksheet in the learning sequence. Both project leaders and teachers referred to the "file cabinet" as the go-to resource for providing diverse learning experiences to classes of as many as 35 students working at heterogeneous levels, so teachers were eager for more options. However, nearly every staff member we interviewed noted that students' limited computer access was a major barrier to using online tools for instructional purposes. One teacher pointed out that having "only one to two computers in class on a daily basis makes online quizzes a challenge," and another said, "the only challenge [to use of the mathematics games] has been a lack of access to technology." One teacher said she had received four iPads from the district and that students could take turns using these to play the Operation Spacewolf mathematics games. Beyond the small numbers of classroom computers and iPads, teachers reported having limited access to a computer lab or laptop carts. The project leaders explained that these resources are almost always tied up with assessments, and that the challenge would only become more severe as the district transitioned to the new Partnership for Assessment of Readiness for College and Careers (PARCC) assessments for the CCSS in the coming years. (Indeed, Adams 50 was the only Project Mastery site in which our student survey had to be administered via paper rather than online, due to students' limited access to computers.) In contrast, these resources appeared easily accessible in the other sites we observed—most of which were also in underresourced urban districts.[3] When asked about the limited resources, project leaders noted that hardware purchases were not eligible expenses under the Project Mastery grant. One teacher even noted that the schools' electrical systems might need updating to support extensive hardware additions. Despite the barriers to upgrading, it remained unclear why the Project Mastery pilot focused on the purchase of online instructional resources for a district in which students' access to computer was so severely limited.

In spite of these limitations, district leaders reported that teachers were referring students to online instructional tools, including a number of free sites teachers had found to be useful, such as IXL (which, as noted above, provided adaptive online teaching and assessments), Raz-Kids, and Khan Academy instructional videos. Teachers had also begun using the free, online

---

[2] Other technical challenges involved the LMS specifications for file formats that could be uploaded in bulk. We note this issue here to highlight the complexity of both technical and logistical issues that districts face in commissioning online LMSs.

[3] For instance, teachers in Sharpstown reported that the district was rolling out a one-to-one laptop program, and laptop carts were in use in every Philadelphia class we observed.

PMI instructional program developed by the New Jersey Center for Teaching and Learning, which included K–12 lessons and assessments. Given its popularity among teachers, its comprehensiveness, and its free online availability, Adams 50 had chosen to formally adopt that curriculum as its districtwide mathematics program starting in the next school year (2013–2014). Project leaders explained that students' use of these online resources was possible, in part, because of what they called a "BYOD model," which they defined to mean "Bring Your Own Device." In other words, students who brought their own devices to class could use them to gain access to online instructional resources for which they could earn academic credit. This presumably helped to make school-owned computer and tablet devices more available for students who could not bring their own. Still, the approach raises the question of how best to provide equal access to learning opportunities within a flexibly paced but resource-constrained instructional environment.

With regard to their own flipped classroom videos, an additional barrier that teachers reported was that some of their videos had to be remade between years as the district's standards were revised to better align with the CCSS. Nevertheless, teachers we interviewed remained enthusiastic about the idea of creating flipped-classroom videos, even if they reported that their time to do so was highly constrained by the demands of delivering highly differentiated instruction. In a competency-based system, said one teacher:

> You have an IEP [Individualized Education Plan] for every single child. . . and when you have 200 students, that is really hard . . . Everything has to be ready in your file cabinet before the year starts.

In short, teachers we interviewed said they were working hard to provide leveled instructional experiences to students working on diverse tasks daily, but that their ability to rely on technology to streamline the process was constrained by the small ratio of computers to students, as well as by the limited infrastructure for integrating data from all of the sites. As noted above, one teacher reported spending 40 percent of his work time on data entry, and project leaders conceded that the record-keeping demands of documenting individualized learning trajectories for each child were quite extensive. It was these demands that made them hopeful for the long-term potential of a data-integration system like LevelUp. As one Adams 50 interviewee explained, "All of this has been built on the backs of teachers, and they've done a remarkable job, but it's not sustainable. I'm not sure how much more you can squeeze out of teachers."

In the Asia Society sites and in Philadelphia, the sustainability challenge primarily involved the human and financial costs of delivering multiweek, project-based learning modules throughout the school year. In the Asia Society sites, the curriculum modules that were built with Project Mastery funds were intended to be exemplars that teachers could download and use, but teachers were generally building—and were expected to build—curriculum modules on their own. This involved aligning their performance tasks (i.e., assignments) with the Asia Society–generated performance outcomes for each subject and grade, some of which were built as part of Project Mastery, and some of which already existed. It also involved adapting the Asia Society–generated rubrics to use concrete language that students could understand and to reflect the components of particular tasks. The system appeared to work, in that teachers described or shared with the research team several innovative projects they had developed. (Teachers in the ISSN were expected to submit at least one curriculum module per year to the

Asia Society for review.) A teacher at Sharpstown, for instance, had developed a mathematics module that required students to build a model soccer stadium. One teacher at Newfound had developed a unit that required students to build a working model car, and another Newfound teacher had developed an art unit that required students to respond to—and demonstrate an understanding of—a pivotal moment in international history. Teachers reported that their challenges were learning to teach in this integrated, project-based way, and finding time to undertake a project-based approach, which involved extensive planning and compilation of materials. One math teacher who emphasized project-based learning said, "I've been working with GPS for three years, so I'm able to [design applied projects]. It would be hard for a first-year teacher who hasn't had too much GPS training to do something like that." Upon describing all that goes into developing a module—creating activities, labs, quizzes, and rubrics—another teacher said, "One of the main things is finding time to do this," and noted that he built one unit while the students were taking state accountability tests. When asked in a group interview about the biggest challenge to competency-based education, Newfound teachers responded: "Time! It takes more time to offer student choice than otherwise. Planning is more intensive. Assessment is more intensive." In DCIS, teachers reportedly received one release day per semester for curriculum development, plus occasional afterschool workshops. One teacher noted that the project-based learning approach was enjoyable but that teachers needed more time for curriculum building.

In the Philadelphia Educurious pilot, teachers were not building their own curricula but were using the units created by Educurious and providing feedback to the company on how to improve them. The challenge in Philadelphia was that, even though the district was instrumental in helping Educurious refine the units, it had rights to use them for only two years, and only in the pilot classrooms. Despite the fact that teachers and students reported liking the units, the district could not afford to roll them out more broadly. With Project Mastery dollars, Philadelphia had also commissioned the PhilWP to build a tenth-grade unit, which the district then owned, but even resources for that type of initiative were limited. (In contrast, Adams 50 retained the right to use the Intific mathematics games from inception forward in exchange for its role in helping Intific refine and pilot the games.) Project leaders explained that most ELA classes in the district were using old textbooks, and that there were no funds available for new curriculum adoptions. Consequently, the question of how to roll out and expand the Educurious pilot more broadly was really a question of financial resources, or of preparing teachers to build their own units, as the Asia Society teachers were expected to do. In other words, to scale the curriculum, the district would have to find a way to purchase it for a longer period of years and number of classrooms, or they would have to increase teachers' capacity and release time to build the curricula themselves.

Promoting Equity

A final tension that emerged across sites involved the question of who benefits most under competency-based models. The qualitative data that emerged from the sites did suggest that effects could be diverse. Educators at several sites suggested that a competency-based approach may disproportionately favor highly motivated learners. One teacher in Adams 50 said a competency-based system "works really well for the kids who want to move faster, the go-getters. For kids who get behind, it becomes a problem for them. They get stuck in their slow-moving pace." In a separate interview a year later, another Adams 50 teacher expressed a similar concern: "[Competency-based education] is really great for those who are motivated. For those who aren't, you have to give them a pace. Otherwise they move at their own pace, which is glacial." One other staff member at the school noted that the new system had "lit a fire" under students who were midlevel performers by taking away the safety net of grade inflation and social promotion. However, he noted that the "reluctant learners" who were three to four content levels behind were the students who had continued to maintain a "won't do" attitude. To counteract this equity problem, he explained that the high school employed a triage approach that placed the most struggling students in smaller classes with stronger teachers, while students who were ahead of the typical level for their age group received much more autonomy, in a more self-directed environment that was "basically study hall." In other words, staff at the school understood that students responded in diverse ways to flexible pacing and said measures were in place to create a safety net for struggling learners. As one Adams 50 teacher clarified:

> Some students take a year or two to understand that "now I'm in trouble" because "I don't' have a credit; I've gone at my own pace." We're trying to push them to see that it is the teacher's pace. It's negative because they are in that position, but it is positive when they realize it's their fault and they change their attitude.

We heard similar concerns about differential motivation patterns at DCIS-M, which, like Adams 50, served a highly disadvantaged student population with highly varied achievement levels. When encountering mastery-based grading, in which students "don't get credit for trying," teachers noted that students responded in one of two ways: "The kids who get it and understand the system, they're running with it. Kids who are willing to put in the extra work, they run with it. Kids who don't want to do work—it's a real struggle." Another teacher echoed that sentiment:

> One thing about the portfolio system . . . is there is a wider gap in grades: a lot of A's and a lot of "off-tracks," and not a lot of kids in the middle. I find that's because a lot of kids know they can submit work months later than they should, because they know they can turn it in at any time [and still receive credit].

Yet some of the same DCIS-M teachers also noted that the competency-based culture was improving students' attitudes toward school. They recounted watching a midyear transfer student throw away his notes from class, and his peers going silent with shock before explaining, "You don't do that here."

Educators at the other sites spoke less frequently about differential student engagement patterns, except to note that some students chose to take advantage of internships, ELOs, and afterschool badging programs, while others did not. The issues of equity may have arisen more frequently at Adams 50 and DCIS-M because these sites had taken the largest steps toward flexible pacing and mastery grading—both practices that affected entrenched patterns of social promotion and thereby threatened to reshape the distribution of student performance. However, just as educators in these sites registered concern about disparate performance trends among their students, they also acknowledged the importance of transparency for enhancing teachers' sense of urgency and students' motivation to learn.

CHAPTER FIVE

Students' Experiences in the Project Mastery Pilot Classes

In this chapter, we discuss students' self-reports from each site about their learning experiences in their Project Mastery classrooms. As discussed in Chapter Two, surveys were distributed in May 2013, near the end of study Year 2, to students in classrooms that the sites reported to be most closely affiliated with the Project Mastery pilot, or (in Asia Society sites) to competency-based approaches. Students were surveyed during class time with assistance from the classroom teachers and project leaders, though participation was optional. Parents received passive informed consent documents in advance of the survey administration, but none chose to opt their children out of the surveys. Students also had a choice of whether to participate. Data were collected in a deidentified form so students' anonymity was assured.

Table 5.1 presents data on the size of the populations surveyed and on the response rates from each site. Taking into account students who opted out of the survey, as well as students in target classes who never logged in (due to absence from school or lack of time provided in the classroom) and students who accepted the survey but did not answer a single question, overall student survey participation rates were 65.3 percent, with some variation across sites.

Respondent Characteristics by Site

Table 5.2 presents the characteristics of survey respondents by site. Most of the students in each site, with the exception of Newfound, identified themselves as racially nonwhite. In DCIS-M, Adams 50, and Sharpstown, most of the students were Hispanic; in Philadelphia, close to one-half were African-American. In DCIS-M and Sharpstown, more than half of

Table 5.1
Student Survey Response Rates

District or Site	Surveyed Population	Responses	Response Rate (Percentage)	Answered at Least One Question	Effective Response Rate (Percentage)
Adams 50	551	340	61.7	340	61.7
Newfound	207	181	87.4	146	70.5
Sharpstown	475	305	64.2	288	60.6
DCIS	151	111	73.5	105	69.5
DCIS-M	231	112	48.5	106	45.9
Philadelphia	528	431	81.6	414	78.4
Total	**2,143**	**1,480**	**69.1**	**1,399**	**65.3**

Table 5.2
Survey Respondent Demographics, by Site

Characteristic	Adams 50*	Newfound	Sharpstown	DCIS	DCIS-M	Philadelphia	Total
Female	0.47	0.39	0.47	0.50	0.46	0.41	**0.44**
Missing gender	0.07	0.30	0.12	0.12	0.07	0.11	**0.12**
White, not Hispanic	0.14	0.51	0.02	0.22	0.03	0.13	**0.15**
Hispanic[a]	0.55	0.08	0.64	0.43	0.71	0.12	**0.39**
African-American	0.07	0.03	0.08	0.06	0.11	0.44	**0.18**
Asian/Pacific Islander	0.09	0.03	0.05	0.05	0.02	0.09	**0.07**
Native American	0.07	0.02	0.06	0.02	0.04	0.03	**0.05**
Other	0.00	0.01	0.03	0.07	0.03	0.08	**0.04**
Missing race	0.06	0.30	0.12	0.14	0.07	0.12	**0.13**
Home language not English	0.37	0.04	0.58	0.23	0.60	0.14	**0.31**
Home language missing	0.06	0.28	0.12	0.11	0.07	0.10	**0.12**
Seventh grade	0.02	0.00	0.00	0.00	0.00	0.00	**0.005**
Eighth grade	0.33	0.00	0.00	0.00	0.00	0.00	**0.80**
Ninth grade	0.31	0.33	0.23	0.00	0.33	1.00	**0.48**
Tenth grade	0.17	0.26	0.22	0.00	0.59	0.00	**0.16**
11th grade	0.14	0.00	0.31	0.47	0.01	0.00	**0.13**
12th grade	0.01	0.00	0.17	0.48	0.00	0.00	**0.70**
Missing grade	0.01	0.41	0.06	0.05	0.07	0.00	**0.08**
Focus classes	Math	ELA	ELA	Passages and Portfolio	Social studies and American history	ELA	
N	340	146	288	105	106	414	**1,399**

NOTE: Adams 50 used levels instead of grades, but we asked students to tell us what "traditional" grade level they were in. These traditional grade levels are reported here. Note that the sample sizes refer to those who answered at least one question on the survey. However, the proportions in this table are calculated using total responses on the particular question under analysis.

[a] We define "Hispanic" as an additional category, even though it is not a race designation. All students who answered "white" to the race question and "yes" to being Hispanic are considered "Hispanic." Students who reported being "African-American" and "Hispanic" (37 students, or about 2.6 percent of the sample) were labeled "African-American." Students who answered "Other" or had missing race information but reported that they were "Hispanic" were assigned Hispanic status.

respondents reported that the primary language spoken in their homes was not English. This was reported to be the case for more than a third of respondents in Adams 50 and for nearly a quarter of respondents in DCIS as well.

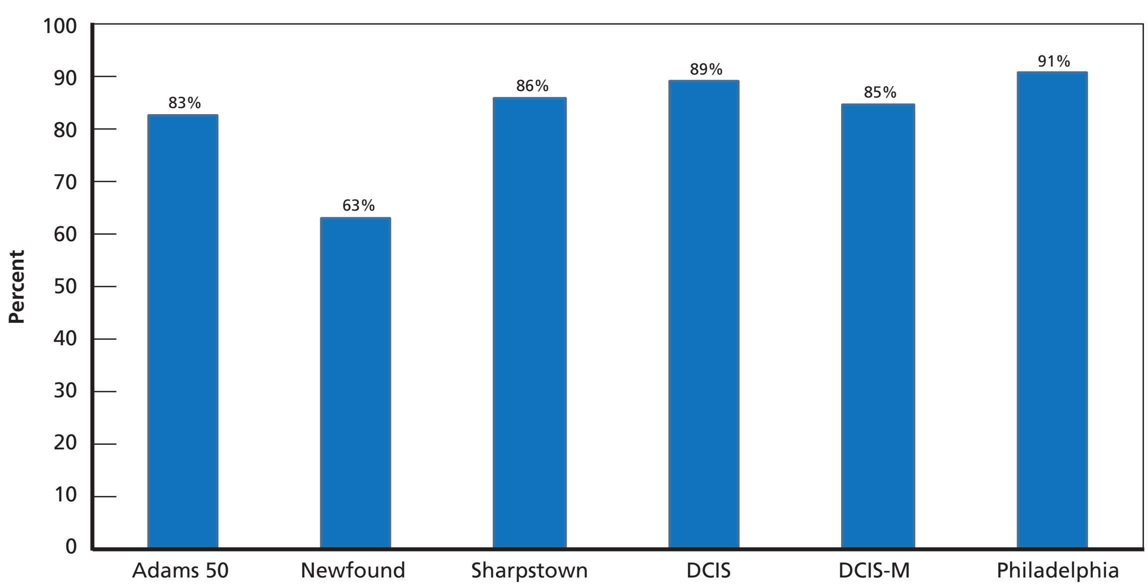
Because the survey focused on different courses and groups of students, the grade levels of student respondents varied by site. As Table 5.2 shows, a majority of respondents were ninth- and tenth-grade students, though a third of respondents in Adams 50 were eighth-graders, and all respondents in DCIS were 11th- or 12th-graders.

Student Perceptions About School Experiences: Common Items

In this section, we describe students' responses to survey questions that asked about several key features of or related to competency-based education, including students' opinions about the value of school, their enjoyment of and engagement with the classes in which they were surveyed, their experiences of choice in the class, their use of technology to track their own academic progress, and their indicators of anytime/anywhere learning. We chose to focus our analysis on these questions because they were common across sites, meaning that students in each site answered the exact same question. However, there were some questions that asked students to focus on one particular content area. In Adams 50, that was mathematics; in Philadelphia, Newfound, and Sharpstown, it was ELA.[1] In DCIS, it was the Passages and Portfolio classes that focused on 11th- and 12th-grade students' independent graduation projects, and in DCIS-M, it was a humanities-focused interdisciplinary class.

In this section, we report students' answers to each question by site. Because we are generalizing only to the populations participating in each site, Figures 5.1 through 5.6 present descriptive response statistics for each site but do not test these comparisons for statistical significance. However, when we disaggregate results by gender and race/ethnicity in Table 5.3 and in Figures 5.7 through 5.10, we do report on the significance of comparisons by gender and racial/ethnic categories, pooling respondents across sites. In these analyses, we give equal weight to each student, so that the pooled estimates reflect the composition of respondents in the sites, but our hypothesis tests adjust for the nesting of students within sites. To test differences by gender, we use chi-squared tests from a regression of each dependent variable on the

Figure 5.1
School Teaches Me Valuable Skills (% Who Agreed or Strongly Agreed)



RAND *RR732-5.1*

---

[1] Students in Newfound and Sharpstown were asked about other courses too, but we present findings only on ELA courses here for simplicity of exposition and because a competency-based approach was reported to be more prevalent in the ELA classes.

dichotomous gender indicator, with a random effect for program site (Raudenbush & Bryk, 2002). For race/ethnicity, we regress the dependent variable on a set of dichotomous racial/ethnic indicators and use a chi-squared test to assess whether the indicators are jointly significant at the 5-percent level, again including a random effect for program site. We report descriptive statistics and hypothesis tests by gender for all of the common items, and we display findings by race/ethnicity for the comparisons that showed statistically significant differences.

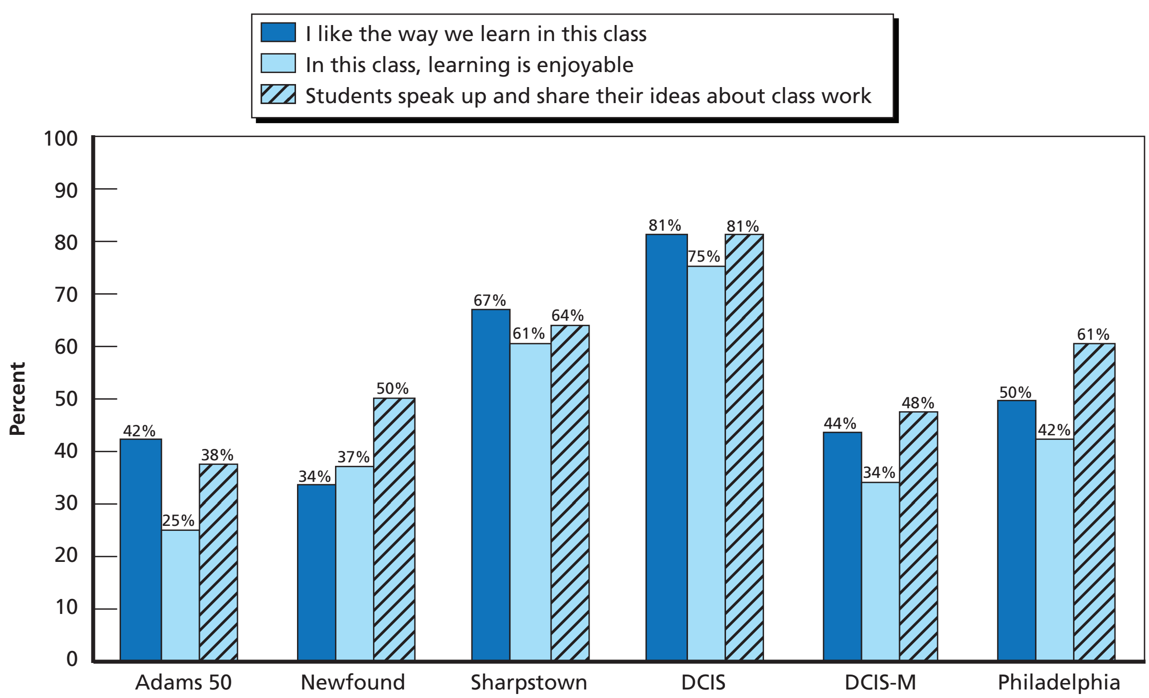
Value of School

With the exception of one site, there was a high degree of similarity across sites in terms of students' self-reported learning experiences. As shown in Figure 5.1, the majority of students at all sites reported that school taught them valuable skills, though the share agreeing with this statement was modestly smaller in Newfound, at 63 percent, than in the other sites, where it ranged between 83 and 91 percent.

Engagement/Enjoyment

In general, students reported that they enjoyed learning in their Project Mastery–related classes, although the sentiment was not overwhelming. As shown in Figure 5.2, about half of students reported that they liked the way they learned in their class and that learning was enjoyable, though the proportions were lowest in Newfound, Adams 50, and DCIS-M, and highest in Sharpstown and DCIS. In the latter two sites, more than half of students also reported that students spoke up and shared ideas about class work (see Figure 5.2).

Figure 5.2
Engagement/Enjoyment Indicators (% Who Said Statement Is Mostly True or Totally True)



RAND *RR732-5.2*

Student Choice

Students in all sites except one (DCIS) reported having limited choice inside the classroom to influence instruction. As Figure 5.3 shows, in four of the six sites, only one-third or fewer students reported having the opportunity to choose instructional materials (books, software, etc.) in their treatment class. Similarly, in five out of the six sites, one-third or fewer students reported having influence over how activities were done in the class. And fewer than one-third of students in five out of the six sites reported having opportunities to choose which topics to focus on in class.

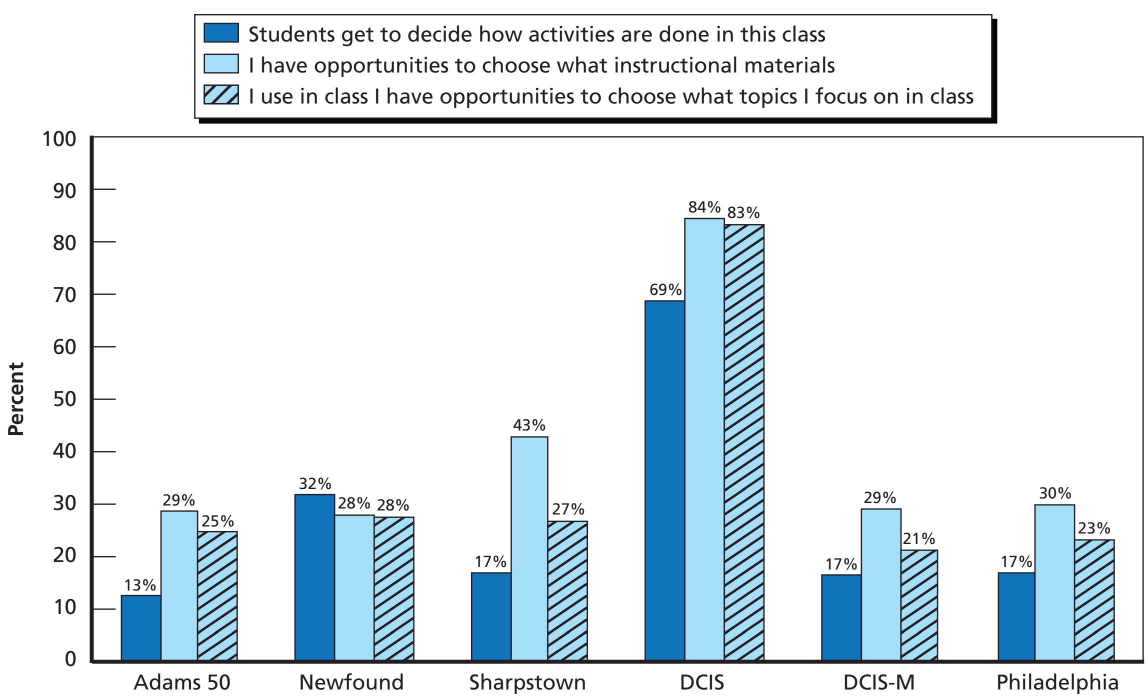
Flexible Pacing

About one-third or fewer students in five out of the six sites reported sometimes working on different topics or skills than their classmates were working on at the same time, as shown in Figure 5.4. In four sites, about 40 percent or fewer students reported being given the chance to work on class material or tasks at a faster or slower pace than other students in the class. Responses from these two questions suggest that most students perceived their progression to be similar to that of peers in the same class.

Use of Technology

In general, use of technology to track learning progress appeared to be far from universal at each of the sites. More than two-thirds of students in two out of the six sites (DCIS and Sharpstown), and more than one-half in two other sites (DCIS-M and Philadelphia) reported being

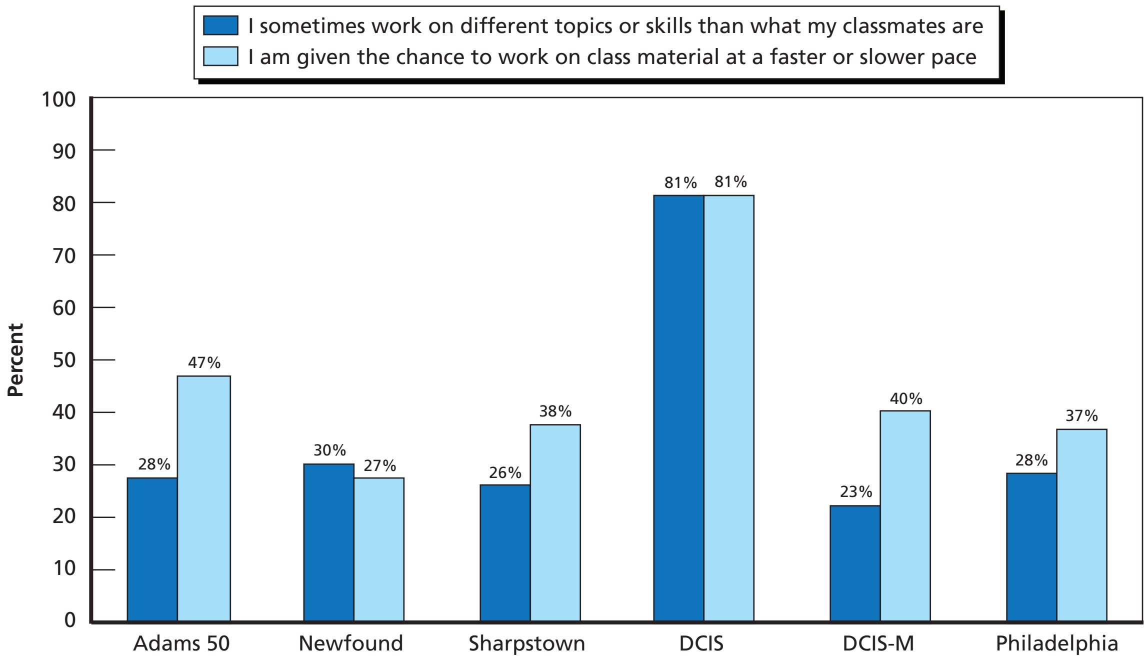
Figure 5.3
Opportunities for Choice in Class (% Reporting Mostly Yes or Always)*



*In the case of "students get to decide how activities are done" the reported frequencies refer to students answering "mostly true" or "totally true."

RAND *RR732-5.3*

Figure 5.4
Flexible Pacing Indicators (% Reporting Mostly Yes or Always)



RAND *RR732-5.4*

able to track their learning progress using technology—for instance, by using an LMS, an online gradebook, a portfolio, or other technology. In the remaining two sites, Adams 50 and Newfound, this proportion was below one-third, as shown in Figure 5.5. A second technology indicator concerned whether students could get help if they ran into trouble academically while using technology. In all sites except DCIS, fewer than 60 percent of students said they could.

Anytime/Anywhere Learning

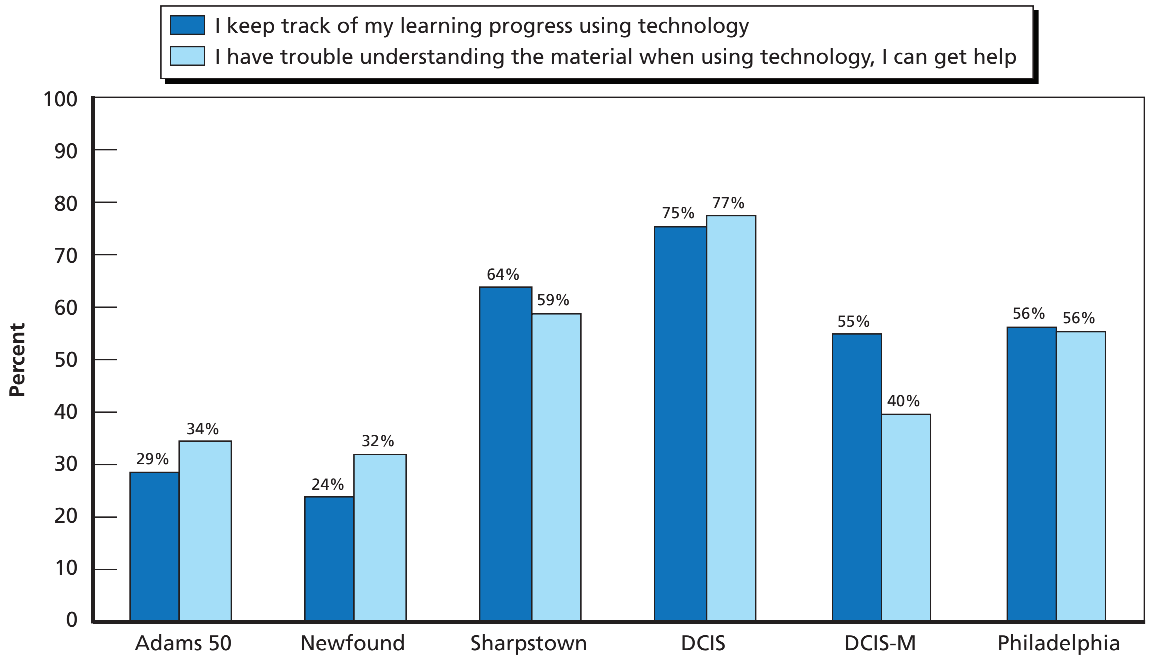
Lastly, a majority of students across all sites reported having access to learning materials outside of school. For example, at least 80 percent of students in every site reported having access to technology or the Internet outside of school whenever they needed it, as shown in Figure 5.6. And at least 74 percent of students in every site reported being able to access their class's course materials and assignments from home or other places.

Taken together, the findings suggest high satisfaction among students with their learning experiences in the surveyed classrooms, as well as strong access to technology and course materials outside of school. However, they may suggest less flexible pacing and student choice than we had anticipated, or simply less student understanding and identification of these practices.

Key Items by Race and Gender Breakdowns

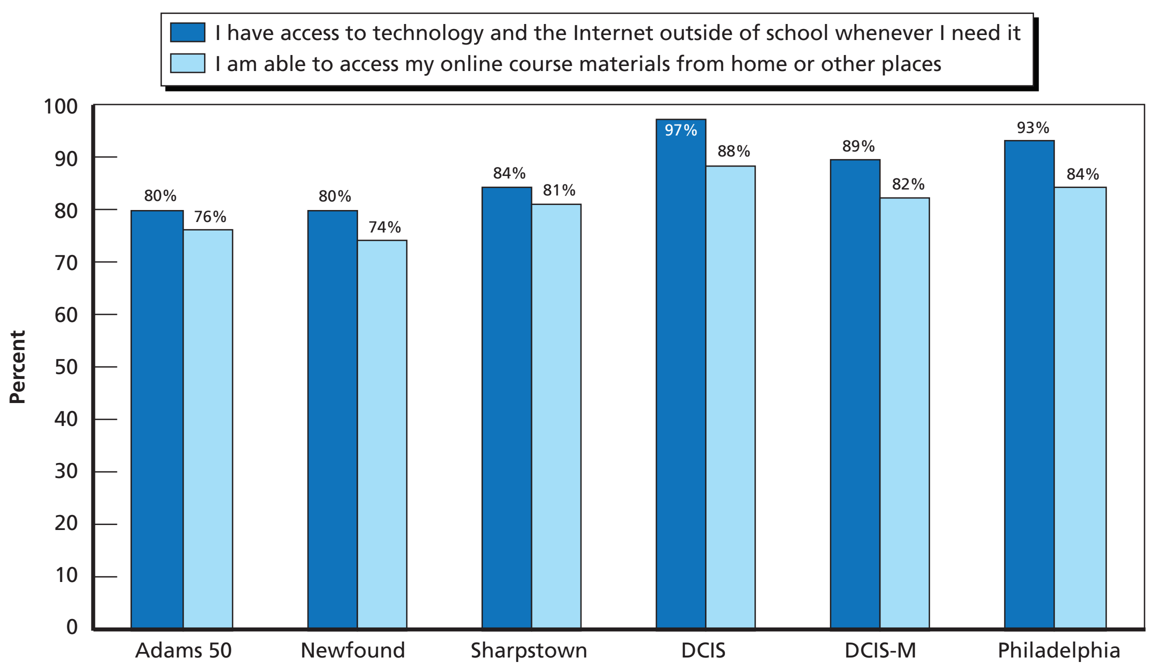
In general, male and female students had similar perceptions of their experiences in Project Mastery–related classes. Our disaggregation of responses by gender in Table 5.3 pools observations across sites. As noted previously, to test the statistical significance of differences by gender, we regressed each variable on a dichotomous gender indicator, adjusting for the nesting

Figure 5.5
Using Technology Indicators (% Reporting Mostly Yes or Always)



RAND *RR732-5.5*

Figure 5.6
Anytime/Anywhere Learning Indicators (% Who Agreed or Strongly Agreed)



RAND *RR732-5.6*

Table 5.3
Student Responses by Gender, All Sites (% Who Agreed or Strongly Agreed)

Statement	Female (n=656)	Male (n=638)	p-value[a]
School teaches me valuable skills	0.83	0.86	0.16
I am able to access my online course materials from home or other	0.82	0.80	0.31
I like the way we learn in this class	0.51	0.52	0.51
Students get to decide how activities are done in this class	0.21	0.22	0.34
In this class, learning is enjoyable	0.42	0.45	0.21
Students speak up and share their ideas about class work	0.59	0.52	0.01
I have opportunities to choose what instructional materials I use in class	0.34	0.39	0.04
I have opportunities to choose what topics I focus on in class	0.28	0.31	0.11
I sometimes work on different topics or skills than what my classmates are	0.32	0.32	0.82
I am given the chance to work on class material at a faster or slower pace	0.42	0.43	0.66
I keep track of my learning progress using technology	0.50	0.50	0.81
If I have trouble understanding the material when I'm using technology, I can get help	0.50	0.49	0.66
I have access to technology and the Internet outside of school whenever I need it	0.85	0.88	0.13

[a] p-values are based on a chi-squared test of the difference between proportions for males and females, adjusting for the nesting of respondents within sites. A p-value below 0.05 is considered statistically significant. significant.

of observations within site, and we report the associated p-values. The only statistically significant differences at the 5-percent level involved the ability to share ideas in class and the choice of materials used in class. Nearly 60 percent of females reported being able to speak up in class and share ideas, relative to only 52 percent of males; meanwhile, 39 percent of males said they had choices of materials to use in class, versus 34 percent of females. Males were slightly more likely to agree that school taught them valuable skills (86 versus 83 percent), that learning was enjoyable in the class (45 versus 42 percent), that they could choose topics in class (31 versus 28 percent), and that they had technology and Internet access outside of school (88 versus 85 percent), but these differences were not statistically significant.

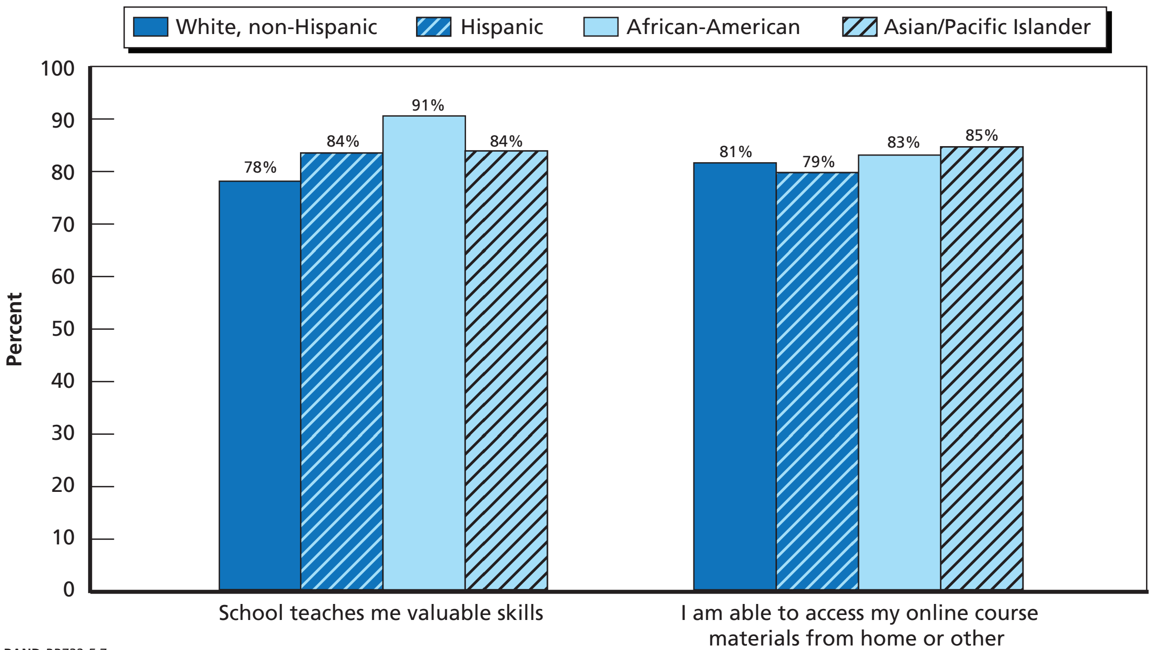
In terms of race/ethnicity, we also found some differences in student perceptions. Figures 5.7 through 5.9 highlight only those survey questions in which we found statistically significant differences at the 5-percent level by students' racial/ethnic category. As with gender, our analysis pooled students across sites, and our hypothesis tests adjusted for the nesting of students within sites.

African-American students were more likely than other groups to report that school taught them valuable skills. As shown in Figure 5.7, 91 percent agreed that it did. White students had the lowest rates of agreement with the statement, at 78 percent.

As Figure 5.7 also illustrates, Asian/Pacific Islander students were the most likely to report having access to online course material outside of school (85 percent), and Hispanic students were least likely to report such access (79 percent).

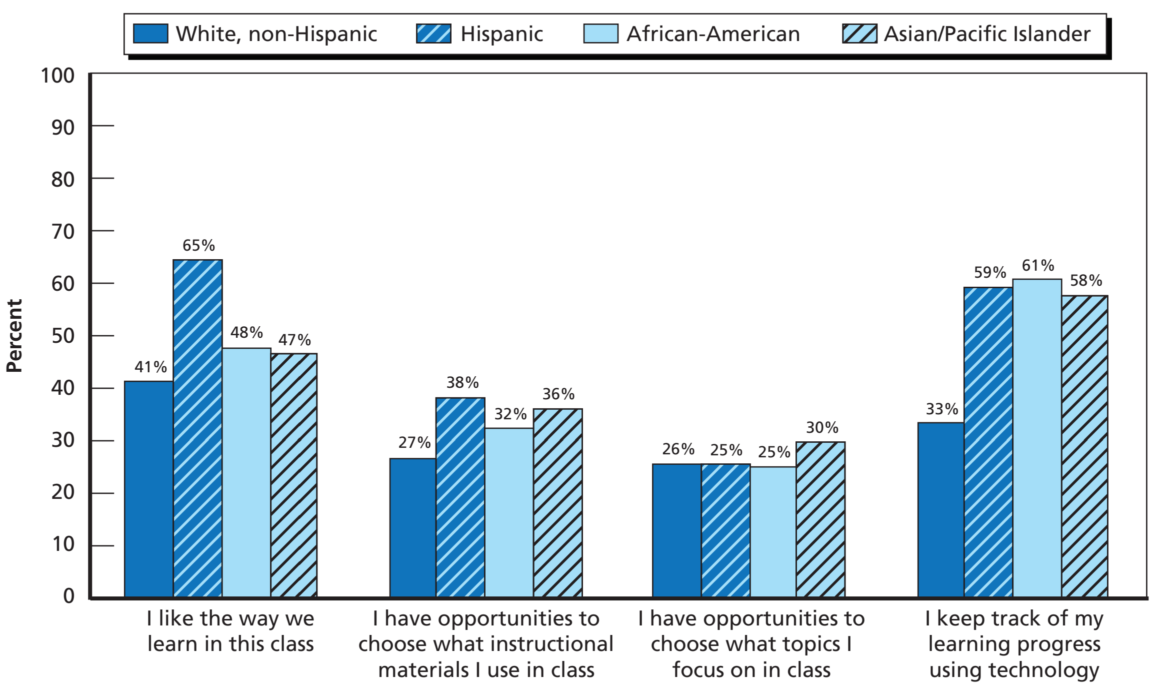
In terms of enjoyment of class work, flexible learning pace, and use of technology, we also found differences by race/ethnicity, as shown in Figure 5.8. Hispanic students were much more likely than all other racial/ethnic groups to report they liked the way they learned in their sur-

Figure 5.7
Student Opinions About Value of High School and Anytime/Anywhere Learning (% Mostly Yes or Always) by Race/Ethnicity, All Sites



RAND *RR732-5.7*

Figure 5.8
Student Opinions About Class Work, Flexible Learning Pace and Use of Technology (% Mostly Yes or Always) by Race/Ethnicity, All Sites



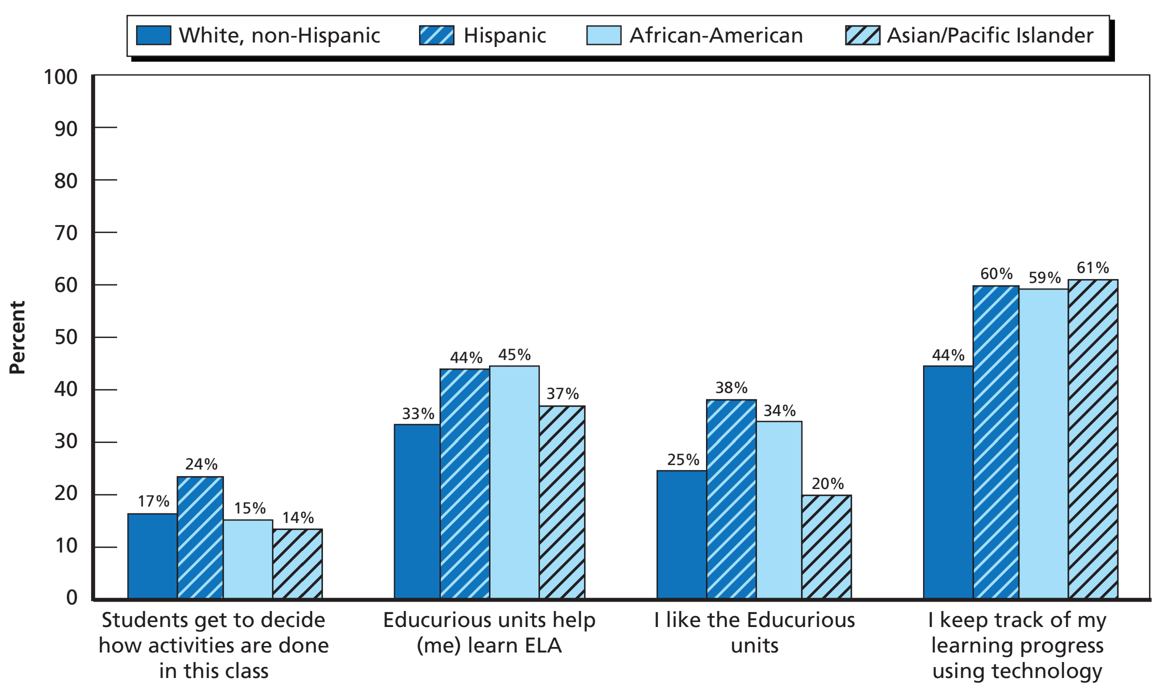
RAND *RR732-5.8*

veyed class, with 65 percent agreeing, as compared with 41 to 48 percent in the other groups.. Hispanic students were also slightly more likely than students in the other groups to report having opportunities for choice of materials. Asian/Pacific Islander students were slightly more likely than the other groups to say they had opportunities to choose topics in class, and white students were markedly less likely than the other groups to say that were able to track their progress using technology.

It is important to note that because these estimates pool across sites, the differences may be affected by differences in demographic compositions between sites. For instance, white students were overrepresented in the Newfound sample, and technology was not a central part of the competency-based reforms in that site.

To investigate site-specific differences in greater detail, we analyzed responses separately for Philadelphia and Adams 50, the two sites that implemented Project Mastery initiatives targeted to a specific subject area (ELA and mathematics, respectively) using well-defined interventions (Educurious units and Operation Spacewolf games). In the case of Philadelphia, Figure 5.9 shows that a higher proportion of African-American and Hispanic students (34 to 38 percent) reported liking the Educurious units, with substantially lower rates reported by Asian/Pacific Islander and white students (20 and 25 percent, respectively). In general, fewer than 50 percent of the students across racial/ethnic categories reported that the units helped them learn ELA, with higher proportions of Hispanic and African-American students reporting some benefit from the materials. White students were less likely than the other groups to report that they kept track of their progress using technology (at 44 percent versus about 60 percent). In terms of choice of how activities were done, Hispanic students were modestly

Figure 5.9
Responses with Statistically Significant Differences by Race/Ethnicity, Philadelphia Only (% Mostly Yes or Always)



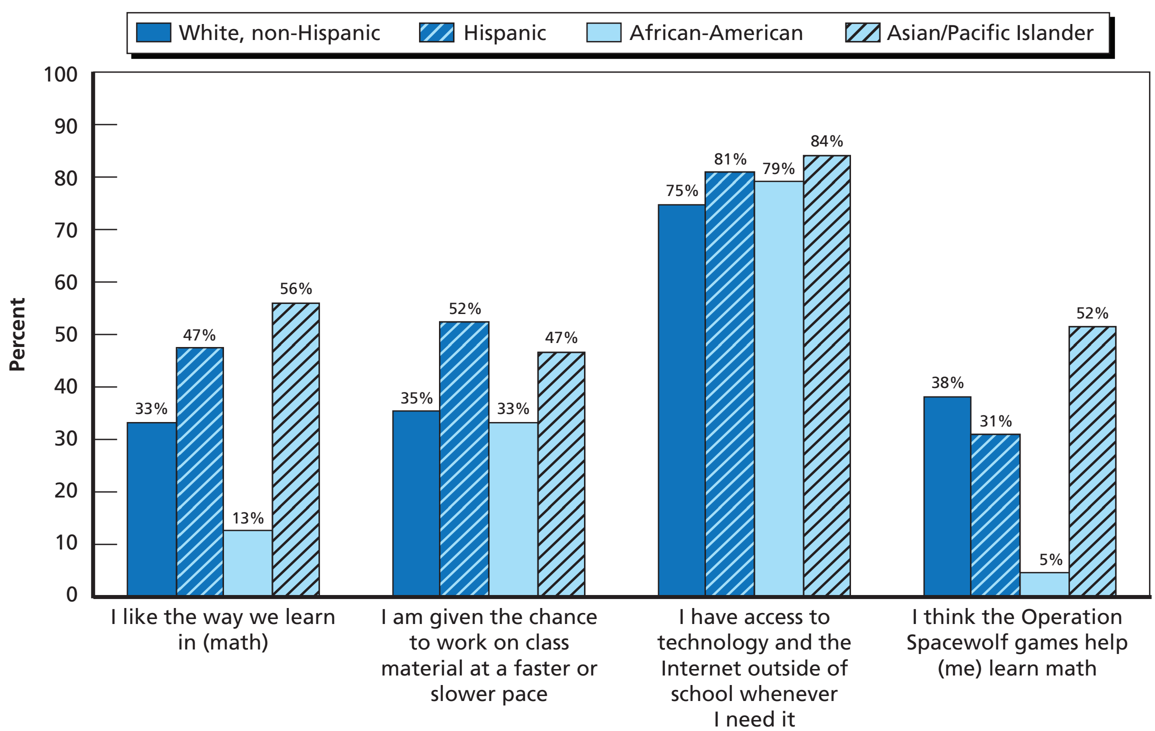
RAND *RR732-5.9*

more likely than students in the other groups to report having choices, at 24 percent versus 14 to 17 percent.

In Adams 50, Asian/Pacific Islander students were more likely than any other racial/ethnic group to report that they liked the Operation Spacewolf mathematics games, with 52 percent agreeing, as shown in Figure 5.10. A extremely low proportion of African-American students—only 5 percent—reported that the games helped them learn mathematics. Similarly, only 13 percent of African-American students expressed satisfaction with the way they were learning mathematics in their mathematics classes, as compared to 56 percent of Asian/Pacific Islander students. Though only 7 percent and 9 percent of respondents were African-American and Asian/Pacific Islander, respectively, these discrepancies are large enough to be of some concern in terms of which students were most responsive to the instructional approach and materials.

As Figure 5.10 also illustrates, Asian/Pacific Islander and Hispanic students were modestly more likely than the other groups to report that they could work at their own pace in their mathematics classes, with 47 to 52 percent agreeing, as compared with 33 to 35 percent of white and African American students. Between 75 percent and 84 percent of students said they had reliable access to technology and the Internet outside of school, with white students being least likely and Asian/Pacific Islander students being most likely to report access.

Figure 5.10
Responses with Statistically Significant Differences by Race/Ethnicity, Adams 50 Only (% Mostly Yes or Always)



RAND *RR732-5.10*

Summary

Because the programmatic details of the Project Mastery sites were heterogeneous, one striking finding from the survey results is the similarity of students' responses across sites to questions about their exposure to competency-based education components. Other than DCIS, where students reported especially high engagement and high exposure to core features of competency-based education, student responses at the other sites were more similar than we had expected. For example, other than DCIS, the rates of students reporting that they had choices of topics and materials were only slightly higher in the project-based learning sites than in Adams 50, where project-based learning received less emphasis. With regard to flexible pacing, the proportion of students reporting that they were often working on different material and at different paces than their classroom peers was nearly as high in the sites that emphasized synchronous learning as in Adams 50, which emphasized flexible pacing. The reasons for the similarities are not entirely clear. It is possible that students working in a flexible-pacing system were not always aware that their own tasks and rates of learning differed from those of their classroom peers, or it is possible that instructional procedures were more synchronous than our interviews and observations would have suggested.

Similarly, it is possible that students differed in what they considered to be examples of choice, or that even in project-based learning sites, choice was limited to a few large assignments and less so to everyday classwork. The outlying response patterns in DCIS are consistent with this interpretation, since the DCIS students were surveyed in Passages and Portfolio classes in which they were tasked with carrying out independent, largely self-directed projects that extended throughout the school year. In other words, the higher self-reports of competency-based education features and student engagement at DCIS are consistent with what we know about the activities of students in the surveyed classrooms.

In addition, surveyed students at DCIS tended to be among the oldest students in the sample, since the Adams 50 and Philadelphia initiatives focused on students in eighth or ninth grades. This means that DCIS students might have been given more learning autonomy than students in the other sites due to their age and years of experience in a competency-based system.

We found few meaningful differences by gender in terms of students' experiences with competency-based features in the sites. We did, however, find a few notable and significant differences by students' race/ethnicity, with African-American and Hispanic students reportedly liking Philadelphia's Educurious materials more than their white and Asian/Pacific Islander peers. In Adams 50, on the other hand, Asian/Pacific Islander students were more likely than Hispanic or white students to praise their mathematics classes and the Operation Spacewolf games, whereas African-American students were far less likely than other groups to express that they found benefit in the classes or games. These findings highlight the usefulness of disaggregating students' perception or performance data by demographic or performance-based risk categories in order to identify and engage students who seem to respond less enthusiastically to competency-based approaches.

CHAPTER SIX

Student Outcomes in the Project Mastery Sites

This chapter describes our findings on student outcome in each of the Project Mastery sites. As described in Chapter Two, we vary our analytic approach for each site due to marked variation in the scope and timing of each intervention. Each of the sections that follows focuses on one of the three Project Mastery initiatives: Adams 50, Asia Society, and Philadelphia. Each discussion of the outcomes is preceded by a detailed discussion of our particular data and methods for that analysis. We conclude the chapter with a cross-site summary of the outcome findings.

Adams 50

In Adams 50, we estimated the effects of competency-based education using a district-level synthetic comparison group (SCG) approach. District-level analysis is suitable for gauging the effects of the conversion to competency-based education in Adams 50 because the transition to competency-based education was districtwide. The SCG method was described first by Abadie & Gardeazabal (2003) in their study of the impact of terrorism in the Basque region of Spain and later by Abadie, Diamond, & Hainmueller (2007) in a study of the effect of a tobacco control program in California. The method, which is also being employed in the Empowering Effective Teachers evaluation supported by the Bill & Melinda Gates Foundation, is called for when there is a single treatment unit—often a district, state, or nation—and a range of possible comparison units, each of which differs from the treated unit in both observable and unobservable ways. The SCG approach compares the dependent variable of interest in the treatment unit to a weighted average of that variable in the comparison units.[1] The weights are designed to create a comparison unit whose preintervention time trend on a set of variables (including the dependent variable) mimics the preintervention time trend for the treatment unit as closely as possible. In other words, the SCG is designed to replicate the historical trend in the treatment unit with regard to the dependent variable of interest and a set of other covariates.

The procedure assigns weights, V, to each of the variables of interest (the dependent variable and available covariates), and it assigns weights W to each possible comparison unit. This process, which here is executed with the *synth* routine in Stata 12.0, iterates to find values for V and W that minimize the following expression:

$$\sqrt{(X_1 - WX_0)'V(X_1 - WX_0)}, \tag{1}$$

[1] The SCG does not estimate a parametric model. The treatment unit is simply compared to the synthetic comparison unit in each of the pretreatment and posttreatment periods.

where X_1 is a vector of treatment-group variables, and X_0 is a corresponding vector of variables for the comparison group. The control variables that comprise X in our Adams 50 analysis include baseline scores, racial composition, subsidized meal eligibility, and district size.

Adams 50, as described above, presented a useful case study because the entire K–12 district undertook a large-scale transformation from a traditional school district to a wholly competency-based district during the 2008–2009 academic year.[2] Given our concern that Adams 50's pilot program intervention (the development of a small number of mathematics games and flipped-classroom videos) offered a fairly low-dose intervention unlikely to yield a student achievement impact, we focused this analysis instead on examining district-level student outcomes of the district-level intervention (which included the pilot program intervention) during the years before and after the competency-based reform. District-level data were publicly available for three years prior to the reform and five years subsequently, meaning that we can examine district-level performance in terms of student scale scores starting in 2005–2006 and extending through the 2011–2013 academic years. We chose mathematics as the focal subject because we knew from the Project Mastery pilot that mathematics has been a focal area in the districtwide move to competency-based education. Additionally, focusing on mathematics enabled us to capture any effects of the pilot intervention, which also focused on that subject. We use district-aggregate scale scores as reported by the state because they are publicly available. We prefer mean scale scores to proficiency rates because the former incorporate information about student achievement throughout the performance spectrum rather than dichotomizing all students with respect to a performance threshold (Betebenner & Linn, 2010). (In three other sites, however, we use proficiency rates because aggregate scale scores are not publicly available.)

Our analysis compared mathematics performance trends in the Adams 50 district to an estimate of what might have happened in the district had the competency-based reform not gone into effect. To do this, we created a synthetic comparison district that is actually a composite of other districts in the state, weighted to exactly reflect student performance trends in the treatment district during the years before competency-based education commenced. The SCG is also weighted to maximize balance with the treatment district in terms of district size and demographic characteristics. Table 6.1 presents the district-level variables (outcome as well as predictor variables) used to create the synthetic comparison group, as well as the extent to which the synthetic and treated groups are similar (averaged across years). We find almost perfect balance on all variables other than percent black (a difference of about 1.17 percentage points), percent English language learners (a difference of 8.82 percentage points), and percent migrant (a difference of about 3.19 percentage points).[3] Standardized math scores are perfectly matched. We note that among this set of covariates, the algorithm assigned the greatest weight (0.912), by far, to baseline scores in its effort to minimize preintervention differences in test scores in each preintervention year. The combined weight of the other terms added to only 0.088. This means that the other variables received far less weight than mathematics scores in the construction of the SCG.

[2] To the best of our knowledge, no other district in Colorado had undertaken a districtwide competency-based reform on a timeline corresponding to that of Adams 50. If other districts had done so, they would conservatively bias our estimate of the effect of competency-based reform toward zero.

[3] Racial/ethnic designations (e.g., black versus African-American) are based on the variable names in the data sets used.

Table 6.1
Balance Between Adams 50 and Synthetic Comparison District on Key Indicators in the Pretreatment Years

Variables	Treated	Synthetic	V-Matrix Weight
Standardized math score	–1.89	–1.89	0.912
District enrollment	10,506.33	10,452.92	0.006
Median growth percentile	46.33	46.29	0.019
% Advanced	7.83	7.42	0.001
% Proficient	23.97	24.07	0.011
% Partially Proficient	35.63	35.55	0.011
% Unsatisfactory	31.07	31.66	0.016
% Hispanic	63.58	63.01	0.001
% Black	1.93	3.10	0.001
% With disabilities	9.81	9.41	0.002
% Gifted	6.27	6.14	0.020
% English language learner	36.48	27.66	0.000
% Migrant	1.81	5.00	0.000

Note that mathematics scores are standardized to have mean zero and standard deviation one at the district level. A district-level standard deviation corresponds to about 20 scale score points, which is just over a quarter of a student-level standard deviation, given that a student-level standard deviation is roughly 77 scale score points. In each year, we are comparing the treatment district to the districts in the composite. Therefore, we are not concerned with statewide differences in scale score points over time, which may be affected by changes over time in test difficulty, test scaling, or the grade-level distribution of students. These differences are factored out because we examine how students in the treatment districts score in comparison to students in the comparison districts in the same year.

Our results are displayed in Table 6.2 and in Figure 6.1. Both illustrate the average mathematics scores (in district-level standard deviation units) in the treatment district and its synthetic comparison district during the three years before the competency-based intervention and in the four years after the intervention went into effect. Table 6.2 also shows the year-by-year differences in these averages between the treatment and comparison groups. These differences are depicted graphically in Figure 6.1. What we find is that the SCG procedure is able to perfectly match the treatment district to its synthetic comparison during the three preintervention years, with only minuscule between-group differences, ranging from 0 to –0.005. In contrast, we find sizable differences in the postintervention years, ranging from about –0.4 during the first competency-based intervention year to –0.9 the second year. The difference then declines to about –0.8 in subsequent years.

In substantive terms, these effects are sizable. Converting these district-level effects to the sale of student-level standard deviation units, we estimate that average student performance in 2011 through 2013 was roughly a fifth of a student-level standard deviation lower in Adams 50 than we would have predicted based on the district's prior performance trajectory and demographic characteristics. For example, a student in a district similar to Adams 50 who would

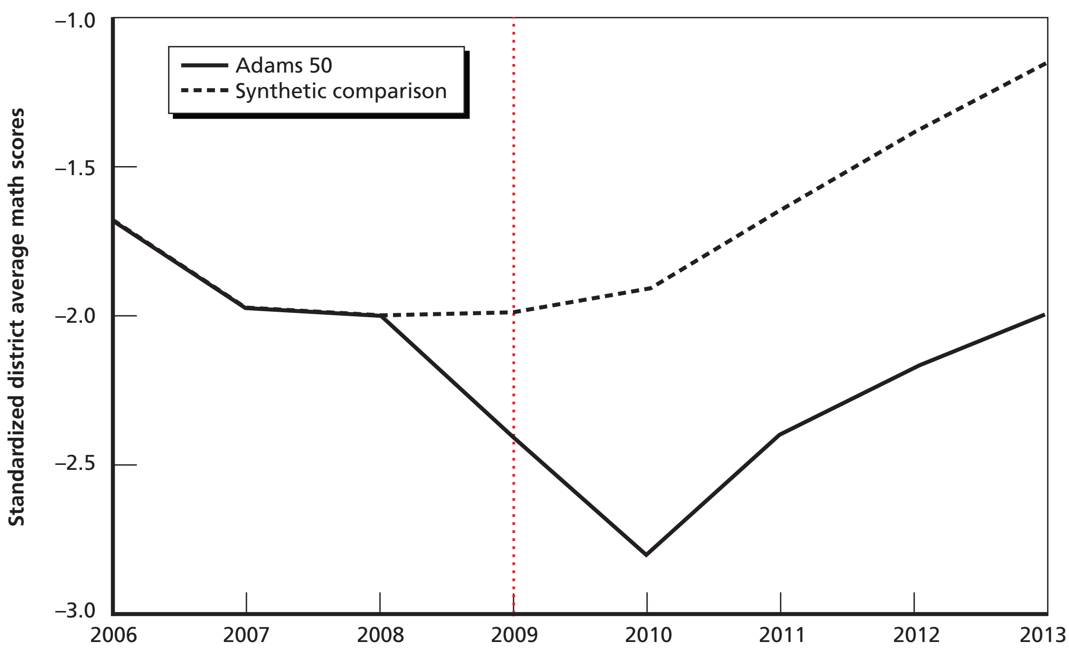
Table 6.2
Standardized Mathematics Test Score Trends for Adams 50 and Its Synthetic Comparison District

Pre/Post Statewide Shift to Competency	Spring Year	Treated	Synthetic	Difference
Pre	2006	–1.692	–1.687	–0.005
Pre	2007	–1.975	–1.973	–0.003
Pre	2008	–2.006	–2.006	0.000
Post	2009	–2.414	–1.993	–0.421
Post	2010	–2.812	–1.915	–0.897
Post	2011	–2.405	–1.652	–0.753
Post	2012	–2.184	–1.391	–0.794
Post	2013	–2.003	–1.156	–0.846

have scored better than 50 percent of the students in the state would have—if in Adams 50—scored better than 41 percent of the students in the state.[4]

We should also note that the district started out already more than 1.5 district standard deviations below the state mean achievement, as did its weighted comparison group. According to our interviews with the district, it was exactly that underperformance that compelled district leaders to embrace a new, competency-based approach. The program involved an extensive

Figure 6.1
Synthetic Comparison of Adams 50 to Other Districts in the State in Mathematics



NOTE: Y-axis is scaled in district-level standard deviation units.
RAND *RR732-6.1*

[4] Percentile conversions assume that student-level scores are normally distributed. They are based on a one-tailed comparison using the z-score calculator available online at www.zscorecalculator.com.

overhaul of teaching and learning, and, with a dearth of competency-based curricula and data systems at the outset, it required an adjustment period. Given evidence from management literature suggesting that performance often dips for a short period after a major organizational change (Herold & Fedor, 2008; Jellison, 2006), it is possible that the district was beginning to emerge from that implementation dip. The positive slope in performance since 2010 indicates that mathematics achievement in the district was improving relative to the state average—a pattern consistent with a rebound from an implementation dip. Still, as of 2012–2013, the gap between Adams 50 and its synthetic comparison district appeared to have stabilized at about 0.8 of a district standard deviation. In other words, the district was still underperforming, by a large margin, from the level we would have predicted based on its test scores prior to competency-based reform.

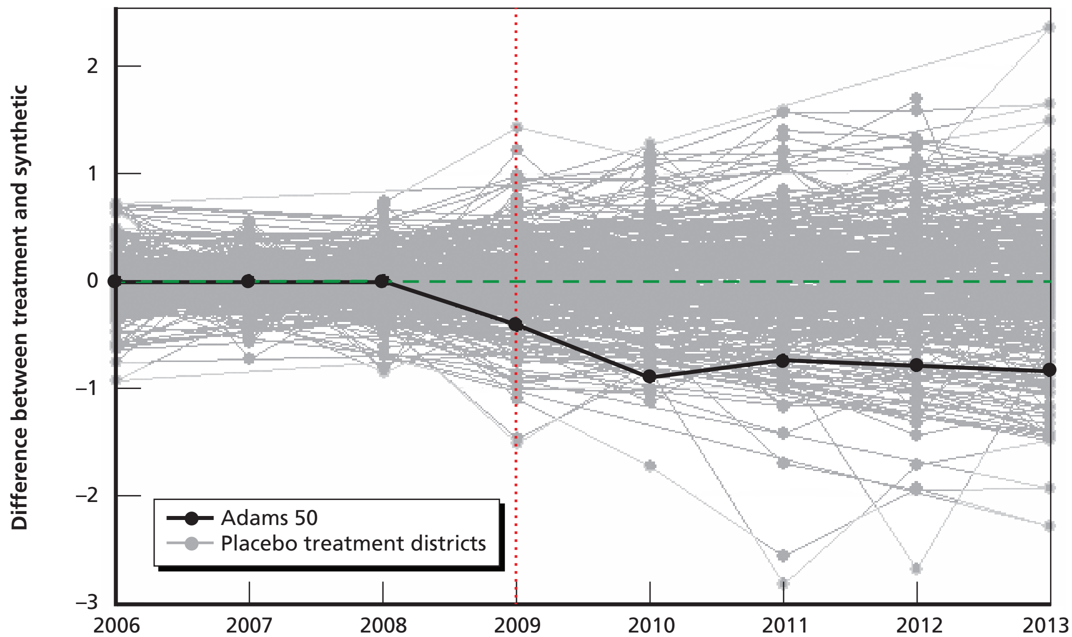
An important consideration is whether this dip in performance might have occurred by chance. With only a single treatment unit, our statistical power is inherently minuscule, and we are not able to use traditional hypothesis testing approaches with this method. However, one way to roughly gauge the extent to which the trend may have occurred by chance is through a placebo test (Abadie et al., 2007; Bertrand, Duflo, & Mullainathan, 2004), in which we apply the SCG method to each of the *untreated* districts and examine how unusual our results for the Adams 50 district are in comparison to running the same test on all the untreated districts. That is, we iterate across each district in the SCG, rerunning the analysis as if that district were the treatment unit. Because the other districts are not randomly drawn, we cannot actually assume that they conform to a normal distribution, and thus we cannot technically apply inferential statistics to the performance of Adams 50's district relative to the others. Still, we can use roughly the same benchmarks we would with a normal distribution to get a sense of how likely it is that the pattern we observe occurred by chance. We present the results of the placebo tests in Table 6.3 and in Figure 6.2. The figure displays gray lines for placebo tests from each of the 177 nontreated districts, so the gray lines can be viewed as effects estimated by chance. Their true mean is zero—meaning no treatment effect—but due

Table 6.3
Difference Between Treatment and Synthetic Comparison Unit for Adams 50 and 177 Placebo Treatment Districts

		Placebo Treatment Districts				
Year	Adams 50	Min	10th Percentile	Median	90th Percentile	Max
2006	0.0	–0.9	–0.3	0.0	0.3	0.7
2007	0.0	–0.7	–0.2	0.0	0.3	0.5
2008	0.0	–0.9	–0.4	0.0	0.3	0.7
2009	–0.4	–1.5	–0.6	0.0	0.6	1.4
2010	**–0.9**	–1.7	–0.7	0.0	0.6	1.3
2011	**–0.8**	–2.8	–0.7	0.1	0.6	1.6
2012	–0.8	–2.7	–0.9	0.0	0.7	1.7

NOTE: Boldface indicates the Adams 50 estimates fall below the 10th percentile of placebo tests.

Figure 6.2
Placebo Test for Adams 50 Synthetic Comparison Group Analysis



RAND *RR732-6.2*

to sampling error, they sometimes yield large effects by chance alone.[5] The dotted horizontal line represents the scenario of no treatment effect—and, as expected, the placebo lines are about evenly distributed above and below that line. The solid black line represents the Adams 50 district, which does appear near the lower extremes of the set of placebo tests during the treatment years. In other words, the magnitude of the effect is large enough that it is unlikely to have occurred by chance alone.

Table 6.3 shows the empirical distribution of the placebo tests in each year. Here we find that the Adams 50 district showed a difference that fell below the tenth percentile of placebo tests in 2010 and 2011. In fact, the district fell below the fifth percentile in 2010, which is akin to a statistically significant difference at the 10-percent level with a two-tailed hypothesis test. It would be difficult to attribute effects of this magnitude and improbability to chance alone, especially given that with one treatment unit, we have statistical power of less than 10 percent to detect an effect as large as one standard deviation.

---

[5] We use the default iteration method in the *synth* package for the placebo tests due to the intensive memory demands of the more-comprehensive "nested" option, though we continue to report on the nested results for the true Adams 50 comparison. The nested option generally results in a better match in the preintervention years because the iteration procedures are less restricted. If anything, using the more-restricted iteration procedure in the placebo tests makes our pseudo-hypothesis test more conservative because placebo tests with the nesting option would likely yield a tighter and less-dispersed estimate cloud. The same approach applies to the placebo tests in our Newfound SCG in the next subsection as well.

Asia Society

As described above, the Asia Society pilots focused on a partnership between a competency-oriented intermediary organization and four schools that shared a competency-based philosophy, with a particular focus on project-based learning and rubric assessments. Newfound was a comprehensive high school in a small-town setting, where it was the only high school in the district. Sharpstown was a school of choice founded within the Houston Independent School District—a large, urban district—in the 2009–2010 academic year. DCIS and DCIS-M were small schools of choice operating within another large, urban district—the Denver Public Schools. Because the three districts—Newfound, Houston, and Denver—were located in different states, the public-use data sets differed for each analysis, and we made use of readily available variables and years in each state.

We take different analytic approaches for Newfound as opposed to DCIS, DCIS-M, and Sharpstown because the curriculum and technology materials that Asia Society developed with the grant funds had not been fully rolled out in classrooms during the 2011–2012 and 2012–2013 academic years. As with Adams 50, we are primarily interested in outcomes corresponding to the sites' competency-based education features, which are summarized in Table 3.2. In Newfound and DCIS-M, competency-based education focused on student choice in the form of project-based learning, and on evaluating students for proficiency rather than effort. In DCIS and Sharpstown, it focused mainly on student choice in the form of project-based learning.

Because Sharpstown, DCIS, and DCIS-M had maintained these foci since their founding, we were not able to establish a preintervention period for them, and we therefore simply compared their performance to their respective state averages and to demographically similar schools in their respective states. For Newfound, however, we examined performance before and after the school's transition to competency-based education, relative to similar high schools in the state during the same pre/post comparison periods. The posttreatment difference from similar schools is interesting because Newfound was reported to have immediately embraced the state's move toward competency-based education to a much greater extent than other schools. It did this by assessing students on a 0–4 competency-based scale rather than an A–F grading system, and evaluating students based not on teacher-made tests but on performance assessments in which students applied what they were learning to real-world problems and were scored by rubrics. This was especially true in high school language arts, so we focus our Newfound analysis on grade 11 reading test scores.

Newfound Regional High School: Synthetic Comparison Group Analysis

To examine the effect of the state competency-based reform shift in Newfound relative to other schools in the same state, we undertook an SCG analysis, but this time at the school level rather than the district level, focusing only on other high schools in the state. We were able to match on only two prior years of student test performance due to data availability constraints; we view this analysis as more descriptive than causal, given the likely relevant unobserved differences between Newfound and other schools in the state that led it to an especially strong embrace of the state reform.

Table 6.4 displays the variables used to construct the SCG. Here, the standardized reading score itself received a weight of 0.21 (much smaller than the outcome variable in the Adams 50 comparison), but the share of students scoring substantially below proficient—a very low

Table 6.4
Balance Between Newfound and Synthetic Comparison Schools on Key Indicators in the Pretreatment Years

Variables	Treated	Synthetic	V-Matrix Weight
Standardized reading score	−0.17	−0.17	0.209
% With disabilities	14.00	15.44	0.000
% Proficient or advanced	57.50	55.57	0.001
% Partially proficient	23.00	23.02	0.050
% Substantially below proficient	6.00	6.01	0.576
School enrollment	468.50	465.96	0.084
% Free/reduced-price lunch	27.42	27.42	0.058
% Hispanic	0.32	0.85	0.000
% African-American	0.00	0.89	0.000
% White	97.66	97.50	0.020

6 percent—receives a weight of 0.58. Balance is very high across all predictors. The process also yields a synthetic comparison group that is nearly identical to Newfound in the two precompetency years. In Table 6.5 and Figure 6.3, we see that Newfound and its synthetic comparison actually scored 0.4 of a school-level standard deviation below the state average in 2008, and very close to the state average in 2009, where the state average is denoted as zero on the y-axis.

During 2010, the competency-based implementation year, the performance of Newfound and its SCG fell with respect to the state average. However, from that year forward, Newfound markedly outperformed its synthetic comparison. Its outperformance increased over time, so that by 2012, Newfound was also outperforming the state as a whole by about 0.4 of a school-level standard deviation. This gap in 2012 represents about a tenth of a student-level standard deviation, or roughly 4 percentile points within a normal distribution.[6] This is a modest but nontrivial positive effect.

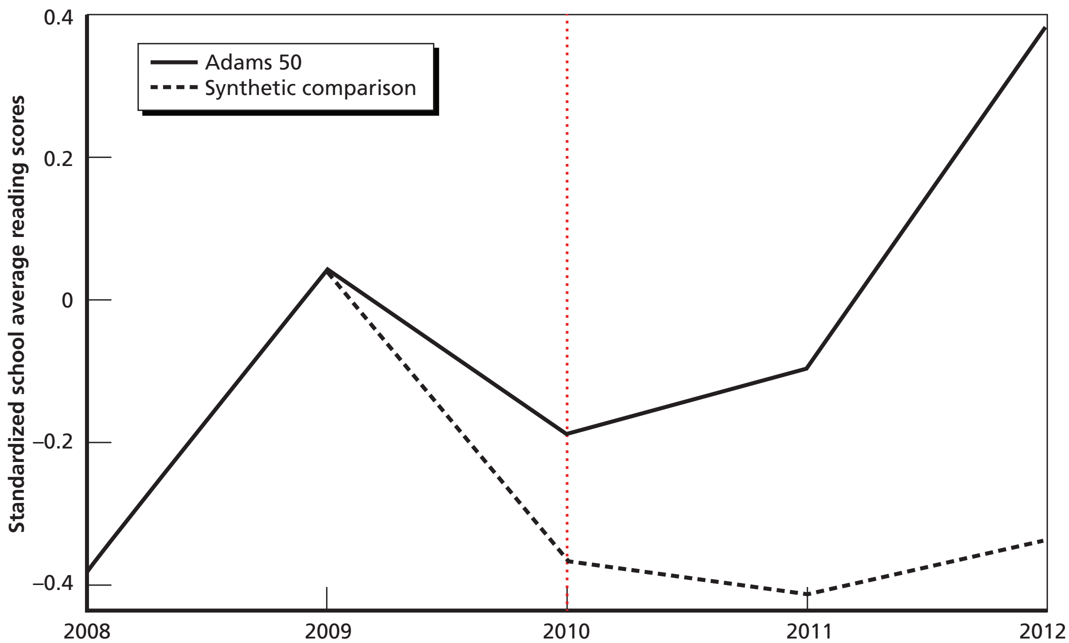
Of course, as was true for Adams 50, we are concerned about whether such a difference might have occurred by chance, since we are focused on a comparison of only a single school to the rest of the state. As with Adams 50, we undertake a placebo test to roughly estimate how likely such an effect would have been by chance. Results of the test are presented in Figure 6.4, where we see that the black line representing Newfound falls well within the cloud of gray lines

Table 6.5
Standardized Reading Test Score Trends for Newfound and Its Synthetic Comparison Schools

Pre/Post Statewide Shift to Competency	Spring Year	Treated	Synthetic	Difference
Pre	2008	−0.384	−0.382	−0.002
Pre	2009	0.041	0.043	−0.002
Post	2010	−0.190	−0.367	0.176
Post	2011	−0.099	−0.411	0.312
Post	2012	0.384	−0.339	0.723

6 This calculation assumes, based on other sites, that the school-level standard deviation is about a fourth of the student-level standard deviation. We were not able to locate the student-level standard deviation for this test.

Figure 6.3
Synthetic Comparison of Newfound in Terms of 11th-Grade Reading Scores



NOTE: Y-axis is scaled in school-level standard deviation units.
RAND *RR732-6.3*

representing the placebo test—except in the final year, 2012, where it approaches the outside of the cloud. As before, a finding that is near the extremes of the cloud is considered less likely to have occurred by chance. Turning to Table 6.6, we find that in no years did the differences between Newfound and its synthetic comparison exceed the 90th percentile among the set of placebo treatment groups. In other words, it is within the range of estimates that we may have expected to see by chance alone and should therefore be interpreted with caution.

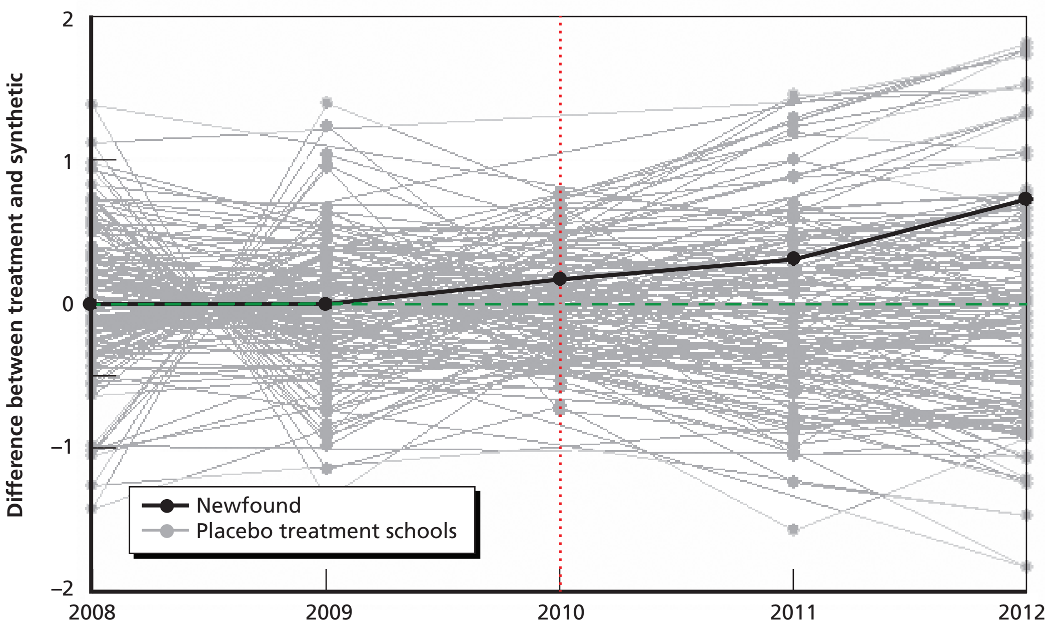
Sharpstown: School-Level Regression Analysis

For the other three Asia Society pilot sites, we did not have a clear preintervention period. These were relatively new schools that had implemented the project-based learning and performance assessment aspects of competency-based education since their inception. Without a pre-competency period from which to construct a synthetic comparison group, we were able simply to compare each school's test scores to schools in its respective state, adjusting for school size, school demographics, and year fixed effects. The regression model was specified as follows:

$$y_{sgt} = \alpha_0 + \alpha_1 z_s + \delta_1' X_{sgt} + \eta_1' T_t + u_s + \varepsilon_{sgt} \,, \qquad (2)$$

where y_{sgt} is the dependent variable of interest for grade g in school s and year t; and z is a time-invariant school-level treatment indicator denoting the Project Mastery school, because we do not observe Sharpstown (or DCIS and DCIS-M) before competency-based model inception. The parameter of interest is α_1, which represents the average difference in y_{sgt} between the Project Mastery and other schools, adjusting for the other terms in the model. Vector X_{sgt} represents a set of time-varying school or school-by-grade demographic characteristics in each year, with

Figure 6.4
Placebo Test for Newfound Synthetic Comparison Group Analysis



RAND *RR732-6.4*

effects given by parameter vector δ_1. T_t is a vector of school-year dummy indicators with fixed effects given by parameter vector η_1. Parameter u_s represents a school-level error term, and e_{sgt} represents a school-grade-year error term, where both errors are assumed to be normally distributed with mean 0 and standard deviation σ.

What is important to bear in mind about this analysis is that, because we cannot adjust for preintervention trends in performance, it is even more subject to selection bias than the SCG analyses presented above. In other words, it is likely that the schools and their students differ in ways that the predictors in our ordinary least squares regression analysis do not fully capture and that are linked to students' test score performance. If, for example, the more-motivated parents of any given race/ethnicity or socioeconomic status choose to send their children to Sharpstown instead of the default school in their district, then adjusting for the percentages

Table 6.6
Difference Between Treatment and Synthetic Comparison Unit for Newfound and 78 Placebo Treatment High Schools

		Placebo Treatment High Schools				
Year	Newfound	Min	10th percentile	Median	90th percentile	Max
2008	0.0	–1.4	–0.6	0.0	0.7	1.4
2009	0.0	–1.4	–0.8	0.0	0.6	1.4
2010	0.2	–0.7	–0.5	0.0	0.5	0.8
2011	0.3	–1.6	–0.9	-0.1	0.9	1.4
2012	0.7	–1.8	–0.9	0.0	1.0	1.8

of students of each race and socioeconomic status in the school will not remove those differences in motivation. With this caveat in mind, we turn to the results.

Table 6.7 presents the coefficients and standard errors from a school-level regression of the proportion of students meeting or exceeding state academic standards in the school on the treatment indicator (Sharpstown), a vector of year fixed effects, and (in column 2) a set of time-varying school-level demographics. We used the percentage meeting or exceeding standards across content areas as the dependent variable here and the DCIS and DCIS-M analyses because school-level scale scores were not available. For Sharpstown, we also used a school-level aggregate of students meeting standards in all core content areas rather than disaggregated by content area, because of data reporting conventions for the state. The scores were school-level averages (student-weighted within schools) for grades nine through 11; the content areas tested were language arts, mathematics, science, and social studies.

In column 1, which simply compared Sharpstown to average school-level performance in the state in each year (not to similar schools), we found that students in Sharpstown had high school proficiency rates that exceeded those of the state by 12.2 percentage points, though with only one treatment school, the effect was not significant at even the 10-percent level. However, when we adjusted for the school-level demographics shown in Table 6.7 for column 2, the positive difference in proficiency rates between Sharpstown and other schools in the state became significant at the 1-percent level, with an estimated difference of 17.9 percentage points. This reflects the fact that, in 2013, students at Sharpstown were markedly more likely than other students in the state to be minority (98 percent versus 66 percent in the state), low-income (94 percent versus 63 percent in the state), at-risk (61 percent versus 49 percent in the state), and limited English proficient (21 percent versus 16 percent in the state). In other words, when these differences between Sharpstown and the state average were controlled, the higher outcomes in Sharpstown became more pronounced.

Figure 6.5 presents these findings graphically. Here, the dashed line represents performance estimates, by year, for demographically comparable schools. The fitted estimates hold covariates constant at the Sharpstown sample means, so the estimates pertain to a school whose demographic characteristics and size are the same as those of Sharpstown in a given year. The dash-dotted line represents actual performance at the average school in the state, without any demographic adjustments. The solid line represents actual (rather than estimated) performance at Sharpstown in the three years it was observed in the data set. The graph clearly demonstrates that students at Sharpstown were outperforming students at similar schools and the state. Of course, as noted above, we cannot disentangle the portion of this outperformance that is due to the motivation of students and families choosing this school as opposed to the effectiveness of the competency-based (or other) instructional features of the school.

DCIS and DCIS-M: Regression Analysis

Finally, we focus on DCIS and DCIS-M, two competency-oriented schools of choice in the Denver Public Schools. The two schools shared a common philosophy, and their staff members regularly collaborated, but they were founded as separate schools. DCIS opened as a stand-alone facility in the 2006–2007 academic year, having expanded from a smaller and long-standing program on a shared campus. DCIS-M was much newer, having opened in the 2011–2012 academic year.[7] In addition, the two schools served demographically differ-

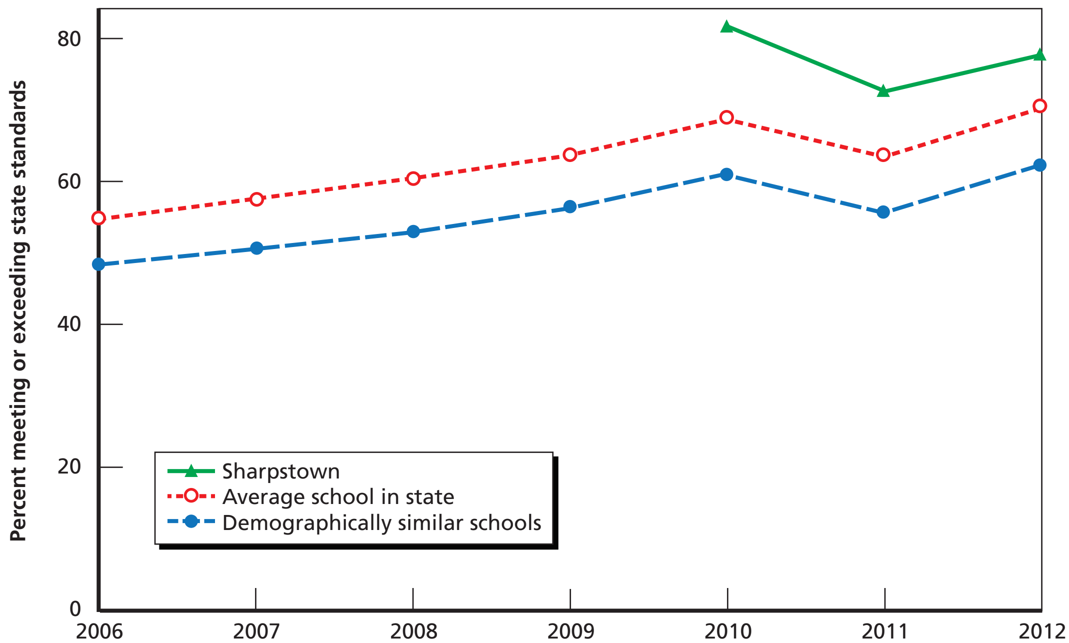
[7] Though DCIS-M was founded to follow in the tradition of DCIS, it was never part of DCIS.

Table 6.7
Coefficients and Standard Errors from Regression Comparing Sharpstown to Other Schools in the State

Characteristic	(1)	(2)
Sharpstown	12.231	17.873**
	(10.401)	(6.424)
2011	2.660***	7.563***
	(0.480)	(0.439)
2012	9.286***	13.633***
	(0.485)	(0.447)
2007		2.353***
		(0.395)
2008		4.619***
		(0.393)
2009		7.962***
		(0.392)
2010		12.748***
		(0.391)
School enrollment		0.002***
		(0.000)
School enrollment squared		–0.000**
		(0.000)
% Asian		0.719***
		(0.139)
% African-American		0.271*
		(0.135)
% Hispanic		0.389**
		(0.135)
% American Indian		0.225
		(0.168)
% white		0.429**
		(0.135)
% limited English proficiency		0.032*
		(0.015)
% at risk		–0.466***
		(0.006)
% free/reduced-price lunch		–0.086***
		(0.007)
Intercept	61.230***	43.121**
	(0.200)	(13.522)
N	11,512	11,512
R^2	0.032	0.632

NOTES: ~ p<0.10; * p<0.05, ** p<0.01, *** p<0.001; standard errors in parentheses. Dependent variable is the percentage meeting or exceeding proficiency in grade nine–11 core content.

Figure 6.5
Percentage Meeting State Academic Standards in Tested Subjects, by Year, in Sharpstown, Versus Average and Demographically Similar Schools in the State



NOTE: Solid circles indicate that the difference from the treatment school is statistically significant at the 10-percent level or better (in this case, 1 percent); hollow markers indicate no significant difference from treatment school.
RAND *RR732-6.5*

ent populations. DCIS, with about 640 students in grades six through ten in 2012–2013, served a population that was about 66 percent minority, and 48 percent of its students were eligible for subsidized meals. At DCIS-M, with a population of about 240 students in grades six through ten in the 2012–2013 school year, the corresponding percentages were 97 percent and 89 percent, respectively. The corresponding percentages for the state were 38 percent and 43 percent, respectively.

For DCIS and DCIS-M, we employed a common data set that provided school-by-grade-by-year records. We considered the two schools in the same analysis because they were the only two Asia Society sites located in the same state and thus subject to the same assessments and present in the same school-level data set. Similar to the Sharpstown analysis, our dependent variable of interest was the percentage of students meeting or exceeding state proficiency standards in a given grade and year, focusing on grades six through ten. We focused in this case on reading performance, since the competency-based work in these schools was reportedly most focused on humanities and language arts. Our analysis employed a multilevel model in which grade-level records were nested within schools. We included a school-level random effect in the regression model, which also included year- and grade-fixed effects. In addition, we included a range of school-level demographic controls.

Table 6.8 presents the regression results. Without adjusting for school-level demographics, we found in column 1 that students in the two treatment schools had a proficiency rate that was 5 percentage points lower in reading than their counterparts in other schools, but

Table 6.8
Coefficients and Standard Errors from Regression Comparing DCIS and DCIS-M to Other Schools in the State

Characteristic	(1)	(2)	(3)	(4)
DCIS and DCIS-M	–5.260 (12.86)	8.882 (7.107)		
DCIS only			16.82~ (9.716)	
DCIS-M only				0.146 (10.18)
2008	0.564*** (0.167)	0.721*** (0.167)	0.721*** (0.167)	0.719*** (0.167)
2009	0.852*** (0.168)	1.509*** (0.169)	1.508*** (0.169)	1.505*** (0.169)
2010	1.032*** (0.168)	1.800*** (0.171)	1.800*** (0.171)	1.798*** (0.171)
2011	0.651*** (0.168)	2.458*** (0.178)	2.457*** (0.178)	2.455*** (0.178)
2012	1.868*** (0.169)	3.787*** (0.180)	3.787*** (0.180)	3.787*** (0.180)
2013	1.766*** (0.168)	3.719*** (0.180)	3.718*** (0.180)	3.719*** (0.180)
Grade 7	–2.645*** (0.199)	–2.784*** (0.198)	–2.789*** (0.198)	–2.780*** (0.198)
Grade 8	–2.918*** (0.201)	–3.082*** (0.199)	–3.085*** (0.199)	–3.075*** (0.199)
Grade 9	–0.595~ (0.346)	–1.741*** (0.328)	–1.739*** (0.329)	–1.678*** (0.330)
Grade 10	0.881* (0.354)	–0.475 (0.332)	–0.481 (0.332)	-0.402 (0.334)
% Free/reduced-price lunch		–0.108*** (0.00791)	–0.108*** (0.00791)	–0.108*** (0.00791)
% Hispanic		–0.383*** (0.0106)	–0.383*** (0.0106)	–0.383*** (0.0106)
% African-American		–0.423*** (0.0216)	–0.423*** (0.0216)	-0.423*** (0.0216)
% Asian		0.208*** (0.0459)	0.207*** (0.0459)	0.207*** (0.0459)
% American Indian		–0.380*** (0.0538)	–0.379*** (0.0538)	–0.380*** (0.0538)
Magnet school		2.143 (1.506)	2.142 (1.506)	2.149 (1.507)
Title I eligible school		–0.0821 (0.276)	–0.0788 (0.277)	–0.0776 (0.277)
Enrollment		–0.00589*** (0.00166)	–0.00589*** (0.00166)	–0.00595*** (0.00166)
Intercept	65.25*** (0.456)	84.25*** (0.473)	84.24*** (0.473)	84.23*** (0.474)

Table 6.8—cont.

Characteristic	(1)	(2)	(3)	(4)
Log of SD of b/w school residual	2.890*** (0.0174)	2.250*** (0.0198)	2.250*** (0.0198)	2.251*** (0.0198)
Log of SD of w/in school residual	2.094*** (0.00395)	2.094*** (0.00398)	2.094*** (0.00398)	2.094*** (0.00398)
N school x grade records	33,874	33,672	33,666	33,640
N schools	1,745	1,698	1,697	1,697

NOTE: p<0.10; * p<0.05, ** p<0.01, *** p<0.001; standard errors in parentheses.
Dependent variable is the percentage meeting or exceeding proficiency in reading.

this difference did not approach statistical significance. When we added school-level demographic controls in column 2, the two treatment schools showed higher proficiency rates than similar schools, but again, this difference was not statistically distinguishable from zero. In columns 3 and 4, we estimated effects for DCIS and DCIS-M separately, excluding the other treatment school from the comparison set in each case. In this case, we did find a marginally significant, positive effect on proficiency rates for DCIS. It had a substantial 17 percentage-point magnitude, but was statistically significant at only the 10-percent level. In contrast, the fitted effect for DCIS-M, which was admittedly a much newer school serving a more-disadvantaged population than DCIS, we found that performance was virtually identical to what the school's demographic attributes would predict, as evidenced by the small and nonsignificant treatment effect.

Figure 6.6 presents these results graphically by year.[8] For simplicity, we hold grade level constant at grade nine in the figure, and we hold other predictors to their sample means. Here we see that DCIS consistently outperformed the state mean performance, as well as the fitted values for a school with the same demographics as DCIS-M. In contrast, in 2012, DCIS-M performed lower than the state mean and similar to the predicted performance for schools with the same demographics, with a slight drop in proficiency rates by 2013.

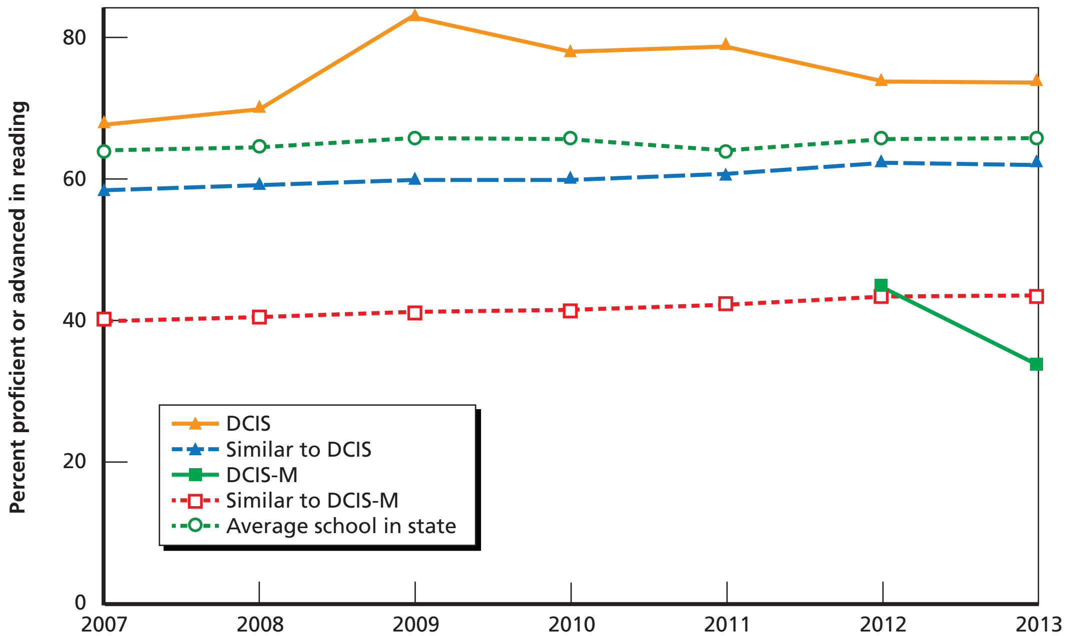
When interpreting these results, the same caveat applies here as applied to the regression estimates for Sharpstown. Specifically, we cannot say that the instructional or assessment models of these schools are responsible for these differences from state and similar-school performance means. There may be a variety of features of the schools that account for these differences, as well as unobserved attributes of the families choosing these schools.

Philadelphia

In Philadelphia, the Project Mastery intervention was available only to the students of eight teachers—distributed across six schools—who elected to take part in the pilot program. Given that these were the only students in the district with access to the Educurious curriculum materials and the Pearson LMS in 2012–2013, we can clearly distinguish pilot participants from nonparticipants, and thus can compare the two groups in terms of key outcomes of interest during the pilot year. Moreover, because the Educurious materials supplanted the traditional

[8] The graph presents observed rather than fitted performance for DCIS and DCIS-M. It presents fitted (i.e., model-predicted) performance estimates for the similar-school and state-level averages.

Figure 6.6
Percent Meeting State Academic Standards in Reading, by Year, in DCIS and DCIS-M, Versus Average and Similar Schools in the State



NOTE: Treatment schools are denoted with solid markers, and the state is denoted with a hollow marker. A solid marker for demographically similar comparison schools indicates that they were performing at levels that were statistically significantly different from their respective treatment school at the 10% level. Estimates are shown for grade 9; a random-effects model adjusts for the nesting of grades within schools.
RAND *RR732-6.6*

ninth-grade curriculum (a textbook of literature typically taught in that grade), the dosage was relatively high, and we might reasonably expect to see a difference in outcomes between pilot and nonpilot ninth-grade classrooms in terms of student achievement, engagement, and attainment. Since ninth-graders in Pennsylvania did not take an end-of-year accountability test in 2013, we measured academic achievement using two midyear benchmark tests—one administered in the fall (November 2012) and the other in winter (January 2013) of the academic year. We used school attendance—i.e., fraction of enrolled days for which the student was present—as a proxy for academic engagement. We measured academic attainment—that is, grade levels completed—in terms of whether students were promoted to tenth grade by the start of the 2013–2014 academic year.

Although Philadelphia was the only Project Mastery site in which we are able to undertake a student-level analysis, we faced an important limitation that was similar to the other sites, in that neither students nor teachers were randomly assigned to their treatment status. In Philadelphia, teachers volunteered to be part of the pilot, and only students who were assigned to those teachers' schools and classrooms were exposed to the pilot intervention. It is therefore possible that Project Mastery teachers and students may have differed from others in Philadelphia in ways that are related to the student outcomes of interest. Because we had data about students' demographic characteristics and prior academic performance, we used regression and propensity score weighting to adjust for a range of variables that may have been confounded with treatment status and thus biased our estimate of the treatment effect. However, even with these adjustments for observed differences between Project Mastery and other students in Philadelphia, it remains possible that students may have differed in unobserved ways that were related to both their treatment status and the measured outcomes.

Moreover, teachers who were part of the pilot used the Project Mastery materials with all of their ninth-grade classes, so we were unable to disentangle the relative effectiveness of these teachers from the effects of the Project Mastery materials. Our estimates of the Project Mastery effect in Philadelphia should therefore be properly interpreted as the effect of being in a Project Mastery classroom as taught by a teacher who volunteered for the program.

Table 6.9 displayed descriptive baseline characteristics for the nearly 7,000 students in the analytic sample, all of whom were in ninth grade in the 2012–2013 academic year.[9] Columns 1 and 2 show the mean and standard deviation, respectively, of the full sample on each descriptive characteristic. Columns 3 and 4 present these statistics just for the 401 students in the analytic sample who participated in Project Mastery, which we term the treatment group. Columns 5 and 6 apply just to the 6,646 students who did not take part in Project Mastery—that is, the comparison group.[10] Column 7 presents the difference in means between treatment and comparison-group students, and column 8 presents the *standardized mean difference*—that is, the difference in means divided by the pooled standard deviation in column 2. In other words, column 8 is reported in standard deviation units. Column 9 represents the p-value from a hypothesis test of the probability that the means in the treatment and comparison groups are identical. A p-value of 0.05 or below indicates that the groups were statistically different on that variable. If the groups were the same on most or all of the observed variables, then we are less concerned about unobserved differences between the treatment and comparison groups, insofar as relevant unobserved characteristics covary with those that are observed. Regardless, we are more concerned about the magnitude of the differences rather than about their statistical significance, since statistical significance is a function not only of the magnitude of the differences but also of the sample size. In Table 6.9, we find that Project Mastery students were about 6 percentage points less likely to be Latino and 5 percentage points more likely to be white than their non–Project Mastery counterparts. They were also about 5 percentage points less likely to be classified as gifted, 9 percentage points less likely to be classified as disabled, and 3 percentage points less likely to have limited English proficiency. They were 6 percentage points less likely to qualify for subsidized meals, to be female, and to be older than we would expect given their ninth-grade status. Their seventh-grade state reading scores were 10 percent of a standard deviation higher than comparison students. Their eighth-grade scores were 11 percent of a standard deviation higher, and their seventh- and eighth-grade mathematics scores were, respectively, 4 and 6 percent of a standard deviation higher than those of the comparison group. We also see a few differences in their school-level characteristics. The schools attended by Project Mastery students had a higher share of white students, a lower share of Latino and subsidized meal-eligible students, and modestly lower lagged reading scores, on average.

---

9 The data set of 2012–2013 ninth-grade students provided by the district includes 10,362 students, but only 6,847 had non-missing data in terms of their Project Mastery status and the baseline characteristics used in our models.

10 Four hundred seventy-five students were identified as Project Mastery participants in the data set, and 7,718 were identified as not exposed to Project Mastery, for a total of 8,203 students whose Project Mastery status was given. We then imputed Project Mastery status for 1,743 students in schools with no Project Mastery teachers, and for four students in schools in which all ninth-graders participated in Project Mastery. This left 412 students whose Project Mastery status remained missing; these students attended schools in which some students and teachers participated in Project Mastery and others did not. These students' English course identifications and teachers were not reported.

Table 6.9
Descriptive Statistics for Unweighted Sample, Overall and by Treatment Status

	1	2	3	4	5	6	7	8	9
Description	**Mean Pooled (n= 6,847)**	**SD Pooled**	**Mean Project Mastery (n=401)**	**SD Project Mastery**	**Mean Comparison (n=6,646)**	**SD Comparison**	**Mean Diff**	**Std. Mean Diff**	**p-value**
American Indian	0.002	*0.044*	0.002	*0.050*	0.002	*0.043*	0.001	0.015	0.78
Asian	0.085	*0.279*	0.095	*0.293*	0.084	*0.278*	0.010	0.037	0.47
African-American	0.569	*0.495*	0.579	*0.494*	0.568	*0.495*	0.010	0.020	0.69
Latino	0.184	*0.388*	0.127	*0.334*	0.188	*0.391*	−0.061	−0.157	0.00
Multi/other	0.018	*0.131*	0.007	*0.086*	0.018	*0.134*	−0.011	−0.081	0.11
White	0.142	*0.349*	0.190	*0.392*	0.139	*0.346*	0.050	0.144	0.01
Gifted	0.065	*0.247*	0.020	*0.140*	0.068	*0.251*	−0.048	−0.194	0.00
Disability	0.165	*0.371*	0.085	*0.279*	0.170	*0.375*	−0.085	−0.229	0.00
Limited English proficiency	0.059	*0.235*	0.035	*0.184*	0.060	*0.238*	−0.025	−0.107	0.04
Subsidized meals	0.662	*0.473*	0.606	*0.489*	0.666	*0.472*	−0.060	−0.125	0.01
Free meals	0.638	*0.481*	0.589	*0.493*	0.641	*0.480*	−0.053	−0.344	0.03
Reduced-price meals	0.024	*0.153*	0.017	*0.131*	0.025	*0.155*	−0.007	−0.014	0.37
Female	0.491	*0.500*	0.436	*0.497*	0.495	*0.500*	−0.058	−0.116	0.02
Over age for grade	0.160	*0.367*	0.102	*0.303*	0.164	*0.370*	−0.061	−0.168	0.00
2012 reading[a]	0.022	*1.001*	0.122	*0.675*	0.016	*1.017*	0.106	0.106	0.04
2011 reading[a]	0.048	*0.989*	0.137	*0.701*	0.042	*1.004*	0.095	0.096	0.06
2012 math[a]	0.024	*1.006*	0.076	*0.714*	0.020	*1.021*	0.056	0.056	0.28
2011 math[a]	0.093	*1.005*	0.129	*0.736*	0.091	*1.020*	0.039	0.038	0.46
2011 testing grade	6.992	*0.112*	6.998	*0.087*	6.992	*0.114*	0.006	0.052	0.31
Size of grade 9 cohort	307.479	*217.072*	218.352	*180.309*	313.023	*217.966*	−94.671	−0.436	0.00
School % African-American[b]	56.407	*26.721*	55.852	*25.206*	56.442	*26.814*	−0.589	−0.022	0.67
School % Latino	18.398	*18.222*	12.843	*10.033*	18.744	*18.559*	−5.901	−0.324	0.00

Table 6.9—Cont.

	1	2	3	4	5	6	7	8	9
Description	**Mean Pooled (n= 6,847)**	**SD Pooled**	**Mean Project Mastery (n=401)**	**SD Project Mastery**	**Mean Comparison (n=6,646)**	**SD Comparison**	**Mean Diff**	**Std. Mean Diff**	**p-value**
Size of grade 9 cohort	307.479	*217.072*	218.352	*180.309*	313.023	*217.966*	–94.671	–0.436	0.00
School % African-American[b]	56.407	*26.721*	55.852	*25.206*	56.442	*26.814*	–0.589	–0.022	0.67
School % Latino	18.398	*18.222*	12.843	*10.033*	18.744	*18.559*	–5.901	–0.324	0.00
School % white	14.613	*15.796*	21.433	*13.304*	14.188	*15.842*	7.245	0.459	0.00
School % subsidized meals	63.885	*11.950*	59.729	*8.940*	64.143	*12.066*	–4.415	–0.369	0.00
School % limited English proficiency	7.040	*6.784*	6.686	*7.727*	7.062	*6.722*	–0.376	–0.055	0.28
School % with disability	16.732	*10.216*	16.613	*6.837*	16.740	*10.390*	–0.127	–0.012	0.81
School % over age	24.427	*15.080*	24.740	*13.686*	24.408	*15.163*	0.332	0.022	0.67
School 2012 reading	0.007	*0.673*	–0.108	*0.312*	0.014	*0.688*	–0.122	–0.181	0.00

[a] Test scores are standardized to have mean 0 and standard deviation 1 within the full ninth-grade district sample.

[b] School-level percentages are scaled from 0 to 100. Individual, categorical characteristics are scaled from 0 to 1.

Propensity Score Weighting of Treatment and Comparison Groups

In an attempt to create a comparison group that was more similar to the treatment group on observable characteristics, we estimated each student's propensity for participating in Project Mastery as a function of both student-level and school-level characteristics. Following Rosenbaum & Rubin (1983), Austin (2011), and Funk et al. (2011), we estimated propensity scores using a logistic regression model, in which the treatment status z_{is} of student i in school s in the study year 2012–2013 (coded 1 for treatment and 0 for comparison), was predicted as a function of student characteristics X_{is}.[11] The fitted values, $\hat{z}_{is}$, were the propensity scores, in that they captured the estimated probability that student i in school s participates in Project Mastery. The propensity score estimation model was specified as follows:

$$\ln\left(\frac{z_{is}}{1-z_{is}}\right)=\alpha_0+\alpha_1'X_{is}+\xi_{is}\,, \tag{3}$$

where α_0 is an intercept term, X_{is} is a vector of student-level characteristics that includes gender and race/ethnicity dummy variables; indicators of gifted, disabled, limited English proficiency, and over age for grade status; indicators of free and reduced-price meal eligibility; the students' lagged mathematics and reading test scores from seventh and eighth grade, as well as an indicator variable for the grade level of the lagged tests, and ξ_{is} is a random error term. Parameter vector α_1 captures the conditional, linear relationships between each element of X_{is} and the dependent variable, z_{is}.

To obtain an estimate of the average effect of treatment on the treated (the ATT), we construct a weighting variable for the full analytic sample, where the individual-level weight, $w_{is,ATT}$, is calculated as follows:

$$w_{is,ATT}=z_{is}+\frac{(1-z_{is})\hat{z}_{is}}{1-\hat{z}_{is}}\,. \tag{4}$$

In equation 4, z_{is} is the observed treatment status (1 or 0) for student i, and $\hat{z}_{is}$ is the student's estimated propensity score. Because this weight is used to estimate the ATT, this approach assigns each treatment-group student a weight of 1 and each comparison group student a weight equal to his or her fitted odds of being in the treatment group. In other words, rather than weighting both groups to be similar to the pooled sample average, which would estimate the average treatment effect, the comparison group is instead weighted to closely approximate the attributes of the treatment group, and each member of the treatment group retains a weight of 1. In this way, the treatment effect is estimated with respect to a group that resembles the treatment group (Austin, 2011).

Our weighted sample achieved very good balance between the treatment and comparison groups in terms of student-level characteristics, as shown in Table 6.10. All student-level differences fell to essentially zero or—in the case of lagged test scores—to no more than half of

[11] We explored several propensity score specifications that included a vector of school-level variables, S_s, in addition to student-level characteristics X_{is}. However, these models made only slight improvements to the school-level balance between treatment and comparison students, while markedly reducing student-level balance, even to levels lower than in the unweighted sample. Because students' Project Mastery status varies both within and between schools, and because our analysis is conducted at the student level, we are especially interested in constructing groups that are balanced on student-level characteristics. By using propensity score weights estimated only with student-level characteristics, we achieve balance on the largest number of variables, and on all of the student-level variables.

1 percent of a standard deviation, and none were statistically significant. Though some modest school-level differences did remain, we found that modifying equation 3 to include school-level characteristics in the propensity score estimation model actually weakened overall balance on all of the individual-level variables and on almost half of the school-level variables, as footnoted. Weighting on propensity scores as estimated in equation 3 yielded the highest level of overall balance across variables of interest. In addition, we explored weighting the sample in order to estimate the average treatment effect rather than the ATT, but this approach also achieved poorer balance than we were able to achieve with ATT weights.

Dependent Variables

As noted, we examined the effects of Project Mastery participation in Philadelphia using four outcomes of interest: the percentage of items students answered correctly on their ELA benchmark tests in November and January, their rate of absence from school during the 2012–2013 school year, and whether they were promoted to the tenth grade by the start of the following school year. The fall benchmark scores ranged from 0 to 96, with a mean of 46.3 percent and a standard deviation of 18.7 percentage points. The winter scores ranged from 0 to 96, with a mean of 40.0 percent and a standard deviation of 17.2 percentage points. Both distributions were symmetric and bell-shaped, consistent with the normality assumptions of linear regression. The percentage of days absent ranged from 0 to 95.7 but was not symmetric; the distribution had a median of 7 and a strong positive skew of 2.6. Consequently, we performed our analysis using the natural log of the percent of days absent plus one—that is, *ln*(percent_absent+1). The logged distribution ranged from 0 to 4.6, with a mean of 2.1 and a standard deviation of 1, and a somewhat bell-shaped distribution with only a small negative skew of –0.1. The grade promotion variable took on a value of 1 to indicate promotion and 0 to indicate otherwise. It had a mean of 0.89 in the analytic sample. Because it was dichotomous rather than continuous, we used the logged odds, or logit, of the dependent variable to estimate the relationship between Project Mastery participation and grade promotion in most analyses.[12]

Estimation Models

To estimate the effect of Project Mastery participation on the student outcomes of interest, we fit several model specifications. The most basic specification was an ordinary least squares linear model in which each of the dependent variables is regressed on the students' Project Mastery status and a vector of student-level and school-level control variables. In the most basic specification, we did not include propensity score weights and did not adjust for the nesting of students within schools. This basic model was specified as follows:

$$y_{is} = \beta_0 + \beta_1 z_{is} + \gamma_1' X_{is} + \lambda_1' S_s + \varepsilon_{is}, \tag{5}$$

where y_{is} is one of the aforementioned dependent variables of interest, z_{is} is Project Mastery treatment status, X_{is} is a vector of the previously described student-level background characteristics, S_s is a vector of school-level characteristics shown in Table 6.9, and ε_{is} is a normally distributed error term with mean 0 and standard deviation σ. The parameter of interest is β_1,

[12] We used a linear probability model in the specification that included propensity score weights in a random effects framework. This allowed us to take advantage of the *xtregre2.ado* file in Stata, which permits the use of propensity score weights in a random effects model (Merryman, 2005).

Table 6.10
Descriptive Statistics for Propensity-Score Weighted Sample, Overall and by Treatment Status

	1	2	3	4	5	6	7	8	9
Characteristic	**Mean Pooled (wtd n=802.17)**	**SD Pooled**	**Mean Project Mastery (wtd n=401)**	**SD Project Mastery**	**Mean Comparison (wtd n=401.17)**	**SD Comparison**	**Mean Diff**	**Std. Mean Diff**	**p-value**
American Indian	0.002	*0.050*	0.002	*0.050*	0.002	*0.049*	0.000	0.001	0.96
Asian	0.095	*0.293*	0.095	*0.293*	0.095	*0.293*	0.000	0.001	0.98
African-American	0.579	*0.494*	0.579	*0.494*	0.579	*0.494*	0.000	–0.001	0.98
Latino	0.127	*0.333*	0.127	*0.334*	0.127	*0.332*	0.001	0.002	0.93
Multi/other	0.007	*0.086*	0.007	*0.086*	0.007	*0.086*	0.000	0.000	1.00
White	0.190	*0.392*	0.190	*0.392*	0.190	*0.392*	–0.001	–0.002	0.95
Gifted	0.020	*0.140*	0.020	*0.140*	0.020	*0.140*	0.000	0.000	1.00
Disability	0.085	*0.279*	0.085	*0.279*	0.085	*0.279*	0.000	0.000	1.00
Limited English proficiency	0.035	*0.183*	0.035	*0.184*	0.035	*0.183*	0.000	0.002	0.94
Subsidized meals	0.606	*0.489*	0.606	*0.489*	0.605	*0.489*	0.001	0.002	0.94
Free meals	0.588	*0.492*	0.589	*0.493*	0.588	*0.492*	0.001	0.002	0.94
Reduced-price meals	0.017	*0.131*	0.017	*0.131*	0.017	*0.131*	0.000	0.000	0.99
Female	0.436	*0.496*	0.436	*0.497*	0.436	*0.496*	0.000	0.000	0.99
Over age for grade	0.102	*0.303*	0.102	*0.303*	0.102	*0.303*	0.000	0.000	0.99
2012 reading[a]	0.124	*0.832*	0.122	*0.675*	0.126	*0.964*	–0.004	–0.005	0.85
2011 reading[a]	0.138	*0.832*	0.137	*0.701*	0.140	*0.946*	–0.002	–0.003	0.91
2012 math[a]	0.078	*0.851*	0.076	*0.714*	0.080	*0.969*	–0.004	–0.005	0.85
2011 math[a]	0.131	*0.858*	0.129	*0.736*	0.133	*0.965*	–0.004	–0.005	0.85
2011 testing grade	6.997	*0.084*	6.998	*0.087*	6.997	*0.082*	0.000	0.002	0.93
Size of grade nine cohort	267.682	*208.967*	218.352	*180.309*	316.992	*223.675*	–98.641	–0.472	0.00

Table 6.10—Cont.

Characteristic	1 Mean Pooled (wtd n=802.17)	2 SD Pooled	3 Mean Project Mastery (wtd n=401)	4 SD Project Mastery	5 Mean Comparison (wtd n= 401.17)	6 SD Comparison	7 Mean Diff	8 Std. Mean Diff	9 p-value
School % African-American[b]	56.234	*25.854*	55.852	*25.206*	56.615	*26.509*	–0.763	–0.030	0.22
School % Latino	15.073	*13.934*	12.843	*10.033*	17.302	*16.670*	–4.458	–0.320	0.00
School % white	18.264	*15.233*	21.433	*13.304*	15.097	*16.353*	6.337	0.416	0.00
School % subsidized meals	61.606	*10.735*	59.729	*8.940*	63.483	*11.987*	–3.755	–0.350	0.00
School % limited English proficiency	6.733	*7.075*	6.686	*7.727*	6.780	*6.368*	–0.093	–0.013	0.59
School % with disability	16.433	*8.669*	16.613	*6.837*	16.252	*10.179*	0.360	0.042	0.09
School % over age	24.093	*14.229*	24.740	*13.686*	23.446	*14.738*	1.295	0.091	0.00
School 2012 reading	–0.032	*0.531*	–0.108	*0.312*	0.044	*0.675*	–0.152	–0.286	0.00

[a] Test scores are standardized to have mean 0 and standard deviation 1 within the full ninth-grade district sample.

[b] School-level percentages are scaled from 0 to 100. Individual, categorical characteristics are scaled from 0 to 1.

which represents the conditional mean difference between Project Mastery and comparison students with regard to the dependent variable, adjusting for the other terms in the model.

The second specification properly adjusted for the nesting of students within schools by including a random intercept term that was permitted to vary at the school level. This adjustment allowed for the similarity of experiences among students in the same school and constrained the degrees of freedom for estimating school-level effects to the number of schools involved in the analysis.[13] The two-level model was specified as above, except that in equation 6, we allow the intercept to vary at both the school and student levels:

$$y_{is} = \beta_0 + \beta_1 z_{is} + \gamma_1' X_{is} + \lambda_1' S_s + u_s + \varepsilon_{is} \,, \tag{6}$$

so that u_s now represents the school-level error term, and ε_{is} represents the student-level error term.

The third specification was identical to equation 6, except that the sample was weighted in terms of the ATT weights described in equation 4. Insofar as either the propensity score estimation model or the regression model is correctly specified, this third specification provides a doubly robust estimate. This means the estimate of β_1 should be unbiased, contingent on correct specification of either equation 3 or equation 6 in adjusting for potential confounds of the treatment effect (Funk et al., 2011). For this reason, this third specification is our preferred method for estimating the effect of Project Mastery on student outcomes.

However, given that three of the six treated schools included students who were classified as being enrolled in either Project Mastery or comparison classes, we were also interested in how these students performed relative *only* to their same-school counterparts, adjusting for the other terms in the regression model. To obtain these within-school comparisons, we fit two school fixed-effect specifications. The first included school-fixed rather than random effects, but was estimated within the unweighted sample:

$$y_{is} = \beta_2 + \beta_3 z_{is} + \gamma_2' X_{is} + \kappa_1' E_s + \varepsilon_{is} \,, \tag{7}$$

where the terms are as specified above, except that the vector of school-level characteristics, S_s, has been replaced by a vector of school indicator variables, E_s, with fixed effects given by parameter vector κ_1. The final model specification is identical to equation 7, except that it employs a doubly robust approach, in that the regression estimates also reflect the application of ATT weights to the sample.

In addition, we investigated the effects of the afterschool digital filmmaking badges with a modification to equation 6, in which we added a dichotomous indicator of whether a Project Mastery student was also a badges student, meaning he or she attended at least 50 percent of the badges sessions. This indicator, b_{is}, was coded 1 for a badges program participant (all of whom were Project Mastery students), and 0 for all other students, whether or not they were part of Project Mastery:

$$y_{is} = \beta_4 + \beta_5 z_{is} + \beta_6 b_{is} + \gamma_3' X_{is} + \lambda_2' S_s + u_s + \varepsilon_{is} \,. \tag{8}$$

[13] Because we observed only two schools in which more than one treatment teacher was nested within the same school, estimating a three-level model with a teacher-level random effect yielded problems with model convergence. Since treatment was assigned at the school/teacher level, we also did not adjust for nesting of classrooms within teachers—a decision consistent with What Works Clearinghouse standards (2014) for estimating treatment effects.

In equation 8, β_5 represents the average difference between Project Mastery and other students, adjusted for the other terms in the model, and β_6 represents the average difference between badges and nonbadges students. Consequently, β_5 alone represents the average performance of a Project Mastery student who is not a badges student relative to a non–Project Mastery student. Meanwhile, $\beta_5+\beta_6$ provide the average difference in performance between a badges student and a non-Project Mastery student.

We first consider just the performance of Project Mastery students relative to non-Project Mastery students. Results for the five estimation models are shown in Table 6.11, where the dependent variables are represented in rows and the model specifications are shown in columns 1 through 5. For each dependent variable, the top row represents the fitted regression coefficient, with the standard error beneath it in parentheses. Beneath the standard error, we present interpretations of the fitted treatment effects. Column 1 presents regression results that include statistical controls but no adjustment for the nesting of observations within schools. Column 2 presents regression results that include a school random effect to adjust for within-school nesting. Column 3 includes not only school random effects but also ATT propensity score weights and is our preferred specification. Columns 4 and 5 include school fixed effects. Though all schools were included in the regression models, the treatment effect in Columns 4 and 5 were estimated only within the three schools that served both Project Mastery and comparison students. Column 4 included regression controls, and column 5 included regression controls as well as ATT propensity score weighting. Though these school fixed-effect models had the distinct advantage of holding school-level factors constant, their disadvantage is that the treatment effects were estimated only within half of the Project Mastery schools, whereas columns 1 through 4 include both within-school and between-school comparison groups. Because we were examining four outcomes simultaneously, we used the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) to adjust for multiple comparisons. Our statistical significance indicators in Tables 6.11 and 6.12 and in Figures 6.7 and 6.8 reflect the adjusted significance thresholds.

Focusing on our preferred specification, we estimated that Project Mastery students scored 2.6 percentage points lower than comparison students on the first ELA benchmark test of the year. In the sample, this means that Project Mastery students had a mean performance of 46.4 percent correct, as compared to 49.0 for the comparison group, and this difference was statistically significant, even when we adjusted our hypothesis testing threshold for multiple comparisons. This difference is represented graphically in the first pair of bars in Figure 6.7. What is important to bear in mind is that students generally completed this assessment in late November, only three months into the school year, at which point they had been exposed to no more than two Educurious units. By the January 2013 benchmark test, Project Mastery students were outperforming their counterparts by about 0.86 percentage points correct (that is, nearly one percentage point) on average, though the effect was not statistically significant. This difference is represented in the second pair of bars in Figure 6.7.

We also examined attendance rates during the full school year, and here we did find a statistically significant difference that favored the Project Mastery group. Using the estimates in column 3, we found that Project Mastery students had an attendance rate that was 1.4 percentage points higher than that of their counterparts. As shown in Figure 6.7, the average rate was 92.4 percent for the treatment group and 91.0 percent for the weighted comparison group, holding the other variables in the regression model constant at their treatment-group sample

Table 6.11
Estimated Coefficients and Interpretive Statistics for Project Mastery Participation in Philadelphia

			Model Specification				
Outcome	**Model**	**Row Descriptor**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Fall benchmark	Linear	Coefficient and significance	–0.872	–2.293	–2.600*	–3.182*	–4.029**
		Standard error	(0.865)	(1.268)	(1.117)	(1.483)	(1.51)
		Fitted % correct (comparison=48.98)	48.11	46.69	46.38	45.80	44.95
Winter benchmark	Linear	Coefficient and significance	0.396	1.022	0.863	1.998	1.152
		Standard error	(0.794)	(1.071)	(0.84)	(1.312)	(1.319)
		Fitted % correct (comparison=40.72)	40.33	39.71	39.86	38.73	39.58
Absence rate	Log-linear	Coefficient and significance	–0.260***	–0.158**	–0.149***	–0.024	0.074
		Standard error	(0.046)	(0.059)	(0.045)	(0.077)	(0.065)
		Treatment-to-comparison ratio of absence rates	0.771	0.854	0.862	0.977	1.077
		Fitted absence rate difference (0–100 scale)	–2.280	–1.456	–1.379	–0.231	0.767
Promoted to tenth grade (n=6,339)	Log-odds (linear probability in column 3)	Coefficient and significance	0.111	–0.0961	0.013	–0.215	–0.434
		Standard error	(0.227)	(0.283)	(0.016)	(0.315)	(0.814)
		Treatment-to-comparison odds ratio	1.117	0.908	1.159	0.807	0.648
		Fitted percentage-point difference in probability (0–100 scale)	0.533	–0.506	1.310	–1.194	–2.662
Regression controls			X	X	X	X	X
School random effects				X	X		
School fixed effects (i.e., within-school estimates)						X	X
Propensity score weights					X		X

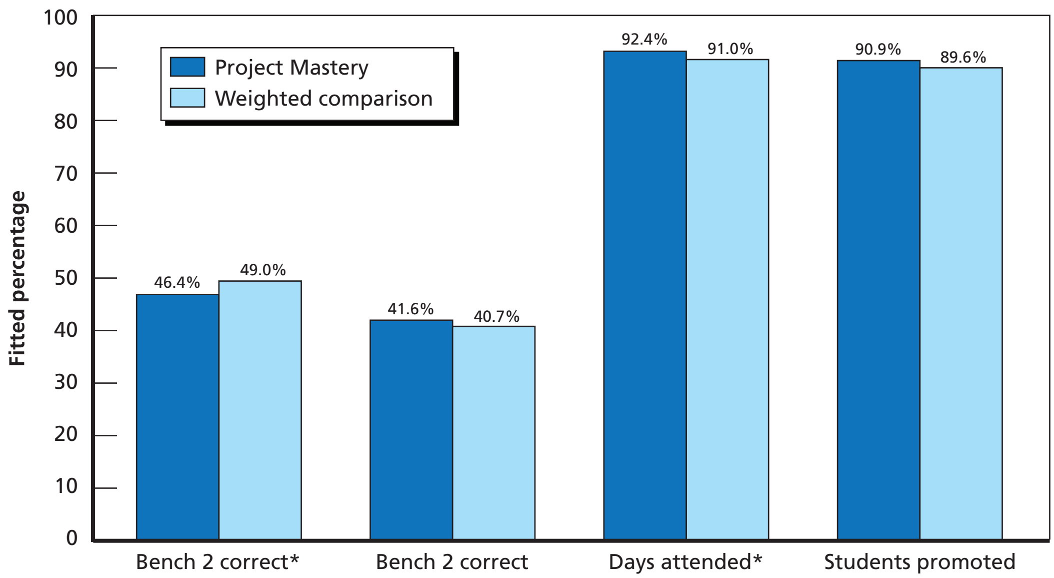
NOTES: ~ p<0.1; * p<0.05, ** p<0.01, *** p<0.001, adjusted for multiple comparisons.
Preferred specification is column 3 because it includes weights and school random effects and compares students across schools, as well as within them. All models include both student-level and school-level controls, except for school fixed-effect models, which include student-level controls. Model 3 uses a linear probability rather than logistic regression model for the on-time promotion indicator.

Table 6.12
Estimated Coefficients and Interpretive Statistics for Project Mastery and Badges Program Participation in Philadelphia

			Model Specification		
Outcome	**Model**	**Row Descriptor**	**(3)**	**(3a)**	**(3b)**
Fall benchmark (n=4,308)	Linear	Coefficient and significance	–2.600*	–2.749*	1.968
		Standard error	(1.117)	(1.129)	(1.145)
		Fitted % correct (comparison=48.98)	46.38	46.23	48.20
Winter benchmark (n=4,535)	Linear	Coefficient and significance	0.863	0.800	1.107
		Standard error	(0.84)	(0.861)	(0.996)
		Fitted % correct (comparison=40.72)	39.86	41.53	42.63
Absence rate (n=6,847)	Log-linear	Coefficient and significance	–0.149**	–0.138**	–0.221**
		Standard error	(0.045)	(0.045)	(0.056)
		Treatment-to-non–Project Mastery ratio of absence rates	0.862	0.857	0.665
		Fitted absence rate difference from non–Project Mastery (0–100 scale)	–1.379	–1.284	–3.009
Promoted to tenth grade (n=6,339)	Linear probability	Coefficient and significance	0.013	0.009	0.057**
		Standard error	(0.016)	(0.016)	(0.020)
		Treatment-to-non–Project Mastery odds ratio	1.159	1.107	2.874
		Fitted percentage-point difference in probability from non–Project Mastery (0–100 scale)	1.310	0.919	6.563
Regression controls			X	X	X
School random effects			X	X	X
School fixed effects (i.e., within-school estimates)					
Propensity score weights			X	X	X

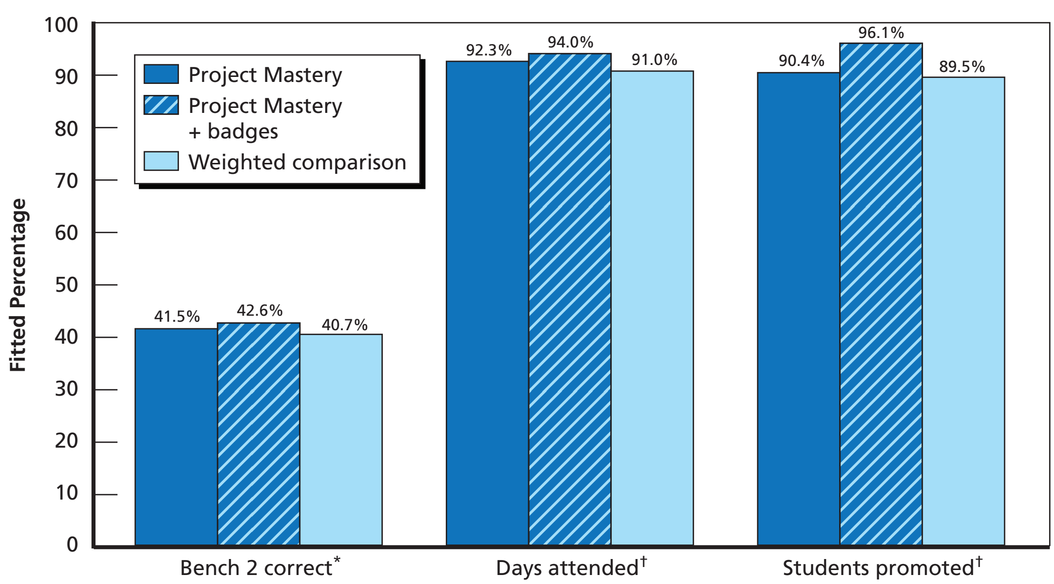
NOTES: ~ p<0.1; * p<0.05, ** p<0.01, *** p<0.001, adjusted for multiple comparisons.
As in Table 6.3, column 3 represents the overall Project Mastery effect with no inclusion of a badging effect in the model. Column 3a now represents the estimated Project Mastery effect net of the badges effect, and column 3b represents the fitted effect of 50 percent or more participation in the badging program, holding constant the Project Mastery effect.

Figure 6.7
Outcome Estimates for Project Mastery vs. Comparison Students in Philadelphia



*Significant at 5% level, adjusted for multiple comparisons.
RAND *RR732-6.7*

Figure 6.8
Outcome Estimates for Badges vs. Project Mastery–Only vs. Comparison Students in Philadelphia



*Project Mastery significant at 5 percent; †Badges significant at 5 percent; both adjusted for multiple comparisons.
RAND *RR732-6.8*

means.[14] When we examined the proportion of students promoted to tenth grade by October of the following school year, we found that the Project Mastery group's promotion rate was about 1.3 percentage points higher than that of the weighted comparison group, though this difference was not statistically significant. This difference is shown in Figure 6.7, where we see that the fitted promotion rate was 90.9 for the Project Mastery group, as compared to 89.6 for the weighted comparison group.

We now turn to an analysis of how students in the afterschool badges program performed relative to their peers. Again, we defined a badges participant as a student who took part in at least 50 percent of the badges sessions offered. This represented 31 of the 47 students in the data set who were shown as taking part in badges at all, though site-based reports indicated that the true number of overall participants was closer to 60.

Our findings are shown in Table 6.12, where column 3 simply duplicates column 3 in Table 6.11 and is included here simply as a basis of comparison. The coefficients in column 3a correspond to parameter $\boldsymbol{\beta_5}$ in equation 8. In other words, they represent the estimated average performance of Project Mastery students who were not badges students (including those with less than 50 percent badges participation) relative to non–Project Mastery students, adjusting for the other terms in the model. The coefficients in column 3b correspond to parameter $\boldsymbol{\beta_6}$ in equation 8. They represent the average performance of badges students relative to other Project Mastery students, so that the sum of the 3a and 3b coefficients represents the performance of badges students relative to non–Project Mastery students in the district.[15]

We found that, though badges students in the sample outperformed other Project Mastery students on both the fall and winter benchmarks, these differences were not statistically significant. In fact, we would not expect to see a badges program have an impact on the fall benchmark, since the badges program was just getting off the ground in November 2012. The fact that we did see some difference (albeit not statistically significant) suggests that there may be some uncontrolled selection bias at play. In other words, it may be that students who participated in at least 50 percent of the badges classes were simply more motivated students than those who did not, and that these motivational differences, rather than the badges program alone, at least partially accounted for differences in the other three outcomes. Though we continued to use a propensity-score weighted regression model in this analysis (using weights for the overall Project Mastery sample), we could not rule out the possibility that the two groups differed in ways that were not fully captured in the data.

Regardless, the estimated differences in absence rates and even in promotion rates indicated that students who took part regularly in the badges program outperformed those who did not, and that these differences were statistically significant. The magnitude of the differences can be interpreted most easily with reference to Figure 6.8, where we display the fitted estimates from columns 3a and 3b. We found that badges students had an estimated atten-

---

[14] In other words, the actual, estimated values shown for all the dependent variables in Figure 6.7 applied to a student whose other characteristics (race, gender, prior test scores, etc.) corresponded to the means in the treatment-group sample. However, the *average differences* in Figure 6.7 between treatment and comparison group students did not depend on the students' background characteristics.

[15] The coefficients in columns 3a and 3b do not sum to the values in column 3 because column 3 reflects the performance of all Project Mastery relative to non–Project Mastery students, not just of the badges students relative to non–Project Mastery students. Instead, the coefficient estimates in columns 3a and 3b bound the estimates in column 3. As specified in equation 6, column 3 includes both badges and nonbadges students as part of the Project Mastery main effect. Columns 3a and 3b correspond to equation 8 and therefore estimate different performance parameters for badges and nonbadges students.

dance rate that was 1.7 percentage points higher than other Project Mastery students, and 3.0 percentage points higher than other students in the district, holding the other terms constant at their sample means. Both the overall Project Mastery effect and the badges effect were statistically significant in this model, even with an adjustment for the fact that we are examining four outcomes. We also noted that badges participants were 5.7 percentage points more likely than other Project Mastery students to have been promoted to tenth grade by the fall of 2013—another statistically significant difference, even adjusting for multiple comparisons. In short, students who took part in the badges program had higher rates of attendance and promotion than other students in the district, including others who took part in Project Mastery. What is less clear is how much of this difference was driven by the intrinsic motivation and interest levels of the badges students and how much was the product of the engagement and learning the badges program generated.

Cross-Site Synthesis

We now consider what lessons for the field can be gleaned from a cross-site synthesis of the student outcome results. A synthesis is challenging due to considerable heterogeneity in terms of not only the features of each site, but also the characteristics of our site-specific analyses, and the dependent variables on which we focused. Thus, rather than meta-analyzing the results to generate a single estimate for the effects of competency-based education, we summarize patterns of alignment between the outcome estimates for each site, the defining features of each program, and the internal validity of each analytic approach. These patterns are represented in Table 6.13. In the table, the key defining features that we identified for each site in Chapter Three (see Table 3.2) are represented in column 2. The gray shading simply distinguishes among the three emphasis categories, where Adams 50 emphasized flexible pacing and evaluation for proficiency; Newfound and DCIS-M emphasized student choice and evaluation for proficiency; and the other three sites emphasized student choice. Column 3 denotes the most likely confounds to the estimates for each site, and column 4 rates the strength of our overall causal inference for that site in terms of our ability to attribute the estimated outcomes to competency-based education per se. The color scheme in columns 3 and 4 is green for better, yellow for moderate, and red for weaker. Column 5 summarizes the estimates in terms of their direction and significance for each measured outcome. Green means positive and statistically significant, yellow means no significant effect, and red means negative and statistically significant. Because none of our approaches leverages randomization to estimate a causal effect, we choose "better" rating in column 4 as our highest basis for causal evidence. This is based on the fact that our estimates for Project Mastery in Philadelphia include statistical controls for a variety of school-level and student-level characteristics and provide a close, weighted match between the characteristics of treatment and comparison students. In fact, our primary Philadelphia analysis would likely meet the U.S. Department of Education's What Works Clearinghouse evidence standards with reservations because we use regression adjustment in a weighted data set that has very strong balance between treatment and control groups in terms of student-level variables (What Works Clearinghouse, 2014).

However, we rate the badges analysis in Philadelphia as having only a moderate basis for causal inference. The reason is that student-level selection on unobserved variables is much more likely in a comparison of students whose treatment status is defined by a choice variable,

Table 6.13
Summary of Sites' Features, Analysis Strength, and Outcome Estimates

Site	Defining Features	Key Possible Confounds	Basis for Causal Interpretation	Estimates
Adams 50	Flexible pacing; proficiency-based evaluation	Unobserved, time-varying district-level characteristics	Moderate	Negative and significant for math
Asia Society—Newfound	Student choice; proficiency-based evaluation	Unobserved, time-varying school and district-level characteristics	Moderate	Positive but nonsignificant for reading
Asia Society—Sharpstown	Student choice	Unobserved student and school characteristics (in a school of choice)	Weaker	Positive and significant for test scores
Asia Society—DCIS	Student choice	Unobserved student and school characteristics (in a school of choice)	Weaker	Positive and significant for reading
Asia Society—DCIS-M	Student choice; proficiency-based evaluation	Unobserved student and school characteristics (in a school of choice)	Weaker	Negative but nonsignificant for reading
Philadelphia: Overall	Student choice	Unobserved student and school characteristics not captured by extensive student-and school-level controls and weighting	Better	Positive and significant for attendance
				Negative and significant for fall reading benchmark
				Nonsignificant for winter reading benchmark and promotion
Philadelphia: Badges	Student choice	Unobserved student characteristics (in a program of choice)	Moderate	Positive and significant for attendance and promotion
				Nonsignificant for winter reading benchmark

NOTE: Gray shading denotes distinctive program features. In columns 3-5, green=advantageous; yellow=neutral; red=disadvantageous.

such as voluntary afterschool participation and persistent attendance in a digital filmmaking program, even when we weight on and control for a host of background covariates. As Steiner, Cook, and Shadish (2011) have shown, the ability of propensity score weights and regression controls to reduce bias depends critically on how well the covariates capture the selection process. If selection occurs on uncontrolled variables that are strongly correlated with the outcome of interest but not with the observed covariates, then some of that correlation may be misattributed to the treatment effect. Still, we would expect that baseline differences in prior academic performance, alongside the other covariates, would help to reduce baseline differences in unobserved motivation that could confound the treatment effect, and that is why we rate the comparison moderate rather than "weaker."

We also give a moderate evidence rating to our SCG analyses for Adams 50 and Newfound, even though the developers of the method characterize it as a strategy that can warrant causal inference (Abadie et al., 2007; Abadie & Gardeazabal, 2003). Our reasoning is that, while the method may estimate a reasonable counterfactual for the two sites, it cannot eliminate any effects of time-varying changes that may have occurred in the treatment or comparison sites after the competency-based reforms took effect in the treatment sites. Though our interviews suggest that these were primarily focused on competency-based reforms, it remains possible that other factors at play in the sites were at least partially responsible for the effects.

Finally, we reserve our lowest evidence rating, "weaker," for the three sites in which we are describing performance in schools of choice and are unable to adjust for student-level background characteristics, or even for precompetency school-level performance. Though these analyses for Sharpstown, DCIS, and DCIS-M do control for time-varying school-level characteristics, they remain potentially confounded by their staffing and instructional attributes that are not related to their competency-based models, and by the fact that they are schools of choice.[16] Consequently, even if we control for school-level demographics, selection remains a threat insofar as we are unable to control for prior student performance and background characteristics.

When we line up the defining features against the estimates, we see that the models that emphasized "student choice" as a key feature have the most positive outcome estimates in general, whereas those that adopted the other approaches had weaker outcomes in most cases. However, this pattern is plausibly explained by differences in research design, since we also find that the weakest analytic designs were associated with the most-positive outcomes, especially in the cases of Sharpstown, DCIS, and (to a lesser extent) the Philadelphia badges program. Meanwhile, the stronger analytic designs (Philadelphia overall, Adams 50, and Newfound) yielded mixed results. One reasonable interpretation of these data is that the estimates from the weaker designs are inflated by selection bias, though of course we cannot say that definitively. What we can say is that the current evidence, though imperfect, favors student choice and project-based learning models more than models that emphasize competency-based grading or flexible pacing. However, no single Project Mastery site serves as an exemplar of any particular approach, and none of the analyses here can claim to be truly causal. In other words, we can present the balance of evidence as it stands, but we do so knowing that the jury must remain out regarding the effects of competency-based models on students' academic performance.

---

16 Interviewees noted that, though DCIS-M and Sharpstown were schools chosen by parents, they were undersubscribed during the admission of the student cohorts included in this analysis, so were not actively turning students away.

CHAPTER SEVEN

Conclusion

Summary: What We Learned

This report presents findings from a mixed-method, multisite study of competency-based education in three pilot programs. The pilots, which were funded by the Bill & Melinda Gates Foundation's Project Mastery grant initiative, were carried out in the 2011–2012 and 2012–2013 academic years. They included 11 schools distributed across five school districts in four states. The pilot initiatives were heterogeneous in terms of both their scope and emphasis, but they shared two main features. First, all focused on technology-enriched models in secondary schools. Second, nearly all focused on urban or suburban schools that served a large share of low-income and minority students.

In each site, we collected interview and observation data about program components and implementation from project leaders, teachers, and relevant administrators and instructional coaches. Our first site visits were near the end of the first implementation year, in May 2012. Our second visits were in November 2012, during the fall of the second implementation year, and our final visits were conducted at the end of the second implementation year, in May 2013. In addition, we surveyed students in their Project Mastery classes in each site during May 2013.

Our surveys revealed that despite marked heterogeneity of the program features, students' experiences in the sites were reasonably uniform. The exception was DCIS, in which students reported disproportionately high levels of engagement, choice, and flexible pacing. It is noteworthy that the DCIS student surveys were conducted in classes that emphasized highly autonomous learning on yearlong projects of students' choosing, on par with a thesis or independent study. According to school staff, DCIS students were prepared to undertake such projects in 11th and 12th grades due to the school's emphasis in earlier grades on project-based learning applied to real-world problems.

Because the sites varied markedly in the dosage and rollout of their Project Mastery pilot interventions, as well as available data, our analysis of student outcomes also needed to vary by site. In all sites but Philadelphia, our outcomes analysis focused on competency-based education as the intervention of interest rather than strictly on the Project Mastery materials alone. Specifically, in Adams 50 and Newfound, we compared each site to others in its respective state using a synthetic comparison approach in which we weighted the comparison sites on a broad range of observable characteristics during the years before adoption of a competency-based approach. In the other Asia Society sites, all of which were relatively new schools of choice with no precompetency period, we compared the schools to demographically similar schools in the same states during each year of their operation. In Philadelphia, where some ninth-graders were exposed to the Project Mastery–funded curricula for a year while others were not exposed at all, we examined student formative assessments, attendance, and promotion for treated versus unexposed students. We also examined differential effects for Project Mastery students in Philadelphia based on participation in the afterschool digital filmmaking badges program offered in three of the schools.

Our findings varied across the sites. In Adams 50, we found that the conversion to a competency-based model was associated with lower mathematics performance on the state accountability test in each of the five years since the transition (by about 0.22 of a student-level standard deviation in 2013)—a difference that was significant at the 10 percent level. In Newfound, we found that conversion to a competency-based model was associated with increased performance of about 0.1 of a student standard deviation in reading, though this was not statistically significant. With regard to the other Asia Society sites, students in DCIS and Sharpstown markedly outperformed demographically similar peers in their respective states, and the differences were statistically significant. Of course, both were schools of choice, which makes it difficult to attribute these differences entirely to the competency-based models. In the other school of choice, DCIS-M, students performed similarly to their demographically comparable peers. In Philadelphia, Project Mastery students performed below their peers on the fall language arts assessment (after three months of exposure). They performed about the same as their peers on the winter assessment (after five months of exposure) and in terms of promotion to tenth grade the following year. Yet across the school year, they had attendance rates that were 1.4 percentage points higher than those not in Project Mastery classrooms—a statistically significant difference. Among those in the badges program after school, school attendance rates and promotion rates were higher than those of ninth-graders who were not in Project Mastery classrooms.

We noted that the most positive effects were seen in analyses that, for the most part, were most subject to selection bias. However, that is not to say that the effects were entirely due to selection bias—only that we cannot know empirically how much was and was not driven by student and/or teacher selection. In addition, the effects were generally more positive for sites that emphasized student choice and project-based learning over flexible pacing and mastery grading. Though we cannot empirically distinguish these program-type effects from research design effects, implementation data revealed that teachers were raising more concerns about equity in the flexible pacing and mastery grading sites than in the sites that emphasized project-based learning. In addition, students' access to the online tools purchased in Adams 50 appeared to be constrained by limited access to computer and tablet hardware, such that the district ultimately adopted a "Bring Your Own Device" policy.

It is important to note that Adams 50 and DCIS-M both faced lower and more varied levels of student achievement than the other sites, as well as high turnover of students, staff, and school leaders. They had adopted flexible pacing and mastery grading in an effort to mitigate the daunting achievement challenges their students faced. It therefore becomes difficult to wholly separate, in a qualitative and quantitative sense, the challenging conditions that gave rise to their models from the challenges that staff faced in implementing the models themselves. Put differently, these educators understood that traditional models of education had not made their students college and career ready, and as the staff in Adams 50 explained it, going back to a traditional schooling model was simply not viewed as an option.

Discussion: Placing the Findings in Context

The competency-based education movement aims to raise students' college readiness by ensuring that students do not advance through the education system without obtaining proficiency in the content required for college-and-career readiness. The movement attempts to accomplish this goal in three ways: (1) through a flexible pace of learning that meets students where they are; (2) through students' choices over their learning experiences and deliverables; and (3) by awarding credit based on evidence of proficiency—what students know and can do—rather than on the time they have spent studying a topic or the effort they put into doing so. The thinking is that this approach will not only motivate students to higher levels of achievement but will also improve the quality of the high school diploma as a signal of skill and knowledge, since students will not be able to obtain it until they have met a certain standard of learning.

In the literature on competency-based education, the relationship between the features of a competency-based approach and improved student achievement has been taken largely for granted, with effectiveness claims relying on anecdote more than on systematic studies. Meanwhile, extant research on the relationship between standards-based reform and high school exit examinations suggests that some students—especially low-income, urban, and minority students—become discouraged and leave school when they are required to meet state standards to graduate, and this appears especially true in states with more rigorous standards (Dee & Jacob, 2006; Holme et al., 2010; Papay, Murnane, and Willett, 2008, 2011; Warren, Jenkins, and Kulick, 2006). Some authors in this literature have speculated that allowing students to use alternative methods—such as portfolio assessments—to demonstrate that they have met standards may ameliorate the dropout effect (Holme et al., 2010). This rationale, though not well tested, lies at the heart of the competency-based movement and the Project Mastery initiatives.

Other econometric studies also raise questions about the rationale. For instance, we know that students who have passed the GED—a direct measure of knowledge and skill that supplants time-based credits—perform worse in the labor market than peers with identical levels of knowledge and skills (as measured by tests other than the GED) who have completed four-year high school diplomas (Heckman & Rubinstein, 2001). In their interpretation of this finding, Nobel-winning economist James Heckman and co-author Yona Rubinstein speculate that it is the "noncognitive" skills measured by the high school diploma—skills such as persistence, attendance, and working within the norms of an institution like high school—that account for the difference. The implications for competency-based education seem clear: As schools shift from time-based to competency-based models, they may do well to consider measures of effort, persistence, and citizenship as components of what it means to be "competent," in the sense of what students need to be ready for college and careers.

Several of the Project Mastery initiatives were consistent with the Heckman & Rubinstein (2001) lesson about defining competence broadly. The Asia Society sites, in particular, emphasized the notion of "global competence" as a feature of global citizenship. In their approaches to project-based learning, service learning, internships, and ELOs, students were required to demonstrate not just that they had learned a skill but that they could apply the skill to complex, real-world problems. This move toward authentic, applied assessment was also evident in the afterschool filmmaking badges program in Philadelphia, even though disagreements persisted over which types of evidence could be used to demonstrate mastery of the CCSS in ELA. Also, it was in the programs that placed the greatest emphasis on project-based learning and authentic tasks—Newfound, DCIS, Sharpstown, and the Philadelphia badges program—that we saw the most encouraging evidence of student performance and engagement, even though our models could not fully adjust for unobserved baseline differences between those who did and did not choose these programs.

Lessons for Policy, Partnerships, and Practice

Because our estimates of student outcome effects in each site represent correlational relationships rather than causal effects of particular models, we cannot make definitive or final pronouncements about the effects of competency-based education on students' achievement or rates of learning. We can only summarize the estimates we found in each program, and contextualize these estimates in terms of research design as well as program design and implementation. To make more-conclusive statements, future studies of competency-based education models would ideally assign students randomly to one or more models—for instance, through lottery-based admission to oversubscribed, competency-based schools. To that end, funders and policymakers might encourage or incentivize oversubscribed programs to employ lottery-based admission policies. Such policies are not only useful for research purposes; they also yield transparent and fair allocation of limited enrollment slots.

In the absence of true randomization, future studies of competency-based education models might continue to examine relative changes in student outcomes at schools before and after they adopt a competency-based approach, similar to the quasi-experimental approach that we employed in Adams 50 and Newfound. Or they might examine students' relative performance before and after entering competency-based schools or classrooms, similar to our approach in Philadelphia. These quasi-experimental designs are not immune to confounds, but when they are carefully conducted with adjustment for important covariates, they can still shed light on the relative performance of students who are and are not exposed to particular personalized learning models.[1] As more competency-based schools—and consequently, more studies—emerge, it will become easier for researchers to synthesize results across studies and thereby to draw farther-reaching conclusions about student outcomes under competency-based and other personalized learning approaches. In the meantime, our study contributes to that emerging effort, and our qualitative data offer implementation insights for educators who wish to implement or scale up these approaches. These include lessons for policy, partnerships, and practice.

Lessons for Policy

1. Competency-based education programs should be assessed on a variety of near-term and longer-term outcomes. This study illustrated that competency-based education models take time to develop and implement, and that benefits may not emerge immediately. The literature on implementing organizational change suggests that performance may decline, at least briefly, before it rebounds (Herold & Fedor, 2008; Jellison, 2006). In an education context, RAND's

[1] For instance, in a study of blended-learning charter schools that RAND is currently undertaking for the Bill & Melinda Gates Foundation, the authors are adjusting for possible selection bias using a national sample of test takers who are matched to the treatment students on prior test scores, test dates, and a variety of other school and student variables. In ongoing analyses, they will also conduct sensitivity tests that limit the comparison sample only to other schools of choice (Pane et al., 2013).

study of charter schools in eight states demonstrated that new schools of choice often take a year or two to stabilize before they begin to improve students' outcomes relative to the sending schools (Zimmer et al., 2009).

Among other sites, this study examined a new school of choice (DCIS-M) and a district that had transformed from a traditional to a competency-based system (Adams 50). It may take time for strong performance to emerge in such contexts. Moreover, by tailoring instruction to the level of each student, it is conceivable that some competency-based approaches would increase the engagement and persistence of students who still continued to struggle academically. For this reason, policymakers may wish to consider outcomes beyond just standardized test scores, insofar as student attendance, behavior, and/or persistence data are available. In Philadelphia, for instance, we found evidence of a positive Project Mastery effect on attendance, but available data did not permit us to examine similar outcomes in the other sites. Finally, important outcomes of interest are not always measurable in the near term, such as younger students' high school graduation rates, as well as their eventual college entry, remediation, and persistence rates. These are important metrics that competency-based education aims to influence, so a complete assessment of a competency-based approach should take them into account.

In Adams 50, for example, it would be valuable to examine high school completion rates for the graduating class of 2015 and beyond to the cohorts that preceded it, because the class of 2015 will be the first one held to a competency-based graduation standard. In other words, it may require at least seven years beyond the inception of the competency-based reform for the most salient findings to emerge. This illustrates why a long time horizon can be useful when considering whether a policy reform is working. It is also consistent with staff members' conviction in Adams 50 that they should stay the course and look for impact on longer-term outcomes such as college enrollment, remediation rates, and persistence.[2]

2. In a competency-based system, flexibly timed accountability tests may provide better measures of progress than fixed, annual tests. Because competency-based systems often favor teachers' subjective assessments of student skills, and because flexibility and choice may increase variation in students' exposure to required content, standardized accountability tests may be especially important for validating students' learning in competency-based systems. Even so, students working in a competency-based system may be studying material at different times than their peers in a traditional system. Even if they eventually learn the same content as their peers, poor alignment between the content they are studying and the content on which they are tested could result in lower test scores strictly as an artifact of timing.

To minimize this problem, students in a flexibly paced, competency-based system would ideally take grade-level accountability tests when they finished meeting learning targets for a specific grade level, rather than at a fixed point of time in the school year. This practice would have two main implications. First, students would take accountability tests that corresponded to the level of work they were studying, not to their age-based grade levels. Second, accountability tests would be administered as students were ready for them, on an individual basis, rather than to all students on a few designated days each year. Schools would still be held accountable for students' mastery of standards-based content, but school accountability criteria

---

[2] In Philadelphia, too, a longer horizon could prove useful, since ninth-graders who were exposed to the Educurious units in 2012–2013 will not take another state accountability assessment in ELA until the spring of 2015 (the Keystone English Literature test), and these students will not finish high school until 2016.

could focus on the average rate at which students advanced between levels (a growth metric) in addition to students' average academic proficiency levels at a given age threshold (a proficiency metric).[3]

In Adams 50, where flexible pacing was a core emphasis, the first criterion was satisfied but not the second. Because students were assigned to academic performance levels rather than to age-based grade levels, their state accountability test scores pertained to the content levels they were studying and not to their ages. However, because students could progress between content levels throughout the year, it was possible that within content levels, students were taking state accountability tests on material they had not yet learned or that they had learned some months previously. Such imperfect alignment between covered content and tested materials can also happen in a traditional classroom setting, but teachers likely have less control over students' exposure to tested content in an environment where students advance at different paces.

Lessons for Partnerships

3. In collaborating with technology developers, schools should negotiate favorable terms and anticipate technical challenges. All of the Project Mastery grants involved collaborations between technology developers and educational organizations. These included partnerships with developers of online LMSs in all of the initiatives, and partnerships with curriculum developers in Adams 50 and Philadelphia. These collaborations revealed two key lessons. First, school districts that agree to pilot technology products with the students they serve are providing a vital context in which developers can test and refine their products. The terms of these partnerships must reflect the indispensability of the schools as pilot sites. In other words, developers would ideally provide pilot sites with long-term access to piloted tools and with discounts on upgrades and expansions, because their businesses are critically dependent on pilot districts and schools as product testers. In Philadelphia, the Educurious pilot agreement did not allow the district to scale the materials they had piloted and on which their teachers had provided detailed feedback. This limited the district's ability to expand the program beyond the pilot teachers and beyond a two-year test period. In contrast, Adams 50 reportedly retained perpetual rights to the Operation Spacewolf games it helped to develop and pilot, which will allow the district to build off the games' popularity during the pilot year. Moreover, Adams 50's essentially pro bono relationship with the LevelUp developers at EffectiveSC suggested the potential for developers to view their alpha-testing districts not just as potential clients but as true thought partners in the development process.

The second lesson that emerged from schools' collaboration with developers was that technical challenges are typical, and that these collaborations intrinsically require schools to test imperfect tools. This was apparent in the limitations that teachers found in the Educate and ShowEvidence learning systems, and in the interoperability problems Philadelphia encountered between its LMS and SchoolNet. In approaching these partnerships, educators should perhaps view them—and present them to teachers—as clinical trials that may benefit future cohorts rather than as just-in-time solutions. Those who expect the latter are likely to become frustrated and perhaps to disengage from using the tools.

---

[3] Such a system would not preclude the use of assessments for teacher accountability either, insofar as measures of teachers' value-added (i.e., impact on learning) were adjusted for the length of time students spent in a particular teacher's class.

4. Collaboration between sites and funders should consider local infrastructure and capacity. In Adams 50, the rationale for purchasing online resources in the form of games and virtual courses seemed clear. In a system of individually paced learning, students needed access to personalized instructional materials, and online resources could plausibly increase their options beyond the file cabinets of worksheets that some teachers said they had come to rely on. However, the district's reliance on the purchased materials (as well as the free online resources their teachers embraced) was greatly impeded by a lack of computers that students could use to go online. The result is that students were asked to bring their own devices to school, exacerbating equity and logistical challenges. Given those circumstances, a critical question is how Adams 50 might have ensured adequate hardware—and possibly electrical upgrades if needed—as a precursor to the purchase of online instructional tools. District staff noted that philanthropic grants often did not cover basic infrastructure, but there may have been other options. For instance, it is possible that the district and Foundation might have leveraged their relationship to find funding from private matching funds or even state funds. There may be an ongoing need for funders interested in competency-based education to help support a basic infrastructure in districts that are working to implement it.

Lessons for Practice

5. Increased student autonomy calls for engagement through skillful teaching. Though approaches to competency-based education were diverse across the Project Mastery grantees, all of the sites gave students more autonomy than one might find in a traditional classroom. In Adams 50, students completed work and demonstrated proficiency at an individualized pace, often receiving instruction through print materials or online. In the Asia Society sites and in Philadelphia's Educurious classrooms and badging program, students had choices about the educational projects they undertook and how they carried them out. Teachers in all of the sites were therefore challenged with preparing students to thrive under increased responsibility and freedom. When we surveyed students, we found the highest levels of engagement when students were undertaking self-directed, long-term projects connected to real-world contexts, as in the DCIS Passages and Portfolio classes. However, their teachers had to create the environment in which students could undertake self-directed, project-based learning successfully. To that end, Asia Society gave its teachers common performance outcomes but expected teachers to create project-based units applicable to real-world problems. In other words, instructional creativity was expected. Philadelphia was moving in a similar direction by having its Philadelphia Writing Project teachers build Educurious-style units for the district, but these efforts remained confined to a small number of teachers. In Adams 50, teachers' creativity was also called upon, but it went largely toward coordinating students' progress through diverse assessment sequences and managing proficiency data from many sources. Across sites, teachers described the challenge of motivating learners—especially struggling learners—to thrive under semi-autonomous conditions. As one Adams 50 teacher noted, there is likely no way around this challenge; it is intrinsic to teaching. Still, given the need to help young people manage their learning wisely, the Project Mastery sites remind us that competency-based education does not reduce the importance of skillful teaching. It heightens it.

6. Competency-based education systems must be vigilant about equity. In a world of flexible and personalized learning, we might expect students' daily educational experiences to become more heterogeneous than under a system of synchronous instruction and group-based learning. Moreover, we might expect students who have fallen behind their peers academically in traditional classrooms to struggle even more when evaluations reduce emphasis on effort and increase emphasis on knowledge and skill. Neither of these shifts is accidental; rather, they are some of the mechanisms by which competency-based education is intended to transform K–12 education. Yet, because they have potential to exacerbate rather than narrow achievement gaps, it is important that educators implementing competency-based models direct additional support and guidance toward students who need the most academic help. It is similarly important that educators monitor the relative performance of students with lower baseline skills and intervene with students who seem to be losing ground. Given that the competency-based education movement aims to improve the college- and career-readiness of students who have traditionally fallen behind, it is incumbent on those in the movement to ensure that it mitigates rather than worsens achievement gaps in America.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2007). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Cambridge, MA: National Bureau of Economic Research.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review,* 93(1), 113–132.

Adams County District 50. (2011). *Assessment,* web page. As of March 30, 2014: http://www.cbsadams50.org/assessment/

Asia Society. (2014a). *Mission and history,* web page. As of May 20, 2014: http://asiasociety.org/about/mission-history

Asia Society. (2014b). *SAGE advice,* web page. As of April 2, 2014: http://asiasociety.org/advice/sage-advice

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological), 57*(1), 289–300.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust difference-in-difference estimates? *The Quarterly Journal of Economics, 119*(1), 249–275.

Betebenner, D. W., & Linn, R. L. (2010). *Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability.* Princeton, NJ: Educational Testing Service.

Bill & Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study.* Seattle, WA.

Bloom, B. S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16.

Carnegie Foundation for the Advancement of Teaching. (2013). *50-state scan of course credit policies.* Menlo Park, CA: Carnegie Foundation for the Advancement of Teaching.

CCSS Initiative—*See* Common Core State Standards Initiative.

Chicago Consortium on School Research. (2012). *Survey documentation,* web page. As of August 10, 2012: http://ccsr.uchicago.edu/surveys/documentation

Chicago Public Schools. (2013, August 14). *CPS students set record high graduation rate of 65.4 percent for school year 2012–2013,* web page. As of September 7, 2013: http://www.cps.edu/News/Press_releases/Pages/PR1_08_14_2013.aspx

Common Core State Standards Initiative. (2014). *English language arts standards: Writing: Grade 9–10,* web page. As of April 2, 2014: http://www.corestandards.org/ELA-Literacy/W/9-10/

Dee, T. S., & Jacob, B. A. (2006). *Do high school exit exams influence educational attainment or labor market performance?* Cambridge, MA: National Bureau of Economic Research.

Dewey, J. (1902). *The child and the curriculum.* Chicago: University of Chicago Press.

Edutopia (Producer). (2011). Salman Khan on liberating the classroom for creativity. *Big Thinkers Series.* As of August 30, 2013:
http://www.edutopia.org/salman-khan-academy-flipped-classroom-video

Funk, M. J., Westreich, D., Wiesen, Stumer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *Practice of Epidemiology, 173*(7), 761–767.

Grossman, T., & Shipton, S. (2012). *State strategies for awarding credit to support student learning.* Washington, DC: National Governors Association.

Guskey, T. R., & Gates, S. L. (1985). *A synthesis of research on group-based mastery learning programs.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Guskey, T. R., & Gates, S. L. (1986, May). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. *Educational Leadership,* 73–80.

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E. B., Supovitz, J. A., Wayman, J. C., et al. (2009). *Using student achievement data to support instructional decision making.* Washington, DC: What Works Clearinghouse, U.S. Department of Education.

Hamilton, L. S., Stecher, B. M., & Yuan, K. (2009). *Standards-based reform in the United States: History, research, and future directions.* Santa Monica, CA: RAND Corporation, RP-1384. As of June 15, 2014:
http://www.rand.org/pubs/reprints/RP1384.html

Hanford, E., & Smith, S. (Producer). (2013). *One child at a time: Custom learning in the digital age,* web page. As of September 6, 2013:
http://americanradioworks.publicradio.org/features/personalized-learning/

Haystead, M. W. (2010). *RISC vs. non-RISC schools: A comparison of student proficiencies for reading, writing, and mathematics.* Englewood, CO: Marzano Research Laboratory.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *The American Economic Review, 91*(2), 145–149.

Herold, D. M., & Fedor, D. B. (2008). *Change the way you lead change: Leadership strategies that really work.* Stanford, CA: Stanford University Press.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2007). *Empirical benchmarks for interpreting effect sizes in research.* MDRC Working Paper on Research Methodology. New York: MDRC.

Holme, J. J., Richards, M. P., Jimerson, J. B., & Cohen, R. W. (2010). Assessing the effects of high school exit examinations. *Review of Educational Research, 80*(4), 476–526.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226–244.

Jellison, J. (2006). *Managing the dynamics of change.* New York: McGraw-Hill.

Koretz, D., Klein, S. P., McCaffrey, D. F., & Stecher, B. M. (1994). *Interim report: The reliability of Vermont portfolio scores in the 1992–93 school year.* Santa Monica, CA: RAND Corporation, RP-260. As of June 15, 2014:
http://www.rand.org/pubs/reprints/RP260.html

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 1990*(60), 265–299.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1979). A meta-analysis of outcome studiesof Keller's Personalized System of Instruction. *American Psychologist, 34,* 307–318.

Lewis, M. W., Adamson, D. M., Steele, J. L., Faxon-Mills, S., Rudnick, M., & Eden, R. (2013). *Proficiency-based pathways: A vision and critical elements for achieving it.* RAND Corporation research; not available to the general public.

Lindsay Unified School District. (2014). *Performance-based system,* web page. As of March 10, 2014: http://www.lindsay.k12.ca.us/District/Department/689-Performance-based-System/13566-Untitled.html

Loadman, W. E., Moore, R. J., Ren, W., Zhu, J., Zhao, J., & Lomax, R. G. (2011). *Striving Readers Year 5 project evaluation report: Ohio: An addendum to the Year 4 report.* Columbus, OH: The Ohio State University.

McCombs, J. S., Kirby, S. N., & Mariano, L. T. (Eds.). (2009). *Ending social promotion without leaving children behind: The case of New York City.* Santa Monica, CA: RAND Corporation, MG-894-NYCDOE. As of June 15, 2014:
http://www.rand.org/pubs/monographs/MG894.html

Meier, D. (1995). *The power of their ideas: Lessons for America from a small school in Harlem.* Boston: Beacon Press.

Meier, D. (2002). *In schools we trust: Creating communities of learning in an era of testing and standardization.* Boston: Beacon Press.

Merryman, S. (2005). *xtregre2.ado.* Washington, DC: Risk Management Agency, U.S. Department of Agriculture.

Nagaoka, J., & Roderick, M. (2004). *Ending social promotion: The effects of retention.* Chicago: Chicago Consortium on School Research.

National Institute of Standards and Technology. (2002, July 13, 2011). *Baldrige Performance Excellence Program: Chugach School District,* web page. As of March 1, 2014:
http://www.nist.gov/baldrige/chugach.cfm

NBC News and the Hechinger Report. (2013). *Carpe Diem: Seize the digital revolution,* web page. As of September 6, 2013:
http://www.educationnation.com/casestudies/carpediem/NBCCaseStudy_CarpeDiem.pdf

Nellie Mae Education Foundation. (2011, November 18). *Proficiency-based pathways initiative one-pagers.* Paper presented at the Proficiency-Based Pathways: Sharing Our Work Convening, Amherst, MA.

New Hampshire Department of Education. (2012). *New Hampshire high school redesign: Narrative history and related documents,* web page. As of April 29, 2013:
http://www.education.nh.gov/innovations/hs_redesign/background.htm

New Hampshire Department of Education. (2013). *New Hamsphire Common Core State Standards–aligned English language arts and literacy comptencies.* Concord, NH: New Hampshire Department of Education.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis, 36*(2), 127–144.

Pane, J. F., Hamilton, L. S., Steiner, E. D., Rudnick, M., Steele, J., & Sales, A. (2013). *Blended and competency-based learning approaches in the Next Generation Learning Challenges (NGLC) Wave IIIa Schools: First year implementation and effects.* RAND Corporation research; not available to the general public.

Papay, J. P., Murnane, R. J., & Willett, J. B. (2008). *The consequences of high school exit examinations for struggling low-income urban students: Evidence from Massachusetts.* Cambridge, MA: National Bureau of Economic Research.

Papay, J. P., Murnane, R. J., & Willett, J. B. (2011). *High-school exit examinations and the schooling decisions of teenagers: A multi-dimensional regression-discontinuity analysis.* Cambridge, MA: National Bureau of Economic Research.

Pennsylvania Department of Education. (2013). *Awarding credit to support student learning: A report to the Governor of Pennsylvania.* Harrisburg, PA.

Priest, N., Rudenstine, A., Weisstein, E., & Gerwin, C. (2012). *Making mastery work: A close-up view of competency education.* Quincy, MA: Nellie Mae Education Foundation.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). London: Sage Publications.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Silvernail, D. L., & Stump, E. K. (2012). *More efficient high schools in Maine: Emerging student-centered learning communities.* Gorham, ME: University of Southern Maine.

Silvernail, D. L., Stump, E. K., Duina, A. A., & Gunn, L. M. (2013). *Preliminary implementation of Maine's proficiency-based diploma program.* Gorham, ME: University of Southern Maine.

Silvernail, D. L., Stump, E. K., McCafferty, A. S., & Hawes, K. M. (2014). *Implementation of a proficiency-based diploma system in Maine: Phase II—District-level analysis.* Gorham, ME: University of Maine.

Sizer, T. R. (2004). *Horace's compromise: The dilemma of the American high school.* New York: Mariner Books.

Slavin, R. E. (1984). Mastery learning reconsidered. *Review of Educational Research, 57,* 175–213.

Soares, L. (2012). *A "disruptive" look at competency-based education: How the innovative use of technology will transform the college experience.* Washington, DC: Center for American Progress.

Steiner, P. M., Cook, T. D., & Shadish, W. R. (2011). On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. *Journal of Educational and Behavioral Statistics, 36*(2), pp. 213–236.

Sturgis, C. (2012). *The art and science of designing competencies: A CompetencyWorks issue brief.* Vienna, Va.: International Association for K–12 Online Learning.

Sturgis, C., & Patrick, S. (2010). *When success is the only option: Designing competency-based pathways for next generation learning.* Vienna, VA: International Association for K–12 Online Learning.

Swanson, C. B. (2009). *Cities in crisis 2009: Closing the graduation gap.* Bethesda, MD: Editorial Projects in Education.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221. doi: 10.1080/00461520.2011.611369

Warren, J. R., Jenkins, K. N., & Kulick, R. B. (2006). High school exit examinations and state-level completion and GED rates, 1975 through 2002. *Educational Evaluation and Policy Analysis, 28*(2), 131–152.

What Works Clearinghouse. (2009). *Read 180* (WWC Intervention Report). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

What Works Clearinghouse. (2014). *Procedures and standards handbook (Version 3.0).* Washington, DC: Institute of Education Sciences, U. S. Department of Education.

Young Women's Leadership Charter School of Chicago. (2013). *The facts: YWLCS academic achievement,* web page. As of September 5, 2013:
http://www.ywlcs.org/academics/StudentAcademicAchievementsv2.php

Zimmer, R., Gill, B., Booker, K., Lavertu, S., Sass, T. R., & Witte, J. F. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration, and competition.* Santa Monica, CA: RAND Corporation, MG-869-BMG. As of June 15, 2014:
http://www.rand.org/pubs/monographs/MG869.html

In 2011, the Bill & Melinda Gates Foundation created the Project Mastery grant program to support competency-based education initiatives in large school systems that serve a high proportion of disadvantaged youth. Competency-based education meets students where they are academically, provides students with opportunities for choice, and awards credit for evidence of learning, not for the time students spend studying a subject. The Foundation asked RAND to evaluate these efforts in terms of implementation, students' experiences, and student performance. This report presents final results from that evaluation, offering an overview of competency-based education and the Project Mastery grant projects and describing the implementation of competency-based educational features under each project. The report concludes with six lessons for policy, partnerships, and practice.

RAND EDUCATION

www.rand.org

$34.95

ISBN-10 0-8330-8725-8
ISBN-13 978-0-8330-8725-6

53495

9 780833 087256

RR-732-BMGF