Improving Teaching Effectiveness

IMPLEMENTATION

The INTENSIVE PARTNERSHIPS *for* EFFECTIVE TEACHING
Through 2013–2014

BRIAN M. STECHER
MICHAEL S. GARET
LAURA S. HAMILTON
ELIZABETH D. STEINER
ABBY ROBYN
JEFFREY POIRIER
DEBORAH HOLTZMAN
ELEANOR S. FULBECK
JAY CHAMBERS
ILIANA BRODZIAK DE LOS REYES

RAND CORPORATION

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

For more information on this publication, visit www.rand.org/t/RR1295

Library of Congress Cataloging-in-Publication Data is available for this publication.
ISBN: 978-0-8330-9221-2

Published by the RAND Corporation, Santa Monica, Calif.
© Copyright 2016 RAND Corporation
RAND® is a registered trademark.

Cover: Teacher Standing in Front of a Class of Raised Hands, Digital Vision.

Print and Electronic Distribution Rights

The trademark(s) contained herein is protected by law. This work is licensed under a Creative Commons Attribution 4.0 International License. All users of the publication are permitted to copy and redistribute the material in any medium or format and transform and build upon the material, including for any purpose (including commercial) without further permission or fees being required. For additional information, please visit http://creativecommons.org/licenses/by/4.0/.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND
Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

Preface

The Bill & Melinda Gates Foundation launched the Intensive Partnerships for Effective Teaching the 2009–2010 school year. After careful screening, the foundation identified seven Intensive Partnership sites—three school districts and four charter management organizations (CMOs)—to implement strategic human-capital reforms over a six-year period.[1] The foundation also selected the RAND Corporation and its partner the American Institutes for Research (AIR) to evaluate the Intensive Partnership efforts. The RAND/AIR team is conducting three interrelated studies examining the implementation of the reforms, the reforms' impact on student outcomes, and the extent to which the reforms are replicated in other districts.

The evaluation began in July 2010 and collected its first wave of data during the 2010–2011 school year; it will continue through the 2015–2016 school year and produce a final report in 2017. During this period, the RAND/AIR team is producing a series of internal progress reports for the foundation and the Intensive Partnership sites. We have also produced two journal articles:

- "Disentangling Disadvantage: Can We Distinguish Good Teaching from Classroom Composition?" (Gema Zamarro, John Engberg, Juan Esteban Saavedra, and Jennifer Steele, *Jour-*

[1] We use the word *site* to describe the three school districts and the four charter management organizations that received funding from the foundation to implement the Intensive Partnerships initiative.

nal of Research on Educational Effectiveness, Vol. 8, No. 1, 2015, pp. 84–111)
- "Implementing Measures of Teacher Effectiveness," (Brian Stecher, Mike Garet, Deborah Holtzman, and Laura Hamilton, *Phi Delta Kappan*, Vol. 94, No. 3, November 2012, pp. 39–43).

In addition, we have produced interim working papers for selected audiences:

- *How Are School Leaders and Teachers Allocating Their Time Under the Partnership Sites to Empower Effective Teaching Initiative?* (Jay Chambers, Iliana Brodziak de los Reyes, Antonia Wang, and Caitlin O'Neil, Santa Monica, Calif.: RAND Corporation, WR-1041-1-BMGF, 2014)
- *How Much Are Districts Spending to Implement Teacher Evaluation Systems? Case Studies of Hillsborough County Public Schools, Memphis City Schools, and Pittsburgh Public Schools* (Jay Chambers, Iliana Brodziak de los Reyes, and Caitlin O'Neil, Santa Monica, Calif.: RAND Corporation, WR-989-BMGF, 2013)
- *Using Teacher Evaluation Data to Inform Professional Development in the Intensive Partnership Sites* (Laura S. Hamilton, Elizabeth D. Steiner, Deborah Holtzman, Eleanor S. Fulbeck, Abby Robyn, Jeffrey Poirier, and Caitlin O'Neil, Santa Monica, Calif.: RAND Corporation, WR-1033-BMGF, 2014)
- *Trends in the Distribution of Teacher Effectiveness in the Intensive Partnerships for Effective Teaching* (Jennifer L. Steele, Matthew D. Baird, John Engberg, and Gerald Hunter, Santa Monica, Calif.: RAND Corporation, WR-1036-BMGF, 2014)
- *Introduction to the Evaluation of the Intensive Partnership for Effective Teaching (IP*) (Brian M. Stecher and Michael Garet, Santa Monica, Calif.: RAND Corporation, WR-1034-BMGF, 2014)
- *Teacher Performance Trajectories in High and Lower-Poverty Schools* (Zeyu Xu, Umut Özek, and Michael Hansen, Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research, Working Paper 101, updated March 2014)

- *Portability of Teacher Effectiveness Across School Settings* (Zeyu Xu, Umut Özek, and Matthew Corritore, Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research, Working Paper 77, June 2012).

The present report summarizes the implementation of the initiative from 2010 through 2014, and it should be of interest to researchers, policymakers, and practitioners who want to understand the potential benefits and challenges of adopting new teacher-evaluation systems and related reforms. Two companion reports present our findings on the initiative's impact on student outcomes and on teaching effectiveness.

Contents

Preface iii
Figures and Tables ix
Summary xiii
Acknowledgments xxix
Abbreviations xxxi

CHAPTER ONE
Introduction 1
The Intensive Partnership for Effective Teaching: Improving Outcomes for Low-Income and Minority Students 1
Launching the Intensive Partnership Initiative 5
Introduction to the Intensive Partnership Sites 9
The College-Ready Promise 12
Approach to Evaluating Intensive Partnership Reform Implementation 14
Limitations of This Study 18
Overall Expenditures on the Intensive Partnership Initiative 19
Organization of the Report 21

CHAPTER TWO
Teacher Evaluation 23
Teacher-Evaluation Lever Implementation 24
Distributions of Teaching-Effectiveness Ratings 28
Teacher and School-Leader Perspectives on Teacher Evaluation 28
Perceptions of the Fairness and Accuracy of the Evaluation Results 44
Cost of Teacher Evaluation 46
Summary 55

CHAPTER THREE
Staffing 57
Staffing Lever Implementation 57
Teacher and School-Leader Perspectives on the Staffing Lever 60
Summary 73

CHAPTER FOUR
Professional Development 75
Professional-Development Lever Implementation 75
Teacher and School-Leader Perspectives on the Professional-Development Lever 81
Summary 86

CHAPTER FIVE
Compensation and Career Ladders 89
Compensation and Career-Ladder Lever Implementation 89
Teacher and School-Leader Perspectives on the Compensation and Career-Ladder Lever 95
Summary 104

CHAPTER SIX
Summary and Conclusions 107
Summary 107
Discussion 113
Future Analyses 118

APPENDIXES
A. Methods for Interview Data Collection and Analysis 119
B. Methods for Coding Implementation Status 123
C. Methods for Survey Data Collection and Analysis 129

References 135

Figures and Tables

Figures

S.1. The Intensive Partnership Theory of Action xiv
S.2. Average Proportion of Intensive Partnership Levers Implemented Between Spring 2010 and Spring 2014 xvi
S.3. Percentage of Teachers Reporting That Evaluation Components Are Valid to a Large or Moderate Extent, School Year 2013–2014 xviii
S.4. Percentage of School Leaders Agreeing That "in the Long Run, Students Will Benefit from the Teacher-Evaluation System," Spring 2011, Spring 2013, and Spring 2014 xix
S.5. Percentage of Teachers Agreeing with Statements About Career-Ladder Positions in 2014 xxiv
1.1. The Intensive Partnership Theory of Action 2
1.2. Intensive Partnership Initiative Expenditures Broken Out by Funding Source, November 2009 Through June 2014 20
2.1. Proportion of the Teacher-Evaluation Lever Implemented, Spring 2010 to Spring 2014 25
2.2. Effectiveness Rating Distributions, by Site, School Years 2011–2012, 2012–2013, and 2013–2014 29
2.3. Percentage of Teachers Reporting That Evaluation Components Are Valid to a Large or Moderate Extent, School Year 2013–2014 32
2.4. Percentage of Teachers Agreeing with Statements About the Effects of Evaluation on Their Teaching, School Year 2013–2014 41

2.5. Percentage of Teachers Agreeing That, "in the Long Run, Students Will Benefit from the Teacher-Evaluation System," Springs of 2011–2014 42
2.6. Percentage of School Leaders Agreeing That, "in the Long Run, Students Will Benefit from the Teacher-Evaluation System," Springs of 2012–2014 43
2.7. Percentage of Teachers Agreeing with Statements About the Fairness of the Evaluation System, School Year 2013–2014 45
2.8. Percentage of Teachers Reporting That Their Prior Year's Overall Evaluation Ratings Were Moderately or Highly Accurate, by Effectiveness Rating, School Year 2013–2014 47
2.9. Funding Sources for Implementing the Teacher-Evaluation Systems in Hillsborough County Public Schools, Shelby County Schools, and Pittsburgh Public Schools, November 2009 to June 2012 49
2.10. Estimated Evaluation System Total Cost per Pupil for School Years 2010–2011 and 2011–2012 54
3.1. Proportion of the Staffing Lever Implemented, Spring 2010 to Spring 2014 59
3.2. Percentage of School Leaders Agreeing That "the Processes by Which Teachers Are Hired to My School Work Well," Springs of 2013 and 2014 62
3.3. Percentage of School Leaders Agreeing That, "More Often Than Is Good for My School, Good Teachers Leave My Staff Because They Perceive Better Opportunities Elsewhere," 2014 64
3.4. Percentage of School Leaders Reporting Satisfaction with the Performance of Teachers Who Transferred to Their Schools, Springs of 2013 and 2014 65
3.5. Percentage of School Leaders Agreeing with Statements About Site Tenure Policies, Springs of 2011–2014 67
3.6. Percentage of School Leaders Reporting That One or More Teachers Experienced Various Outcomes, Springs of 2013 and 2014 71
4.1. Proportion of the Professional-Development Lever Implemented, Spring 2010 to Spring 2014 77

4.2. Percentage of Teachers Responding to the Question, "Has Support (Coaching, Professional Development, etc.) Been Made Available to You to Address the Needs Identified by Your Evaluation Results?" School Years 2012–2013 and 2013–2014 82
4.3. Percentage of Teachers Agreeing That "I Have Had Easy Access to a Catalog of Professional Development Opportunities Aligned with My District/CMO Teacher Observation Rubric," School Years 2012–2013 and 2013–2014 83
4.4. Percentage of School Leaders Agreeing with Statements About Coaching and Mentoring the Whole Staff, School Year 2013–2014 85
4.5. Percentage of Teachers Indicating the Usefulness of School-Based Workshops and In-Services, District- or Charter Management Organization–Based Workshops and In-Services, and School-Based Teacher Collaboration, 2014 87
5.1. Proportion of Compensation and Career-Ladder Lever Implemented, Spring 2010 to Spring 2014 91
5.2. Percentage of Teachers Agreeing That "a Teacher's Base Pay Should Be Based on Seniority," 2014 97
5.3. Percentage of Teachers Agreeing That "the Way Compensation Decisions Are Made in My District/CMO Is Fair to Most Teachers," Springs of 2011, 2013, and 2014 ... 98
5.4. Percentage of Teachers and School Leaders Responding to "This Year, Does Your District/CMO Have in Place a 'Career Ladder' for Teachers, or Specialized Instructional Positions That Teachers May Take on If They Are Considered Qualified?" 101
5.5. Percentage of Teachers Agreeing with Statements About Career-Ladder Positions, 2014 103
5.6. Percentage of Teachers and School Leaders Agreeing That "the Teachers Who Hold Higher-Level Positions at My School Deserve the Additional Compensation (Bonuses or Higher Salaries) They Are Receiving," 2014 105

Tables

1.1. Characteristics of the Intensive Partnership Sites, 2009–2010 and 2013–2014 School Years 10
2.1. Overview of Expenditures on the Teacher-Evaluation Systems in Hillsborough County Public Schools, Shelby County Schools, and Pittsburgh Public Schools, November 2009 to June 2012 50
2.2. Teacher-Evaluation System and Overall Intensive Partnership Initiative Expenditures, per Pupil and Percentages 53
A.1. Number of Central-Office Administrators and Stakeholders Interviewed 119
A.2. Number of School-Level Staff Interviewed 121
B.1. Levers, Practices, and Definitions 124
C.1. Number of Schools Surveyed 130
C.2. Number of Teachers and School Leaders Surveyed 130
C.3. Teacher Response Rates, Surveys Completed, and Teachers Sampled 132
C.4. School-Leader Response Rates, Surveys Completed, and Leaders Sampled 133

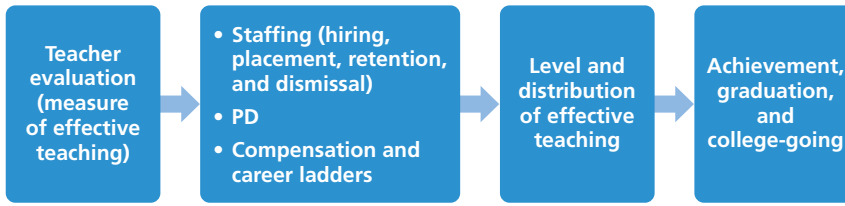
Summary

To improve the nation's education system through more-effective classroom teaching, in the 2009–2010 school year, the Bill & Melinda Gates Foundation announced four Intensive Partnership for Effective Teaching sites. The Intensive Partnership initiative is based on the premise that efforts to improve instruction can benefit from high-quality measures of teaching effectiveness. In its Measures of Effective Teaching study, the foundation found that it is possible to build a high-quality measure of teaching effectiveness by combining information about student achievement growth, direct observation of teaching practice, and student feedback.

The Intensive Partnership initiative seeks to determine whether each school can implement a high-quality measure of teaching effectiveness and use it to support and manage teachers in ways that improve student outcomes. This approach is consistent with broader national trends in which performance-based teacher evaluation was increasingly being mandated at state and local levels.

Figure S.1 shows the theory of action for the Intensive Partnership initiative. The process starts with adoption of valid measures of teaching effectiveness. These measures are incorporated into a management system to improve the teacher workforce over time. The elements of this system include staffing policies to improve the effectiveness of new teachers, professional-development (PD) practices to promote the effectiveness of current teachers, and compensation and differentiated careers for retaining highly effective teachers. Together, these practices are designed to improve overall teaching effectiveness and promote a

Figure S.1
The Intensive Partnership Theory of Action



SOURCE: Bill & Melinda Gates Foundation.
NOTE: Career-ladder positions are those for teachers with extra responsibilities and extra pay.
RAND *RR1295-S.1*

more-equitable distribution of effective teaching across schools and students. This, in turn, should lead to better outcomes for all students.

To test the theory in practice, the foundation sought partnership sites. It selected three school districts—Hillsborough County Public Schools (HCPS) in Florida, Memphis City Schools in Tennessee (which merged with Shelby County Schools, or SCS, during the course of the initiative), and Pittsburgh Public Schools (PPS) in Pennsylvania—chosen from the population of districts enrolling at least 25,000 students, with at least 40 percent of students eligible for free or reduced-price lunch, and located in states that did not have low thresholds for granting tenure. The foundation also selected four charter management organizations (CMOs)—Alliance College-Ready Public Schools, Aspire Public Schools, Green Dot Public Schools, and the Partnerships to Uplift Communities (PUC Schools), all in California—representing a mix of large and midsize schools with diverse populations.

The foundation pledged $290 million to support the initiative over the seven-year period from 2009 to 2016; from November 2009 through June 2014, it awarded more than $160 million to the seven sites. The foundation also required that each site provide matching funds from other sources, such as local foundations or federal grants, to support the initiative. The sites also allocated some general-fund resources to support the initiative. Total expenditures have ranged from $3.1 million in PUC Schools to $144 million in HCPS, while total expenditures per pupil ranged from $473 in Green Dot to $2,149

in PPS. Foundation funds ranged from 40 percent of initiative funds in HCPS and Aspire to 74 percent in Alliance. The foundation also stayed engaged with the sites providing expertise, convening cross-site meetings for learning and sharing and assigning each an active program officer to help keep the initiative on track.

To evaluate Intensive Partnership implementation, we interviewed annually central-office staff at each site and teachers and other staff in a sample of schools for each site. We also used data from annual teacher and school-leader surveys and documents that the sites and the foundation provided. This report summarizes the implementation status of key reform elements at each site when the Intensive Partnership initiative launched through the spring of 2014.

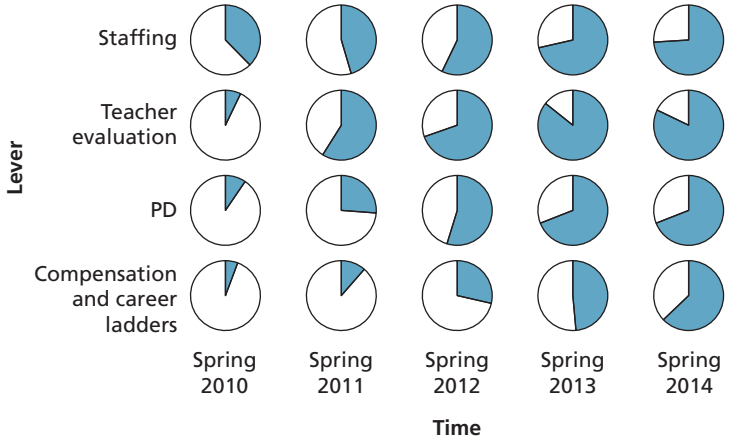
It is important to note that the Intensive Partnership initiative was not implemented in a vacuum, i.e., the sites all had to deal with the usual changing conditions and pressures under which schools and school systems must operate. Every site faced changes in state policy (e.g., declining budgets, new accountability rules), changes in local context (e.g., enrollment shifts, union concerns, more-competitive teacher labor markets), and changes in leadership (e.g., new superintendents, new senior staff) that influenced their plans and affected their attention to the Intensive Partnership initiative.

Implementation

To summarize the level of implementation in each site and its progress over time, we identified specific policies and practices sites adopted as part of their Intensive Partnership reforms and grouped them into four broad categories or levers, corresponding to the elements of the Intensive Partnership theory of action: (1) teacher evaluation, (2) staffing, (3) PD, and (4) compensation and career ladders.

Figure S.2 shows the average proportion of practices associated with each lever that were implemented annually from the spring of 2010 to the spring of 2014. Sites did not necessarily plan to implement all of the practices associated with each lever, but they were expected to implement many of them. The important features to notice when

Figure S.2
Average Proportion of Intensive Partnership Levers Implemented Between Spring 2010 and Spring 2014



RAND *RR1295-S.2*

looking at the figure are when actions began, how quickly they progressed, and when they attained stability. The report presents more details about the implementation of each of the four levers in each of the sites. The sites were expected to enact a cohesive set of teacher-evaluation, staffing, PD, and compensation and career-ladder policies that were compatible with state laws and local contexts. Although the sites shared the same goal of improving student achievement through improving teaching, there was no expectation that the sites would enact the same policies on the same schedule. Therefore, it would be incorrect to interpret the Intensive Partnership initiative as a competition and to overemphasize comparisons among the sites.

Teacher Evaluation

Each of the sites took about two years to design and implement its teaching-effectiveness measure, including engaging stakeholders, defining the component measures, training observers to rate classroom practice accurately and reliably, determining weights for the compos-

ite measure, and producing the effectiveness scores. HCPS and PPS already had some measures in development at the time of the initiative launch; by the spring of 2013, all sites had most of the relevant measures in place.

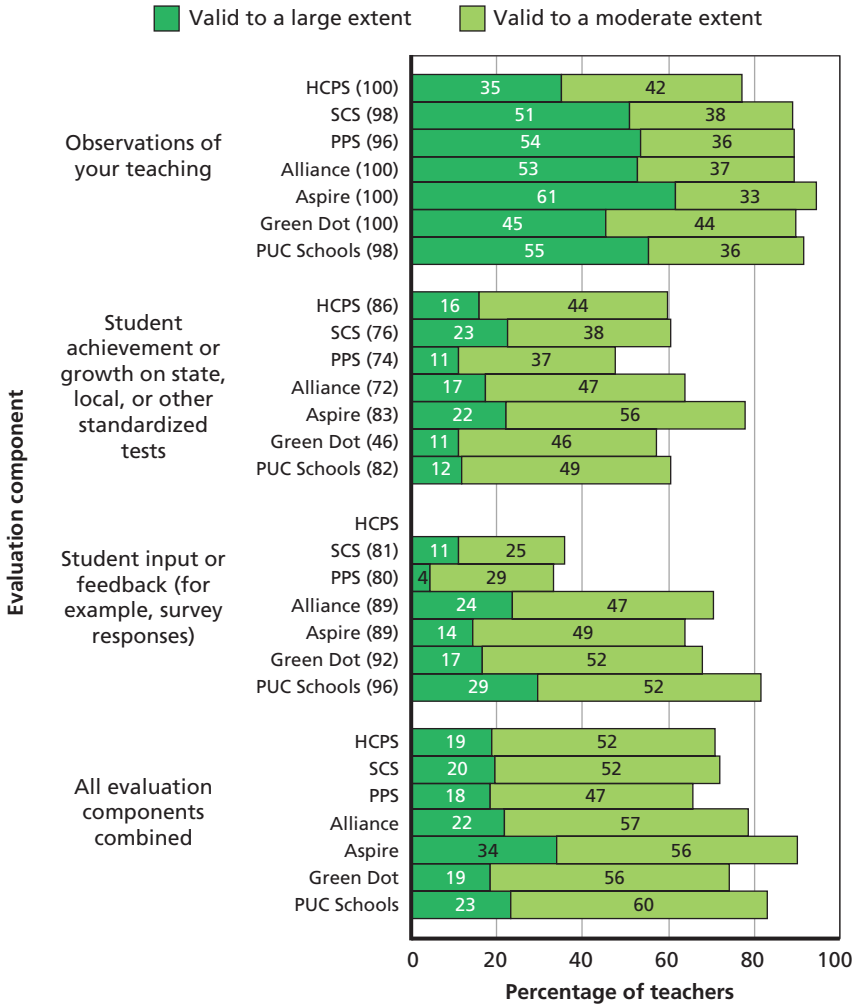
Our surveys and interviews suggested that teachers and school leaders thought that the effectiveness measures were valid indicators of the quality of teaching, although support was stronger for some components (such as observations of teaching) than for others (such as student achievement or student feedback surveys) (Figure S.3). Most teachers agreed that they had "a clear sense of what kinds of things the observers are looking for," that observers were "well qualified," and that they received "useful and actionable feedback." At the same time, some teachers we interviewed expressed reservations about the observation rubrics, saying they could lead to "a dog and pony show," that one observation might not be sufficient, and that observers might "purposely only find things wrong."

Overall, teachers indicated that their site's evaluation system was fair, although they were more likely to say that the system had been fair to themselves personally than to all teachers generally. Not surprisingly, teachers with high ratings were more likely than teachers with low ratings to report that their prior year's evaluation rating was accurate. At each site, at least 80 percent of teachers rated highly effective thought that their rating was accurate, but in none of the sites did more than 40 percent of teachers rated low think that their rating was accurate.

In general, teachers were concerned about attaching serious consequences—e.g., termination, salary bonuses—to the evaluation results. Fewer than 40 percent of teachers thought that such consequences were reasonable, fair, and appropriate. School leaders were more likely than teachers to agree that using the evaluation results for such purposes was fair.

Given teachers' concerns about the use of evaluation results, it is interesting to note that large majorities of teachers in each site received ratings equivalent to "effective" or "highly effective," and the percentages of teachers rated in these categories increased over time. Thus, the potential for negative consequences falls on few teachers.

Figure S.3
Percentage of Teachers Reporting That Evaluation Components Are Valid to a Large or Moderate Extent, School Year 2013–2014

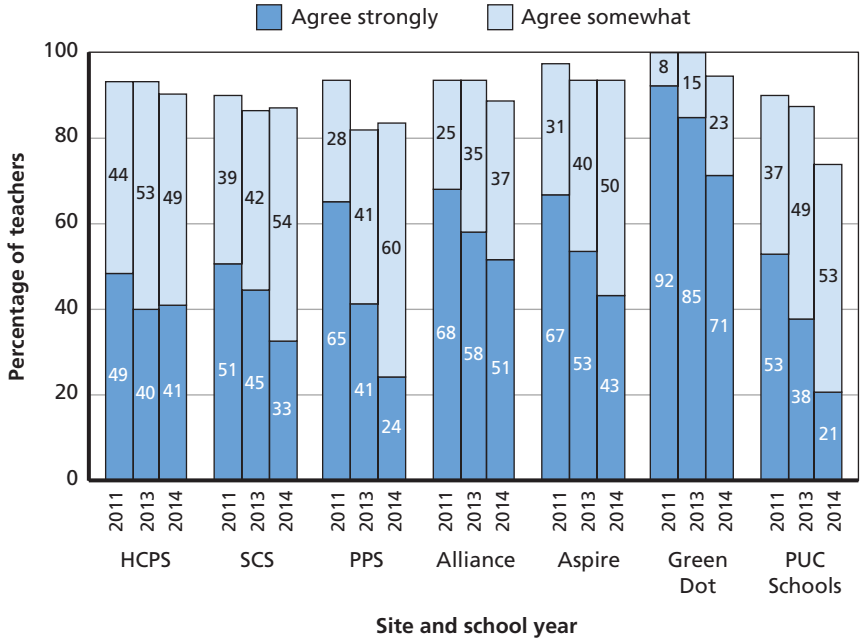


NOTE: The numbers in parentheses next to the site names are the percentages of teachers (among those who reported being evaluated) who indicated that the component was part of their evaluation. The numbers on the bars are those who rated the component valid to a moderate or large extent. All of the Intensive Partnership sites except HCPS include student input as a component in teachers' effectiveness ratings.

RAND *RR1295-S.3*

Overall, large majorities of school leaders agreed that students would benefit from the evaluation system in the long run (Figure S.4). Yet, the strength of school leaders' responses declined over time, although they still generally agreed with the statement. The percentage of teachers agreeing with this statement was 10 to 20 percentage points lower, and it declined a bit in the past year in most sites. This could be due to growing frustration with the mechanics of the evaluation system; in interviews, teachers reported that observations were too time-consuming, required too much preparation, and required them to teach to a checklist that is not best for students.

Figure S.4
Percentage of School Leaders Agreeing That "in the Long Run, Students Will Benefit from the Teacher-Evaluation System," Spring 2011, Spring 2013, and Spring 2014



NOTE: Each bar with a date reflects a different school year.

Costing Teacher Evaluation

To understand the magnitude of effort required to design and implement new teacher-evaluation systems, we conducted three case studies in the larger sites: HCPS, SCS (or, more precisely, Memphis City Schools before the merger occurred), and PPS. The three Intensive Partnership sites were at different points in the implementation of their new evaluation systems because of how their local districts were structured, their existing capacity, and the strategies they selected.

The total estimated evaluation-system expenditures in the three districts from November 2009 to June 2012 were $38.9 million. Per-pupil expenditures identified for implementing the system were $54 in school year 2010–2011 and $61 in school year 2011–2012 for HCPS, $21 in school year 2010–2011 and $51 in school year 2011–2012 for SCS, and $290 in school year 2010–2011 and $257 in school year 2011–2012 for PPS. These estimates do not capture the value of the time that school leaders spent observing and evaluating teachers or the value of the additional time teachers might have devoted to the evaluations.

To incorporate the cost of school-leader and teacher time into these estimates, we drew on time-allocation data from the teacher and school-leader surveys.[1] We estimated the value of the additional time school leaders spent on teacher evaluation to be $8 per pupil for HCPS, $46 per pupil for SCS, and $74 per pupil for PPS. We estimated the value of additional time teachers spent in mentoring and evaluation activities to be $119 per pupil for HCPS, $146 for SCS, and $213 for PPS. This large variation reflects the different approaches each site took to implement the teacher-evaluation system. For example, in PPS, school leaders bore the primary responsibility for conducting the teacher observations as part of their existing duties, so no new expenditures were incurred, whereas HCPS employed reassigned teachers

[1] Our estimates include time that teachers and school leaders spent attending training to conduct teacher evaluations, observing classroom instructions, preparing and providing feedback to teachers as part of their evaluations, other activities related to evaluating teachers, and time spent evaluating nonteaching staff. Estimates do not include time spent in professional development to improve practice that the evaluations prompted.

as full-time peer evaluators to conduct most of the observations and incurred additional costs.

Data from the surveys also indicate that new teacher-evaluation systems caused staff, particularly school leaders, to shift their responsibilities and restructure the ways they spend their time. New activities, such as classroom observations, communications among school leaders and teachers about performance, and PD, appeared to replace time previously spent on certain routine administrative tasks. Interviews with central-office staff across the Intensive Partnership sites seem to suggest the reallocation of certain central-office administrative responsibilities as well.

Staffing

To try to improve the overall effectiveness of their teacher workforces, the sites made several changes to their procedures for recruiting, hiring, placing, granting tenure to, and dismissing teachers. These changes included earlier identification of vacancies, more-aggressive recruitment, better screening of candidates for effectiveness, more-strategic referrals to high-need schools, interview training for principals, and more-effective recruitment, hiring, and orientation for new hires.

The specific changes varied from site to site, reflecting local laws and contractual agreements. By 2012, most sites had most practices in place. (The CMOs and HCPS had many of these practices in place at the start of the initiative.) School leaders were generally satisfied with teacher-hiring practices but were less pleased with transfer and dismissal practices. School leaders reported that low-performing teachers were unlikely to be dismissed directly and that most teachers were not very worried about being dismissed. However, teacher mobility remained an issue in the sites: The CMOs in particular struggled to retain effective teachers, and some principals in the districts expressed dissatisfaction with teachers who were assigned to them without their agreement.

Professional Development

In the initial years of the reform, the sites made few changes to their PD practices, in part because they lacked comprehensive data on teaching quality on which to base changes. By 2012 and 2013, when the teaching-effectiveness measures were operational, the sites had begun to explore strategies for supporting teachers based on their identified needs.

However, for several reasons, the sites have found it challenging to customize PD to the needs of individual teachers. First, much of the formal PD traditionally provided is delivered in a group format that can be difficult to individualize. Second, the new observation rubrics identify elements of effective teaching that do not always align with existing PD offerings. Third, it has been difficult to implement information systems allowing sites to link specific PD opportunities to identified dimensions of practice (or keep track of which teachers participate in which PD). Fourth, some sites expanded their PD offerings to include videos, online training, and individual readings, but not everyone knew of these options.

Nevertheless, school leaders reported that the evaluation information was helpful in focusing mentoring and support, and majorities of teachers in some sites reported having access to coaching or other PD that addressed their needs. In some sites, teachers were responsible for seeking the PD they needed, while others made the principal responsible for overseeing teachers' PD. Many sites emphasized individualized coaching or mentoring, usually conducted by other teachers who had been recognized as effective, rather than formal workshops or courses. As one teacher said, "working with colleagues, we both teach biology; when we have time to work with each other or share resources, it's more useful than a lot of the professional-development time we spend waiting for someone to finish talking."

Those teachers who received customized support linked to their evaluation results found the support useful for improving their instructional practice. Specifically, most teachers in all sites except PPS who received support to address needs identified by their evaluation reported that this support helped them address these needs.

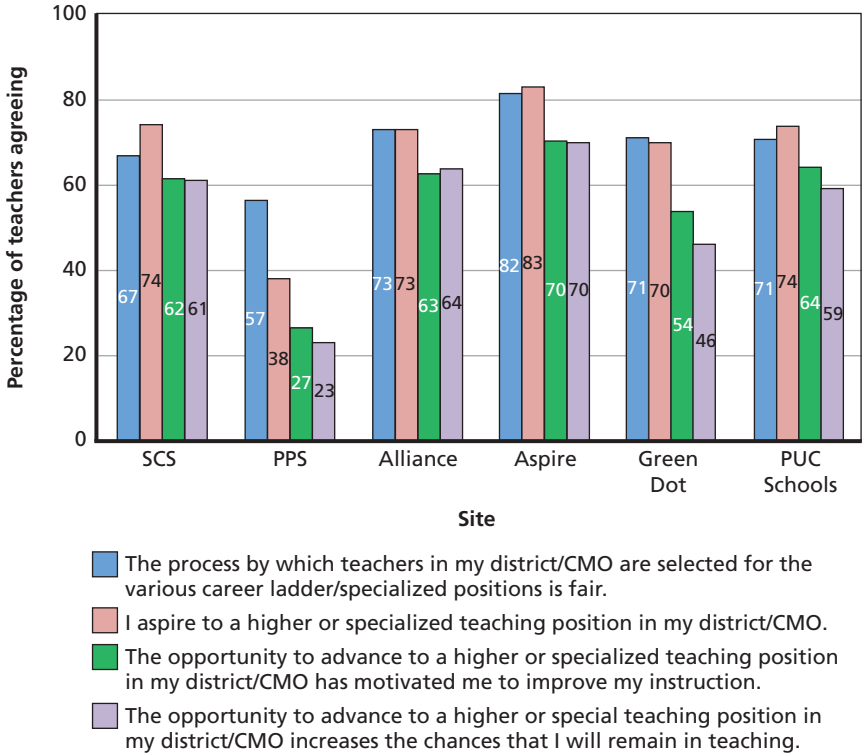
Compensation and Career Ladders

Compensation reforms and specialized career-ladder positions were implemented later than the other levers at most sites, in part because the teaching-effectiveness measures needed to be in place to facilitate these reforms and in part because, in many sites, these reforms required changes to negotiated contracts with teachers. By school year 2013–2014, all of the sites had adopted some form of effectiveness bonus, i.e., awarding extra compensation to teachers who receive the highest effectiveness ratings.

Teachers' responses were mixed regarding the fairness and incentive effects of their site's compensation system, with teachers in the CMOs slightly more positive than teachers in the districts. Although most teachers thought that base pay should be based on seniority, a majority also thought that teachers should receive additional compensation for demonstrating outstanding teaching skills and for teaching in low-performing schools. Some teachers said that they preferred extra compensation tied to increased responsibilities; some liked having extra compensation linked to performance on the teaching-effectiveness measures; and some thought that both were fine.

All of the sites have developed some form of career ladder in which effective teachers take on new roles, such as coaching or mentoring, and receive a permanent increase in their base salary or a one-time additional stipend for these responsibilities. Some of these positions are full time, but most expect teachers to continue spending some of their time in classroom instruction. Most teachers at most sites thought that the selection process for specialized career-ladder positions was fair, reported aspiring to a specialized position, indicated that career-ladder positions motivated them to improve instruction, and reported that the career-ladder positions increased the chances that they would remain in teaching (Figure S.5). The lower proportions of PPS teachers agreeing with these statements might be attributable to relatively few openings for career-ladder positions in PPS, as well as to perceptions about such positions (e.g., amount of work, need to move, uncertainty about future).

Figure S.5
Percentage of Teachers Agreeing with Statements About Career-Ladder Positions in 2014



NOTE: We exclude HCPS from this because there were not many such positions in that system and too few respondents said that they knew anyone in a career-ladder position.

RAND *RR1295-S.5*

Conclusions and Future Questions

Although it is too soon to draw overall conclusions about implementation—the site implementation grants extend through school year 2015–2016, and the sites are still modifying policies and procedures as they gain experience, as well as in response to changing contextual factors—the information we have gathered to date could

reveal themes and questions that other sites adopting similar reforms are likely to face.

Time to Implement Reforms

In developing the Intensive Partnership initiative, the foundation recognized that change takes time, particularly to modify core policies and practices. As a result, it funded the sites for six years. It appears that the sites benefited from time devoted to initial planning and early stakeholder engagement. The Intensive Partnership experience also suggests that it was wise not to change all related human-resource (HR) policies at the outset but to focus on teacher-evaluation metrics first and then address reforms, such as customized PD and career ladders, which build from the teacher-evaluation metrics. A newly adopting site could probably learn useful lessons from the experiences of the Intensive Partnership sites in many areas, including training observers, defining terms in the observation rubric, combining multiple measures into a single effectiveness score, creating specialized career-ladder positions, and developing PD to support identified needs.

Interpreting Improvements to Date

The growing percentages of teachers performing at the highest levels on teacher-evaluation measures might provide some evidence that the reforms are improving teacher quality. Nevertheless, other explanations for these improvements are also possible. For example, observers might be getting more generous in their evaluations as stakes attached to teachers' performance are intensified.

Keeping Reforms on Track

Each site had to cope with unanticipated changes in state policy, most had to address stakeholder complaints or criticisms, and nearly all had to deal with unanticipated changes in leadership. Despite these challenges, for several reasons, the sites were more or less able to stay on course. Long-term strategic plans, including the time taken to implement the reforms, helped them do so. Foundation program officers also provided support, helping the sites develop strategies to manage change. Foundation program officers also responded as necessary to

concerns about measurements and their weighting. Community, philanthropic, and other groups also helped sustain the reform in some sites through their endorsements.

Role of the Bill & Melinda Gates Foundation

The foundation played a stronger role in guiding implementation than is typical in many grant programs. In addition to providing funding, the foundation helped sustain the vision, convened the sites and experts for relevant learning and dialogue, and served as a critical friend, which included providing an engaged program officer able to reflect on the initiative from the perspective of an outsider.

Achieving and Maintaining Acceptance from Teachers and School Leaders

All the sites have sought teacher support for the reforms, promoting the reforms as means to help teachers improve their practice. Most teachers perceive the primary purpose of the reform to be providing feedback to help improve instruction, and many teachers reported that the reforms have made them more reflective about their teaching. Yet, some sites are encountering resistance as stakes intensify, and some teachers are beginning to see the reforms as stressful and punitive, rather than helpful and informative—suggesting that the stakes might, at some point, undermine teacher support.

Supporting Teachers in Improving Their Practice

The sites have implemented several strategies to help teachers improve their effectiveness, including centralized PD targeting common challenges, customized workshops, local coaching, and collaborative communities of practice. The sites have struggled with developing data systems to monitor teacher participation in this diverse collection of activities. In addition, the sites have not yet determined which of these strategies is most effective, but teachers and school leaders perceive some to be more valuable than others.

Future Analyses

We will continue evaluating the implementation and impact of the Intensive Partnership initiative through the 2015–2016 school year. This will include updates on implementation, the cost of the initiative, and specific steps the sites are taking to sustain the reforms. We will also investigate the quality of the teaching-effectiveness measures and assess the effects that the reform has on student outcomes.

Acknowledgments

We are grateful to the large number of central-office staff, school leaders, and teachers who gave generously of their time to share their insights and experiences with the Intensive Partnership initiative. In particular, we appreciate the efforts of site staff who reviewed our tables and charts and helped us characterize each site's implementation of the various Intensive Partnership levers correctly. These individuals include Anna Brown and Tracy Schatzberg of Hillsborough County Public Schools; Jessica Lotz and Kristin M. Walker of Memphis City Schools; Tara Tucci and Ashley Varrato of Pittsburgh Public Schools; Judy Ivie Burton, Harris Luu, and Vireak Chheng of Alliance College-Ready Public Schools; Elise Darwish and James Gallagher of Aspire Public Schools; Cristina de Jesus and Julia Fisher of Green Dot Public Schools; and Jacqueline Elliot and Jonathan Stewart of Partnerships to Uplift Communities Schools. We also appreciate the foundation staff's willingness to engage with us; we are especially grateful to David Silver and Eli Pristoop for their advice and responsiveness throughout the project. We thank Cathleen Stasz of RAND, Julie Marsh of the University of Southern California, and Leslie M. Anderson of Policy Studies Associates for reviewing the document and providing constructive feedback. Finally, we acknowledge other members of the RAND team, including Catherine H. Augustine, Courtney Ann Kase, Beth Katz, Stephanie Lonsinger, Nate Orr, Mollie Rudnick, and Jennifer Sloan, and of the American Institutes for Research team, including Jennifer Ford, Kaitlin Fronberg, Gur Hoshen, Jesse D. Levin, John Mezzanotte, and Antonia Wang.

Abbreviations

AIR	American Institutes for Research
CBA	curriculum-based assessment
CMO	charter management organization
EET	Empowering Effective Teachers
ELA	English language arts
FY	fiscal year
HCPS	Hillsborough County Public Schools
HR	human resource
LAUSD	Los Angeles Unified School District
LIM	low-income minority
MCS	Memphis City Schools
MET	Measures of Effective Teaching
PD	professional development
PPS	Pittsburgh Public Schools
PUC Schools	Partnerships to Uplift Communities Schools
RISE	Research-based Inclusive System of Evaluation
RTT	Race to the Top

SCS	Shelby County Schools
SGP	student growth percentile
TCRP	the College-Ready Promise
TE	teaching effectiveness
TEM	Teacher Effectiveness Measure
TIF	Teacher Incentive Fund
TVAAS	Tennessee Value-Added Assessment System
VAM	value-added model

CHAPTER ONE

Introduction

In the 2009–2010 school year, the Bill & Melinda Gates Foundation and seven school systems began a six-year effort to improve student outcomes by reforming how the sites interact with their teacher workforces. This report summarizes the status of those reforms five years after the grants were announced, based on a comprehensive evaluation that researchers from RAND and the American Institutes for Research (AIR) are conducting. This introduction begins by describing the theory of action that motivates the Intensive Partnership initiative. Following that, we recount the competitive process to select the seven Intensive Partnership sites and present brief introductions to each site. Next, we describe our approach to evaluating the Intensive Partnership initiative, including the limitations of this study. The Intensive Partnership initiative represents a huge investment by the foundation, which has pledged $290 million to the effort, and we briefly summarize overall expenditures from the foundation and other sources before outlining the rest of the report.

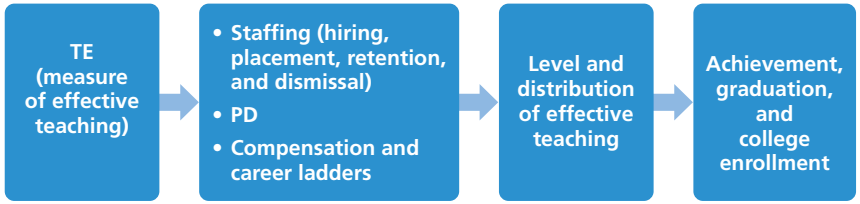
The Intensive Partnership for Effective Teaching: Improving Outcomes for Low-Income and Minority Students

The Intensive Partnership initiative is based on the premise that efforts to improve instruction can benefit from high-quality measures of teaching effectiveness (TE). In the Measures of Effective Teaching (MET) study (Kane and Staiger, 2012), the foundation found that it is

possible to build a high-quality measure of TE by combining information about student achievement growth, direct observation of teaching practice, and student feedback. The authors concluded that "better evidence should lead to better decisions" about teachers. The Intensive Partnership initiative is an attempt to put that conclusion to the test, i.e., to see whether a district or charter management organization (CMO) can implement a high-quality measure of TE and use it to manage its teacher workforce in ways that dramatically improve student outcomes. The foundation and the sites are particularly interested in improving outcomes for low-income minority (LIM) students. This approach is consistent with broader national trends in which states and localities are increasingly mandating performance-based teacher evaluation (Doherty and Jacobs, 2013; Master, 2014; Rotherham and Mitchel, 2014).

Figure 1.1 shows the theory of action for the Intensive Partnership initiative (adapted from a figure prepared by the foundation). The developers of the initiative believe that, by strategically managing its human capital, a district or CMO can provide all students with effective teaching, which will better prepare them for college and careers. The process starts with the adoption of a valid measure of TE; in the Intensive Partnership sites, these measures are all weighted combinations of classroom observations, growth in student achievement, and other information, such as student surveys of learning conditions or indicators of teacher professionalism. The TE measures are incorpo-

Figure 1.1
The Intensive Partnership Theory of Action



SOURCE: Bill & Melinda Gates Foundation.
NOTE: Career-ladder positions are those for teachers with extra responsibilities and extra pay.
RAND *RR1295-1.1*

rated into a comprehensive management system designed to improve the teacher workforce over time. The elements of this system, which are all linked to the TE measures, include staffing policies (such as hiring practices, placement, and retention and dismissal decisions), professional-development (PD) practices, and policies related to compensation and the creation of career ladders for teachers.

In particular, these new policies and practices are expected to improve the overall effectiveness of teaching in various ways. First, new staffing practices should lead to more-effective teachers for all students, particularly low-income and minority students. These staffing policies include better hiring and mentoring informed by a clear understanding of the elements of effective teaching; this should improve the effectiveness of new teachers. Similarly, more-strategic placement policies should result in effective teachers being assigned to students in greatest need. Finally, through informed tenure and dismissal policies, districts, CMOs, and schools should encourage ineffective teachers who cannot improve their practice to exit the system. Second, customized PD should promote growth in the effectiveness of existing teachers. PD that is tailored to teachers' strengths and weaknesses as identified by the evaluation measures should be particularly helpful for "teachers in the middle," those who are not yet highly effective but who are not performing at the lowest level. These teachers should be able to become more effective by addressing the areas of weakness in their instruction. Third, through compensation and career-ladder policies (i.e., policies that create new teaching positions with added responsibilities for mentoring, support, and leadership), highly effective teachers should be retained at higher rates, and the opportunity for new career opportunities or compensation should motivate teachers to improve their performance. Taken together, this set of strategies is expected to improve overall TE and promote a more-equitable distribution of effective teaching across schools and students. This change, in turn, will lead to better outcomes for all students, including higher graduation and college attendance rates.

This theory of action is consistent with much of the recent literature on teacher evaluation and compensation. Most fundamentally, studies of pay-for-performance policies suggest that relying exclusively

on increased compensation as a means to promote TE is unlikely to improve student learning or teacher behaviors (Fryer, 2011; Marsh et al., 2011; Springer, Ballou, et al., 2010; Springer, Pane, et al., 2012; Yuan et al., 2013), but there is some evidence that comprehensive, multiple-measure, teacher-evaluation systems can improve outcomes (Dee and Wyckoff, 2013). Research also suggests that the *information* that the TE measures provide, rather than the incentives attached to those measures, can play an important role in raising student achievement. Principals are likely to perceive direct measures of practice, such as observations, as most useful for improving teacher performance (Goldring et al., 2015), and their use has been associated with increased student achievement. For example, Taylor and Tyler, 2012, found that midcareer Cincinnati teachers who received evaluative feedback on their practice were able to improve student outcomes as measured by value-added scores and that the information they received about their practice was a likely mechanism contributing to this improvement. The implementation of a pilot evaluation system in Chicago Public Schools, based on classroom observations that were assessed using the Danielson Framework for Teaching, was also associated with improved student achievement (Steinberg and Sartain, 2015). Research on the effects of using teacher-evaluation results to inform staffing and career-ladder decisions is limited, but these uses are consistent with some prior empirical evidence and theory about human capital reforms (Goldhaber, 2015).

Each site has developed its own version of the approach depicted in Figure 1.1, adopting different metrics, emphasizing somewhat different components, and modifying the template in different ways over time in response to local considerations and changing priorities. Nevertheless, this basic theory of action is visible in all seven sites. Importantly, all of the sites are concerned not only with managing teachers—using evidence of effectiveness to guide decisions about teacher compensation, placement, dismissal, and other administrative actions—but also with developing teachers—using measures to identify areas for improvement and providing mentoring and support to improve performance.

In the chapters that follow, we describe the extent to which each of the seven sites implemented the features identified in Figure 1.1 and

examine the ways in which teachers responded to the reforms. There is some evidence that teachers' understanding of, and support for, classroom-level changes are important predictors of quality and intensity of implementation and of the extent to which the changes lead to improved teaching quality (Coburn, 2005; Schmidt and Datnow, 2005; Spillane, Reiser, and Reimer, 2002). Thus, we expect that teachers' views on the validity of teacher-evaluation measures and their beliefs that the evaluations will be used appropriately are likely to be important predictors of implementation success. Recent research suggests that teachers tend to express fairly positive opinions about some aspects of evaluation systems, particularly the classroom-observation component, and they are typically skeptical of the validity of achievement growth measures (Jiang, Sporte, and Luppescu, 2015). Moreover, school leaders report finding classroom observations more useful for decisionmaking than student achievement growth measures (Goldring et al., 2015). We draw on our data to examine whether teacher and leader perceptions in the seven Intensive Partnership sites are consistent with these results.

Launching the Intensive Partnership Initiative

In 2009, the foundation decided to commit resources to test its theory of action in practice.[1] Foundation staff anticipated that the human-resource (HR) changes associated with their theory of action might not be easy to implement, and they developed a plan for selecting and supporting sites to increase the chance of success. For example, the foundation assumed that the changes would likely require agreement from local stakeholders, including each site's central administration, teachers, school board, and local community. Furthermore, most sites would need additional resources to undertake such a transformation. They would need external expertise to help with design, communication, and data infrastructure, and they would need financial support to

[1] The description of the launch of the initiative is based on interviews with foundation staff and a review of foundation documents.

enable existing staff to participate in planning activities, to retrain staff to implement new practices, and to hire additional staff to perform new functions. With the decision to undertake the Intensive Partnership initiative, the foundation committed to providing participating sites with money, expertise, consultants, and ongoing monitoring to test the efficacy of its approach. The foundation envisioned the relationship with each site as a partnership, hence the term *Intensive Partnership*.

Thus, the foundation began a search for districts and CMOs (sites) willing to transform their HR practices to align with the Intensive Partnership theory of action.[2] The foundation initially cast a wide net in its search for partnership sites. The foundation began with the full national population of districts and CMOs, selectively filtering down the number of candidates based on enrollment (targeting districts or CMOs with between about 25,000 and 250,000 students), student background (wanting more than 40 percent of students eligible for free or reduced-price lunch), and state policy environment (excluding states with low thresholds for granting tenure).

This process produced a set of about 70 potential districts. Staff reviewed this list based on their own knowledge and identified about two dozen districts that had demonstrated a commitment to improvement efforts. They contacted each of these sites to judge their interest in the goals of the initiative and to learn about previous reform efforts that might demonstrate district leaders' ability to work effectively with key stakeholders and maintain a focus on improvement. They visited each site and met with district administrators, school board members, and union representatives to assess whether key stakeholders would commit to the reform. They also sought a commitment from community leaders that they would "stay the course" to ensure that the reforms were sustainable after the foundation's commitment expired. After the visits, the foundation selected ten districts to receive a formal request for proposals to be part of the partnership.

The foundation was also interested in the role that CMOs could play in educational improvement, and it had worked with some CMOs in the past. In response to interest from some Southern California

[2] In this report, we use the term *site* to refer to a district or a CMO.

charter schools, the foundation decided to test their ability to work collectively. The foundation encouraged forming a collaborative CMO effort to prepare a proposal.

The sites received a variety of supports from the foundation to develop their proposals. For example, the foundation paid for technical assistance—each site was allowed to choose to work with one of three consulting firms to conduct analyses of local data and help in the development of its plan. In addition, the sites were convened three times between April and July 2009 to share ideas and report on progress; the goal of the meetings was to establish a "collegial competition" in which the sites pushed each other to improve their plans.

The foundation considered a variety of factors when it reviewed the proposals and made its final selections. The most important included the scope and comprehensiveness of the plan, the level of collaboration among local stakeholders, and the site's capacity to execute the plan and accomplish its stated goals. A nonbinding memorandum of understanding signed by district, union, and community leaders was a required part of each site's proposal. The foundation also wanted to fund a set of sites that differed in size and demographics, so the overall portfolio of partners would contain archetypes that other districts would see as similar to themselves. If successful, the foundation hoped the sites would serve as "proof points" for many others.

Although we are jumping ahead in the sequence of events, this is an appropriate place to mention another distinctive feature of the Intensive Partnership initiative. After the award of the contracts, the foundation continued to play an active role in supporting the sites' efforts. For example, the foundation assigned a dedicated program officer to each site, whose role includes active engagement in local strategic planning and helping to identify resources (e.g., consultants on instructional technology, strategic communication) to meet each site's individual needs. Initially, the program officers spent roughly one week per month on site working directly with the site's implementation lead. The foundation holds annual or semiannual "convenings" to provide information and support services and give the sites opportunities to share successes and challenges. The foundation also hired a contractor to build a data warehouse to monitor progress toward meeting the

goals of the initiative. The contractor assembles site-level data on students and teachers and prepares annual "dashboards" with quantitative indicators of progress. These results are discussed in detail with each site at an annual "stocktake" event designed to foster reflection and improvement. However, these details were not explicit at the time of the selection of sites.

Informed by the proposals received, the foundation chose three districts and a collaborative of four CMOs (the College-Ready Promise, or TCRP).[3] As the foundation noted in its press release,

> Each of the selected communities demonstrated a broad-based commitment to raising student achievement, with an emphasis on reforming how teachers are recruited, evaluated, supported, retained, and rewarded. They also represent a mix of large and mid-size urban school systems with diverse populations. (Bill & Melinda Gates Foundation, undated)

The three traditional school districts that received awards were Hillsborough County Public Schools (HCPS) in Florida; Memphis City Schools (MCS), which, in 2013, merged with Shelby County Schools (SCS) in Tennessee,[4] and Pittsburgh Public Schools (PPS) in Pennsylvania. The CMOs under the TCRP umbrella included Alliance College-Ready Public Schools, Aspire Public Schools, Green Dot Public Schools, and Partnerships to Uplift Communities (PUC) Schools. Working through TCRP, the CMOs pooled their efforts and created common observation rubrics and a common value-added methodology; they operate independently when it comes to creating an overall measure of effectiveness and using it to manage and develop their teacher workforces.

---

[3] A fifth CMO, Inner City Education Foundation Public Schools, was originally part of TCRP but left the collaborative after the first year.

[4] With the exception of the historical summary paragraph below, we refer to the district as *SCS* in the report.

Introduction to the Intensive Partnership Sites

Each of the sites presented opportunities and challenges for reforming teacher policies and practices, and each site's approach was influenced in part by the state and local context. Table 1.1 presents basic information about the size and student demographic characteristics of the sites at the start of the initiative.

The Three District Sites

The three district sites—HCPS, SCS, and PPS—differ in several ways that are particularly relevant in interpreting the results that follow. In particular, when Intensive Partnership funding was awarded, HCPS was already engaged in several of the reforms emphasized in the Intensive Partnership approach, including salary bonuses based on individual- and school-level effectiveness measures along with customized PD. This district therefore had a head start over the other Intensive Partnership sites on at least some of the activities funded as part of the initiative. HCPS also participated in the Bill & Melinda Gates Foundation–funded MET project from 2009 to 2013, which, as described above, developed measures of effective teaching. PPS and SCS, by contrast, had been utilizing standard step-based salary schedules and had not attempted to customize PD to teachers' needs. During the year prior to the Intensive Partnership funding, PPS began collaborating with union leaders to develop a system for observing teacher practice for guiding teacher professional growth; the Intensive Partnership funding contributed to further development of this system. SCS also participated in the MET project and won a federal grant to support performance-based pay; the district also had been receiving value-added scores on its teachers for many years through the Tennessee Value-Added Assessment System (TVAAS). Additional background information on the districts is provided in Box 1.1.

Table 1.1
Characteristics of the Intensive Partnership Sites, 2009–2010 and 2013–2014 School Years

	2009–2010				2013–2014			
Site	Number of Schools	Total Enrollment	Percentage of Students Who Are Low Income	Percentage of Students Who Are Minority	Number of Schools	Total Enrollment	Percentage of Students Who Are Low Income	Percentage of Students Who Are Minority
District								
HCPS	251	187,411	56	57	260	188,917	64	59
SCS	205	106,934	87	92	246	120,590	81	88
PPS	74	25,986	72	63	66	24,358	76	62
CMO								
Alliance	16	5,058	94	99	23	9,988	91	99
Aspire	25	7,632	72	80	37	13,653	79	85
Green Dot	13	7,118	92	99	16	10,183	94	99
PUC	10	2,598	87	98	13	4,211	73	97

NOTE: *Low income* entails being eligible for free or reduced-price meals. *Minority* consists of black, Hispanic, American Indian, or multiracial (PPS only). Only students enrolled in the schools on October 7, 2009, and October 2, 2013, were included. SCS information reflects MCS for school year 2009–2010 and SCS for school year 2013–2014.

Box 1.1

Brief Summaries of the Three Intensive Partnership Districts at the Start of the Intensive Partnership Initiative

Hillsborough County Public Schools

HCPS is the eighth-largest school district in the country and second-largest in Florida. In 2009, when HCPS submitted its Empowering Effective Teachers (EET) proposal to the Bill & Melinda Gates Foundation, it was already engaged in a variety of efforts to improve TE, including delivering PD and paying for high performance. The prior evaluation system did not incorporate student outcome measures, relied solely on principals as the evaluators of teacher performance, and did not link PD to evaluation results. In September 2009, HCPS contracted with Charlotte Danielson to develop a new classroom evaluation rubric based on her Framework for Teaching. Through the EET initiative, HCPS developed a value-added model (VAM) based on student test scores for all grades and subject areas. This was possible because, prior to this initiative, HCPS had developed its own tests for all subjects and grades. In February 2014, the Florida State Board of Education adopted a modified version of the Common Core State Standards, now referred to as the Florida Standards. These new standards necessitate retraining of educators, retooling of curriculum and pacing guides, and revisions to Common Core–based instructional modules and units.

Pittsburgh Public Schools

PPS is the second-largest school district in Pennsylvania. In 2008, prior to the Intensive Partnership initiative, PPS administrators, union leaders, and teachers began work on the Research-based Inclusive System of Evaluation (RISE), a new system for observing and evaluating teacher practice and for guiding teacher professional growth. In 2010, PPS and the Pittsburgh Federation of Teachers jointly developed and passed a collective bargaining agreement that codified new career-ladder roles associated with additional compensation, bonuses largely based on effectiveness measures, and a merit-based salary schedule for teachers hired after the agreement was passed. PPS's initiative, which the district has labeled the Empowering Effective Teachers initiative, builds on the RISE framework. Initially, it also emphasized improving the quality of its teaching workforce through hiring and attracting some of the district's best teachers to the neediest schools. However, budget shortfalls have forced the district to cut services, furlough teachers, and reduce the number of central-office staff. As a result, PPS has shifted its emphasis to coaching and PD targeted to the areas in which teachers need to improve. In 2012, Pennsylvania enacted Act 82, which required new multiple-measure rating systems to evaluate teachers and principals, specified policies for teacher dismissal, and mandated participation in PD for all teachers who receive low performance ratings. PPS was granted a three-year approval to use its own rating system to meet the Act 82 requirements; this approval will expire at the end of the 2016–2017 school year.

Memphis City Schools (now Shelby County Schools)

MCS was the largest school district in Tennessee prior to its merger in 2013 with the surrounding SCS. SCS remains the largest district in the state despite several municipalities in the county leaving SCS to form their own school districts in 2014. Prior to the Bill & Melinda Gates Foundation proposal (2008–2009 school year), MCS relied primarily on a traditional, step-based salary schedule and did not link PD or other HR decisions to teacher performance, although the district had been experimenting with some new policies that were consistent with the goals of the Intensive Partnership initiative. For example, MCS participated in the MET project and piloted classroom-observation rubrics as a measure of TE. MCS also received federal funding to provide group-based bonuses to teachers and administrators in district schools with high achievement gains. TVAAS, the state's value-added

Box 1.1—Continued
assessment system for teachers, has been in place for more than 20 years. After the merger with SCS in 2013, the MCS initiative was adopted by the new district and renamed the Teacher and Leader Effectiveness initiative. SCS incorporated its existing tiered coaching teacher support system, which includes career-ladder roles, into Teacher and Leader Effectiveness. State-level policies have also shifted during the course of the Intensive Partnership reforms. As part of Tennessee's Race to the Top (RTT) effort, the state revised its teacher-evaluation policies and measures, as well as its hiring, placement, and tenure policies; these changes were consistent with the goals of the Intensive Partnership initiative. Although SCS's current measure of TE differs slightly from the state's measure, SCS adheres to state policy for teacher hiring, placement, and tenure.

The College-Ready Promise

The four CMOs that collaborated to form TCRP are all nonprofit organizations whose schools serve low-income students in high-need communities. As CMOs, they were not bound by the staffing, teacher-evaluation, PD, or compensation policies that applied to traditional school districts in California. For example, prior to the Intensive Partnership initiative, principals in the CMOs had complete hiring authority for their schools, none of the CMOs awarded tenure, and PD was the province of each school principal. One important feature of these four CMOs is that all are expanding their operations, opening new schools that require additional staff, so recruitment and hiring have been a central concern. Despite their commonalities, each of the CMOs had its own culture and began the Intensive Partnership reforms with its own perspective. Additional background information on each of the four CMOs is provided in Box 1.2.

Before the Intensive Partnership initiative, there was considerable variation in the CMOs' teacher-evaluation policies, particularly the extent to which each central office set rules for teacher evaluation. When they formed TCRP and were awarded an Intensive Partnership grant, the CMOs each agreed to adopt TE measures that included a common teacher-observation measure and a common student growth measure. In addition, they agreed to incorporate both a student satisfaction survey and a family satisfaction survey in their TE measures. In presenting the initiative to their staffs, the CMOs emphasized the opportunity it offered for effective teachers to earn higher salaries

Box 1.2
Brief Summaries of the Charter Management Organizations at the Start of the Intensive Partnership Initiative

Alliance College-Ready Public Schools

Alliance, with schools located in Southern California, holds college readiness as a core goal, and it saw the Intensive Partnership initiative as a way to obtain additional funding to promote this goal and bolster some of its efforts. The Alliance organizational structure was the most decentralized of the CMOs. Principals were responsible for recruiting, screening, and hiring staff; conducting teacher evaluations; coaching teachers; and providing PD. Thus, they operated with a great deal of autonomy. The establishment of a specific common teacher-observation rubric with a set number of observations and observation process was a cultural shift for Alliance.

Aspire Public Schools

Aspire is the largest of the four CMOs, with schools in southern, central, and northern California and, since school year 2013–2014, in Memphis, Tennessee. Prior to the Intensive Partnership initiative, Aspire had its own teacher-observation rubric that was similar to the rubric that TCRP adopted. This similarity facilitated the adoption of the new TCRP evaluation process. The grant was a catalyst for training principals on capturing objective evidence and calibrating their observations. In 2009, Aspire began a major initiative to develop a teacher data portal and a teacher resource library, with the long-term goal of creating a fully integrated student, teacher, and HR data platform. Prior to the initiative, Aspire also had a team of coaches organized by region and school level particularly focused on new teachers. Aspire has always had a merit pay system integrated with a step-and-column compensation structure.

Green Dot Public Schools

Green Dot schools are primarily located in Southern California, but the CMO has one school in Tennessee that opened in 2014 and is planning to open additional schools in Seattle and Tacoma. Green Dot is unique among the four CMOs in having a teacher union. Each of Green Dot's Intensive Partnership reforms required union approval. To encourage teacher ratification of the initiative, Green Dot adopted a strategy of intensive communication with teachers. Like the other CMOs, Green Dot described the TCRP initiative to teachers as a structure that would lead to increased pay for effective teachers. As a result, the teachers' union was cautiously welcoming of the initiative. Prior to the Intensive Partnership initiative, principals and a teacher team at each school led PD. At monthly meetings with the director of teacher support, principals received guidance on PD planning based on data from benchmark assessments.

Partnerships to Uplift Communities Schools

PUC Schools are located primarily in Southern California, but one school recently opened in Rochester, New York. PUC Schools is a close-knit organization highly responsive to school-leader and teacher feedback; for example, all principals meet weekly with home-office staff to discuss operations and initiatives. Though principals have full hiring authority, recruitment and screening is done centrally, and applicant interviews are also centrally organized. PUC Schools has a strong culture around performance management. Prior to the Intensive Partnership initiative, PUC Schools had used a common TE rubric in all its schools for three years, and school leaders conducted extensive teacher observations and informal coaching sessions. PD was provided on a schoolwide basis. PUC Schools has always had a coaching staff that provided support to new and struggling teachers. Prior to the Intensive Partnership initiative, PUC Schools had already started to consider tying compensation to teacher performance, so this policy did not represent a dramatic shift for the organization.

rather than the possibilities it would offer to improve teaching practice. Unfortunately, from its inception, TCRP has had to weather a huge downturn in state funding, which began in school year 2009–2010 and continued through school year 2013–2014. During this period, the CMOs stopped giving raises to teachers, cut back on development, and postponed some of their planned initiatives, such as a pay-for-performance compensation system.

Approach to Evaluating Intensive Partnership Reform Implementation

The evaluation is examining both implementation and outcomes associated with the Intensive Partnership reforms, as well as the degree to which the reform policies are replicated in other districts. The chapters that follow focus exclusively on implementation and address two broad research questions:

- What policies and practices were implemented in each site as part of the Intensive Partnership initiative, and when?
- How did teachers and school leaders respond to the Intensive Partnership reforms?

The following sections describe the data we collected to answer the research question and our approach to measuring implementation. Additional detail can be found in Appendixes A through E.[5]

Data Collection

To address the first question, we relied primarily on annual interviews with central-office staff in each site. We collected additional informa-

[5] Appendix A summarizes the methods used for collecting and analyzing the interview data. Appendix B describes methods for coding Intensive Partnership implementation status, and Appendix C summarizes methods for the survey data collection and analysis. Appendix D provides detailed discussions of lever implementation for each site along with the detailed coded lever tables. Appendix E summarizes the responses of teachers and school leaders to survey questions about the allocation of their work time. Appendixes D and E are available online only (Stecher, Garet, Hamilton, et al., forthcoming).

tion about plans for implementation and about implementation itself by reviewing the annual "stocktake" documents that the sites prepared for the foundation and other documents that the sites and foundation prepared.

To address the second question, we administered web-based surveys to school leaders and teachers each spring beginning in 2011.[6] We surveyed all school leaders and a sample of teachers from every school within each site. School leaders included principals, assistant principals, and other staff holding equivalent titles (e.g., director, instructional leader, dean). We used a stratified random sampling procedure to select the teachers, taking into account subject area taught and years of teaching experience;[7] the number of teachers selected in each school varied by site and school level. Teacher survey response rates ranged from 61 percent to 86 percent across years and sites, and school-leader response rates ranged from 56 percent to 83 percent. We applied sampling and nonresponse weights to the final survey responses so the results would reflect each site as a whole. We also conducted interviews annually with teachers and other staff in a sample of seven schools in each district and one to two schools in each CMO.

Measuring Implementation Status

Each of the next four chapters begins by summarizing the implementation status of key reform elements in each of the sites beginning in the 2009–2010 school year when sites submitted their proposals to the foundation and ending in the spring of 2014. The analyses of implementation status draw on the central-office interviews and documents

---

6 Teachers were not surveyed in 2012.

7 Specifically, we stratified based on core and noncore subject areas, in order to ensure adequate representation from teachers of all types. We defined core teachers as general-education teachers of reading and English language arts (ELA), mathematics, science, social studies, and (at middle and high school levels) foreign languages. We defined noncore teachers as teachers of other subject areas and special-education teachers. Our samples typically consisted of approximately 80 percent core teachers and 20 percent noncore teachers. In addition, we oversampled novice teachers in the districts (which have high proportions of experienced teachers) and experienced teachers in the CMOs (which have high proportions of novice teachers) to ensure adequate representation from each group.

collected from each site. Although the policies and practices that the sites have adopted vary, this report's primary purpose is to examine the initiative as a whole. Thus, these chapters focus on documenting implementation across sites, with some site-specific examples included to illustrate what implementation looked like on the ground. The report is not intended to compare sites with one another; the sites emphasized different reform elements and adopted different schedules for implementing them, so an emphasis on how sites compared with one another could create a false idea that sites are in competition with each other. Moreover, in this report, we do not offer explanations for cross-site differences in implementation except in a few noteworthy cases. Our final report, which will examine the evolution of implementation over the full course of the grants, will provide more explanations of cross-site differences.

To summarize the status of implementation in each site and the progress of implementation over time, we identified specific policies and practices that sites adopted as part of their Intensive Partnership reforms and grouped them into four broad categories of reform (or levers) corresponding to the elements of the Intensive Partnership theory of action.[8] The four levers are (1) teacher evaluation, which focuses on the development of high-quality measures of effective teaching; (2) staffing, which includes recruitment and hiring, placement, tenure, and dismissal; (3) PD, particularly support that is customized to meet teachers' identified needs; and (4) compensation and career ladders. Within each lever, we identified detailed policies or practices that were consistent with the foundation's conceptualization of the key elements of the reforms. To develop the list of specific practices, we reviewed each site's proposal and materials that the foundation produced that described the reform, and research team members who were familiar with each site identified what they thought were the core, reform-aligned practices for that specific site. This activity was informed by interviews with site central-office staff and foundation staff. We compared the lists across sites and modified the descrip-

[8] To simplify reporting, we use the term *practice* to refer to both policies and practices that were enacted to implement the levers; the bulk of the actions were changes in practice.

tions where necessary in light of site-specific terms and situations so they were general enough to apply to all sites (although all sites did not intend to implement all of the identified practices). For example, the teacher-evaluation lever included whether the site had developed an evaluation metric and whether the site's metric included (separately) scores derived from structured classroom observations, scores from student or parent surveys, and scores from value-added or other growth models. The teacher-evaluation lever also included a few other practices, such as the creation of a data warehouse to facilitate collection and use of evaluation data. The relevant chapters provide complete lists of specific practices included under each lever. Additional information about how we define each of these elements appears in Appendix B.

To describe and compare each site's progress in implementing the levers, we classified each site as implementing or not implementing the practice at each of five time points,[9] spanning the period from the time the Intensive Partnership initiative funding was awarded, in the spring of 2010, through the spring of 2014.[10] We classified the practice as implementing if it was in effect for all intended staff[11] or if it was being formally piloted for later use. Otherwise, we classified the practice as not implementing. Each year, we assigned one point for practices that were classified as implementing and zero points for practices that were classified as not implementing. We summed point values for each of the four levers over each time period and then converted them to percentages. We relied on two data sources for coding implementation status: site-produced documents, including annual stocktake reports for the foundation, as well as other Intensive Partnership reform status

[9] After we developed the implementation tables, we shared them with site leaders to confirm their accuracy. In a few cases, we made changes to our classifications in response to additional information that these site leaders provided.

[10] *The spring of 2010* describes the practices the sites had in place at the beginning of the initiative (as described in their proposals); *the spring of 2011* summarizes implementation as of April–May 2011, at the end of the first full school year after the initiative was launched. Our most recent summary, from the spring of 2014, describes implementation status as of April–May 2014, the end of the fourth school year of the initiative.

[11] We consider an evaluation measure implementing when it is obtained, or calculated, for all intended teachers regardless of when consequences were attached to the measure.

updates and interviews with central-office staff in each site from the fall of 2010 through the fall of 2014. We asked interviewees to confirm and elaborate on judgments about the status of implementation based on each site's stocktake report and Intensive Partnership status updates.

The following four chapters summarize the status of implementation by year for each of the four levers. Although the individual practices discussed in each chapter address the major components of the Intensive Partnership initiative, they should not be interpreted as the definitive list of the essential features of an Intensive Partnership reform. We have tapped the key features that were intended as part of the initiative, but we have not tried to capture every policy or practice that the Intensive Partnership sites adopted as part of the reform, and individual sites made different choices about which of these approaches to emphasize. The levers' relevance to sites also varies based on local context. For instance, CMOs that do not offer tenure could not have adopted policies linking tenure to effectiveness. As noted above, we tried to define the indicators broadly so that they would be applicable across sites (e.g., by referring to the linking of tenure or retention decisions to effectiveness), but the reader should still keep in mind that we would not necessarily expect each site to have adopted all of these practices.

Chapters Two through Five also present teacher and school-leader perspectives on the reforms. We used surveys and interviews to gather information about teachers' and school leaders' experiences with, and opinions about evaluation, staffing, PD, compensation, and career ladders. The main findings draw on our survey data. Where applicable, we supplement these results with information obtained through interviews with central-office and school-level staff.

Limitations of This Study

Because this report focuses on reform implementation through the spring of 2014, it does not provide evidence of what the fully implemented reforms will look like once the Gates Foundation funding ends, and it does not tell us anything about how the reforms have influ-

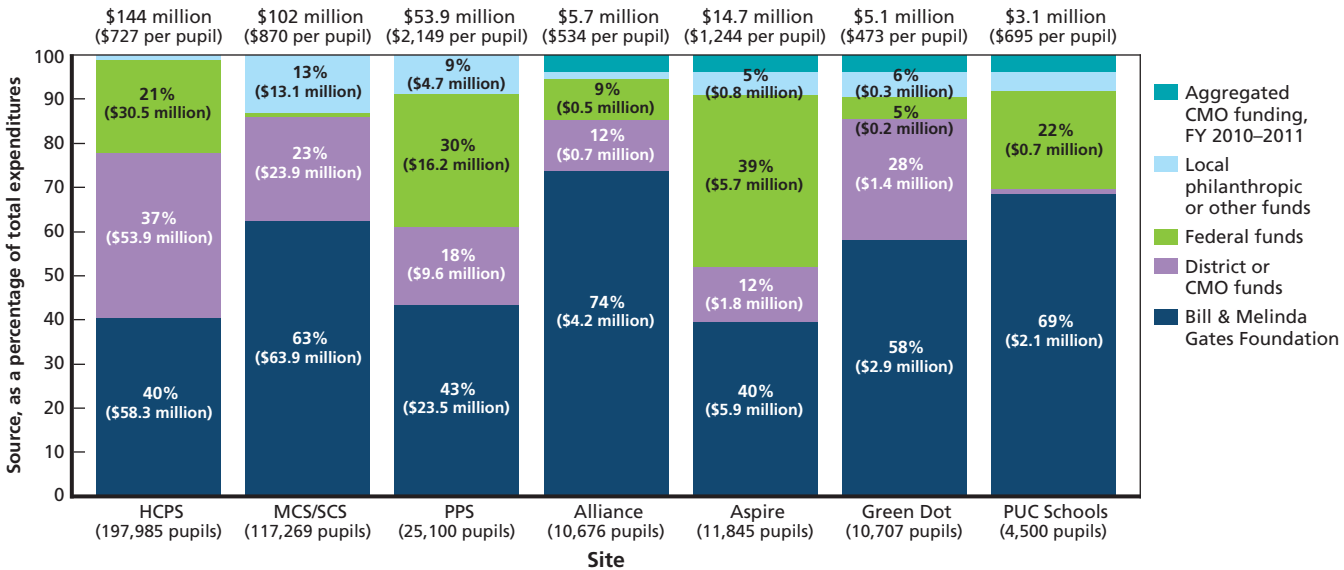
enced student outcomes. Forthcoming reports from this evaluation will address these gaps. In addition, much of the data presented in this report are drawn from self-report surveys and interviews with school and district staff, so we lack independent verification of how instructional practice and other aspects of the schools or districts actually changed. We also lack information on the quality of implementation of specific reform components. Finally, as noted above, the purpose of this report is to provide an overview of what Intensive Partnership implementation looked like four years into the initiative, and we do not attempt to understand the specific reasons for cross-site differences in implementation.

Overall Expenditures on the Intensive Partnership Initiative

The foundation pledged $290 million to support the Intensive Partnership initiative overall; from November 2009 through June 2014, it awarded more than $160 million to the seven Intensive Partnership sites. However, that is not the total cost of the effort. The foundation required that each site provide matching funds from other sources, such as local foundations or federal grants, to support the initiative. The sites also allocated some general-fund resources to support the initiative. These investments were used to pay for a variety of things, including the purchase or modification of computer data systems, the development of new procedures and management systems for evaluating teachers, changes to PD practices, payment of incentives to reward effective teaching performance, and compensation premiums for working in more-challenging school environments.

We estimated the overall site-level expenditures on the Intensive Partnership initiative from November 2009 through June 2014 (see Figure 1.2), based on a review of site financial reports, fiscal records, and interviews with administrators. The total expenditures ranged from $3.1 million in PUC Schools to $144 million in HCPS. Much of this variation is related to differences in enrollment across the sites.

Figure 1.2
Intensive Partnership Initiative Expenditures Broken Out by Funding Source, November 2009 Through June 2014



SOURCE: Intensive Partnership sites' financial reports for the fall of 2014; Chambers, Brodziak de los Reyes, and O'Neil, 2013.
NOTE: The figure displays only percentages above 5 percent. Detailed financial reports were not available for the CMOs prior to fiscal year (FY) 2012. We estimated the aggregated funding for each CMO for FY 2010–2011 by prorating the total TCRP funding in those years by each CMO's share of all CMO funding for FY 2012–2014. For SCS, we estimated federal funds using methods described in a case study conducted by Chambers, Brodziak de los Reyes, and O'Neil, 2013.

RAND *RR1295-1.2*

However, district size did not entirely explain variation because the total expenditures per pupil ranged from $473 to $2,149.[12]

During the initial five years of the initiative, funding from the Bill & Melinda Gates Foundation supported the largest portions of the expenditures, but other funds played a significant role. In relative terms, the foundation funding accounted for between 40 percent of the total funds allocated to the Intensive Partnership initiatives (in HCPS and Aspire) to 74 percent (in Alliance). Federal funds, such as RTT, Teacher Incentive Fund (TIF), and School Improvement Grants, were a significant source of funding in PPS, Aspire, and PUC Schools, and district and CMO funds were a large source of funding in HCPS, SCS, and Green Dot. The proportion of local philanthropic funding ranged from 1 percent in HCPS to 13 percent in SCS.

Organization of the Report

Chapter Two presents our findings about the sites' teacher-evaluation practices, including the effectiveness measures they implemented, the distribution of TE, teacher and school-leader reactions to the evaluation system, and estimates of the cost of implementing it in each site. Chapter Three explores changes in staffing practices, including hiring, placement, tenure, and dismissal. We also present teacher and school-leader perspectives on these practices. Chapter Four focuses on the PD practices and describes sites' efforts to customize PD and improve the effectiveness of all teachers. Chapter Five examines the implementation of compensation reforms and career ladders, describing the types of policies adopted and teacher and school-leader responses to them. Finally, Chapter Six summarizes the findings and presents conclusions about the status of the Intensive Partnership initiative to date.

[12] The large per-pupil expenditure estimates in PPS are related to a substantial decline in enrollment during this period.

CHAPTER TWO

Teacher Evaluation

Measuring and supporting effective classroom teaching is the core focus of the Intensive Partnership initiative, and the foundation's first strategic priority for the sites was developing a meaningful measure of TE. This measure is essential for the other levers to work effectively. The foundation suggested that the measure should "include growth in student learning over time, teachers' knowledge and skill, observed teaching practices, and student perceptions and levels of effort in the classroom" (Bill & Melinda Gates Foundation, 2009, p. 3). Subsequent to awarding the Intensive Partnership grants, the foundation's MET project demonstrated that it is possible to build a measure of TE with reasonable reliability and validity by combining information about student achievement growth, direct observation of teaching practice, and student feedback. Although the sites were not required to use a specific combination of measures, they were encouraged to use many of these elements. This approach to teacher evaluation is consistent with systems that states and districts have adopted to comply with RTT and other initiatives (e.g., No Child Left Behind waivers, TIF).

The sites began developing new teacher-evaluation policies as one of their first actions when they received their Intensive Partnership grants. This chapter documents the implementation of teacher-evaluation practices in each site and the distribution of effectiveness ratings these systems produced. It then examines teacher and school-leader responses to, and opinions about, the evaluation system based on results from surveys and interviews. We also present information about the cost of developing each site's teacher-evaluation system.

Teacher-Evaluation Lever Implementation

The teacher-evaluation lever includes eight specific practices:

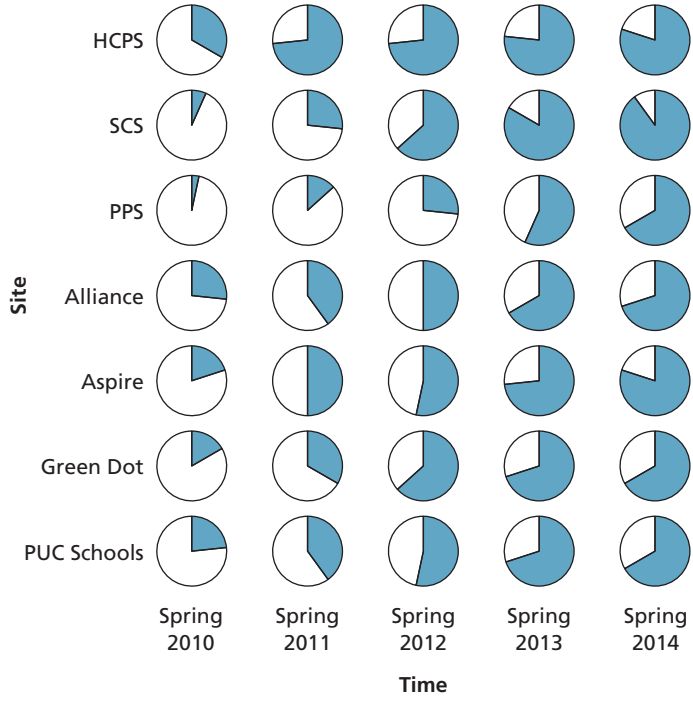
- observation by principals or other administrators
- observation by an additional set of observers (e.g., other school leaders, content-area specialists, peers, central-office administrators, coaches) for at least some teachers
- student or parent surveys
- other measures of TE (e.g., content knowledge, professionalism, peer survey)
- individual VAM or student growth percentile score for subjects and grades with state tests
- individual VAM or student growth percentile score for subjects and grades with no state test, or other measures of student growth
- multiple measures combined using weights
- data warehouse established for teacher-evaluation data.

Appendix B provides additional information about how we defined each of these elements.

Sites Took Approximately Two Years to Develop and Refine Their Teaching-Effectiveness Measures and Implement Systems to Operationalize Them

Figure 2.1 shows the proportion of practices in the teacher-evaluation lever that each site implemented annually from the spring of 2010 to the spring of 2014. As we mentioned earlier, sites did not necessarily plan, and were not expected, to implement all of the practices included as part of this lever. For example, a site could have a rigorous TE measure without input from parents or students. Thus, we should not expect all sites to achieve 100-percent implementation (i.e., a fully colored circle in Figure 2.1). However, sites were expected to redesign their teacher-evaluation systems to include multiple measures of effectiveness. The important features to notice when looking at the figure are when actions began, how quickly they progressed, and when they attained stability.

Figure 2.1
Proportion of the Teacher-Evaluation Lever Implemented, Spring 2010 to Spring 2014



RAND *RR1295-2.1*

As shown in Figure 2.1, when the initiative began, most sites had none or very few teacher-evaluation practices that were consistent with the Intensive Partnership initiative. Sites spent a year or two carefully building their teacher-evaluation systems. This time was required because site leaders wanted to ensure that teachers and other stakeholders endorsed the system; they also wanted to be sure that the measures had adequate technical quality and produced data that would support the other levers and the decisions that would be made on the basis of evaluation scores. Sites also wanted to encourage stakeholder participation and buy-in, which they believed would be facilitated by a careful, gradual rollout. To this end, all sites engaged teachers in the planning process and conducted pilot tests to refine the rubrics

and the observation process. Even though the sites adapted rubrics that had been developed elsewhere (e.g., Danielson's Framework for Teaching, the District of Columbia Public Schools' Teaching and Learning Framework), the process of adaptation was time-consuming because it entailed discussing each dimension and considering whether modifications were appropriate. For example, HCPS's revised teacher-evaluation rubric is organized around the 22 components of professional practice from Charlotte Danielson's Framework for Teaching. HCPS began working with Danielson in September 2009 to develop its new observation rubric, which it approved in June 2010 and began implementing for teachers in the 2010–2011 school year. The CMOs developed a teacher-observation rubric in school year 2009–2010, and each CMO piloted it with a few teachers in several schools in school year 2010–2011 and then refined it. It was not until school year 2011–2012 that the teacher-observation rubric was used for all CMO teachers. Similarly, observer training was an important consideration for all the sites, and they devoted considerable time to developing standards of rigor that had to be met before observers could be certified. In addition, HCPS decided to use peers as observers, and training and certifying peer observers added to the time it took to initiate the observation component.

By the second year (the spring of 2012), all sites except PPS had implemented a majority of the practices.[1] PPS originally had not planned to adopt multiple measures of effectiveness but instead planned to rely exclusively on the classroom-observation rubric, RISE. As it became clear that the Bill & Melinda Gates Foundation was emphasizing multiple measures and that Pennsylvania was going to require them, PPS adopted the Tripod student feedback survey and focused on developing its VAM.

Most sites could not compute a value-added measure for teachers working in nontested grades and subjects because their students did

[1] In some cases, sites were coded as implementing a practice even though it was still being piloted and had not yet been rolled out to its full extent. For example, in the spring of 2011, SCS was piloting its teacher-evaluation system but had not fully implemented it yet because the state law was not in effect.

not take annual statewide achievement tests. Most sites did not have the capacity or resources to develop local tests for these students, so several opted to compute a school-level growth score and assign it to teachers of untested students. For example, most of the CMOs use a school-level student growth percentile (SGP) for teachers of nontested classes. Most of the CMOs intend to develop individual measures for nontested subjects and grades, but, for the most part, lack of resources has postponed their development. In contrast, HCPS has administered local examinations in all classes for many years, and it opted to continue this practice and use these results to compute student growth scores for teachers whose students did not take the state tests.

All Sites Included Classroom Observations and Measures of Teachers' Contribution to Student Achievement Growth in Their Evaluation Systems

Together these two components make up between 75 percent and 100 percent of a teacher's score. All sites but HCPS also included student surveys (Tripod) in their formal evaluation systems, but scores from these surveys receive much less weight than the student achievement and observation measures. For example, PPS assigns 15-percent weight to Tripod scores, and SCS assigns only 5 percent.

By the spring of 2014, all sites had implemented all of the teacher-evaluation practices that they intended to implement. The fact that some 2014 circles are not completely filled in reflects site intentions and local conditions. For example, HCPS did not implement student surveys as part of its formal evaluation system because of teacher concerns about validity of these data, and PPS did not implement additional measures of TE beyond observations, student surveys, and VAM because district leaders believed that the measures they did adopt were adequate (see Figure 2.1). The proportion of evaluation practices implemented in the CMOs (except for Alliance) declined from 2013 to 2014 because California did not report results from a new statewide test in the spring of 2014, so SGPs could not be calculated for teachers that year.[2]

[2] Alliance used 2013 state test scores along with previous years' scores to calculate an SGP.

Distributions of Teaching-Effectiveness Ratings

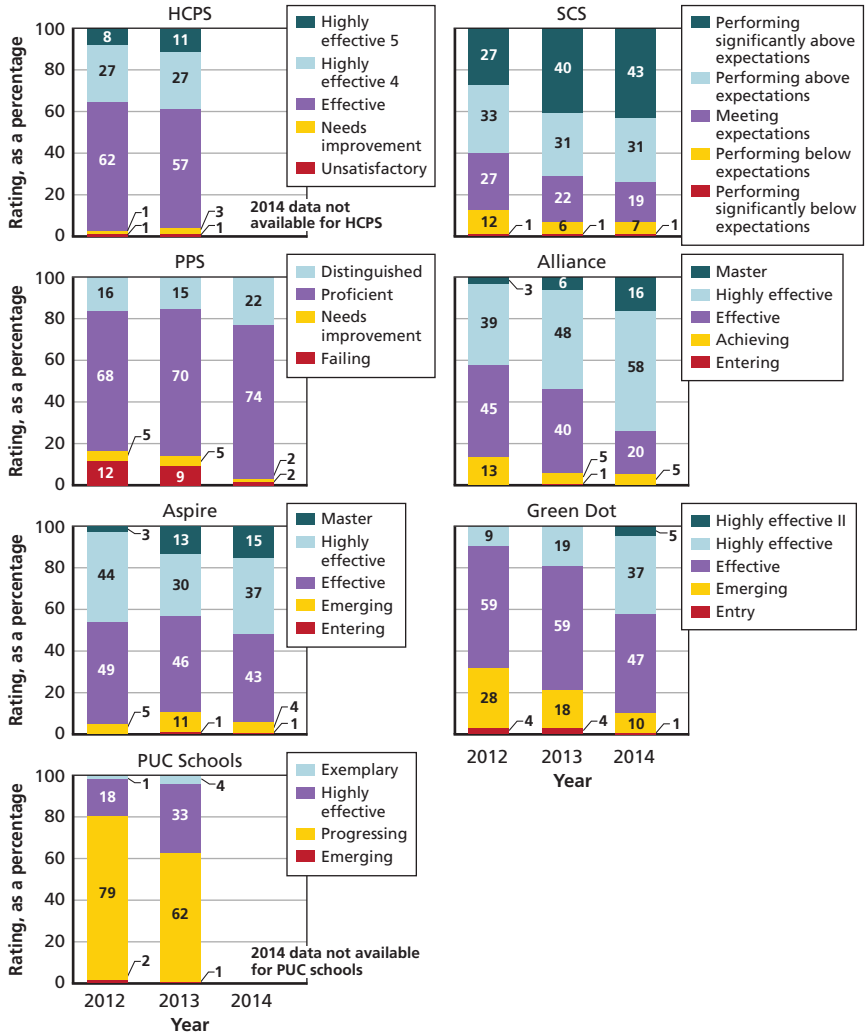
Across Sites, a Substantial Majority of Teachers Received Ratings Equivalent to Effective or Highly Effective, and the Percentage of Teachers Performing as Effective or Higher Increased from 2012 to 2014 in Most Sites

Figure 2.2 shows the percentage of teachers classified at each performance level over time by site, for the 2011–2012, 2012–2013, and 2013–2014 school years. Two of the sites reported four levels of effectiveness, while the others reported five levels. It is worth noting that, at the beginning of the initiative, the sites were similar to the rest of the country, i.e., few teachers were assigned to the bottom category, where they would be placed on improvement plans or be at risk of negative consequences, such as termination (Weisberg et al., 2009). Figure 2.2 shows that, in all sites, the percentage of teachers in the top category has increased over time, while the percentage of teachers in the bottom category has declined. One of the goals of the Intensive Partnership initiative is to increase the prevalence of effective teaching, and these data are consistent with that goal. On the other hand, during this period, some of the sites made changes to the computation of TE scores or to the cut points associated with each category, and this might have affected the distribution of effectiveness ratings. Future analyses will explore the shifts in the distribution and try to understand the mechanism through which change has occurred (e.g., more-effective recruitment and induction, elimination of the lowest-performing teachers, better retention of highly effective teachers, improvement among all teachers).

Teacher and School-Leader Perspectives on Teacher Evaluation

A rigorous measure of effective teaching is a core element of the Intensive Partnership initiative; this measure informs most of the other elements. Examining teacher and school-leader perspectives about the effectiveness measure—and the other elements—provides insights on how well implementation is going from the practitioner point of view

Figure 2.2
Effectiveness Rating Distributions, by Site, School Years 2011–2012, 2012–2013, and 2013–2014



SOURCE: Our tabulations of TE scores reported by each Intensive Partnership site.
NOTE: The ratings were assigned to teachers in fall of 2012, 2013, and 2014 based on data collected during the prior two years. Each bar with a date reflects a different school year.

RAND *RR1295-2.2*

and whether the sites might be facing challenges related to buy-in. As noted in Chapter One, research suggests that educator support for reform initiatives can influence the quality of implementation and the extent to which reforms achieve their goals. In the sections that follow, we draw on survey and interview data to describe teachers' and leaders' opinions about the validity of these measures, how they were used, their effects on instruction, and whether they were fair to all teachers.

Teachers and School Leaders Were More Likely to See Observation Ratings as Valid Indicators of Effectiveness Than Either Student Achievement Gains or Student Survey Responses

Most teachers reported that their performance was being evaluated, particularly in the 2012–2013 and 2013–2014 school years. Observations of teaching, student achievement or growth, and student input (e.g., from surveys) were among the most–commonly reported components of teachers' evaluations.[3] Among these three components, teachers were more likely to report that observations were valid indicators of the effectiveness of their teaching than that the other two components were or, notably, than all components combined (see Figure 2.3). One interpretation of this is that teachers might think that the combined measure is only as valid as its least-valid component.

Observation Component

Teachers Had Generally Positive Perceptions About the Observations, but Some Expressed Concerns About the Suitability of the Observation Rubric for Measuring Different Forms or Styles of Good Teaching and the Qualifications of Particular Observers

Nearly all teachers surveyed in all seven sites reported that observations of their teaching were part of their evaluation. In general, reactions toward the observations were positive on a range of dimensions. In addition to the positive response about the validity of observations

[3] Other commonly reported components included "your professional conduct, behaviors, and responsibilities" and, in the CMOs particularly, parent input and feedback.

shown in Figure 2.3, in both 2013 and 2014, most teachers agreed with the following statements included on the survey:

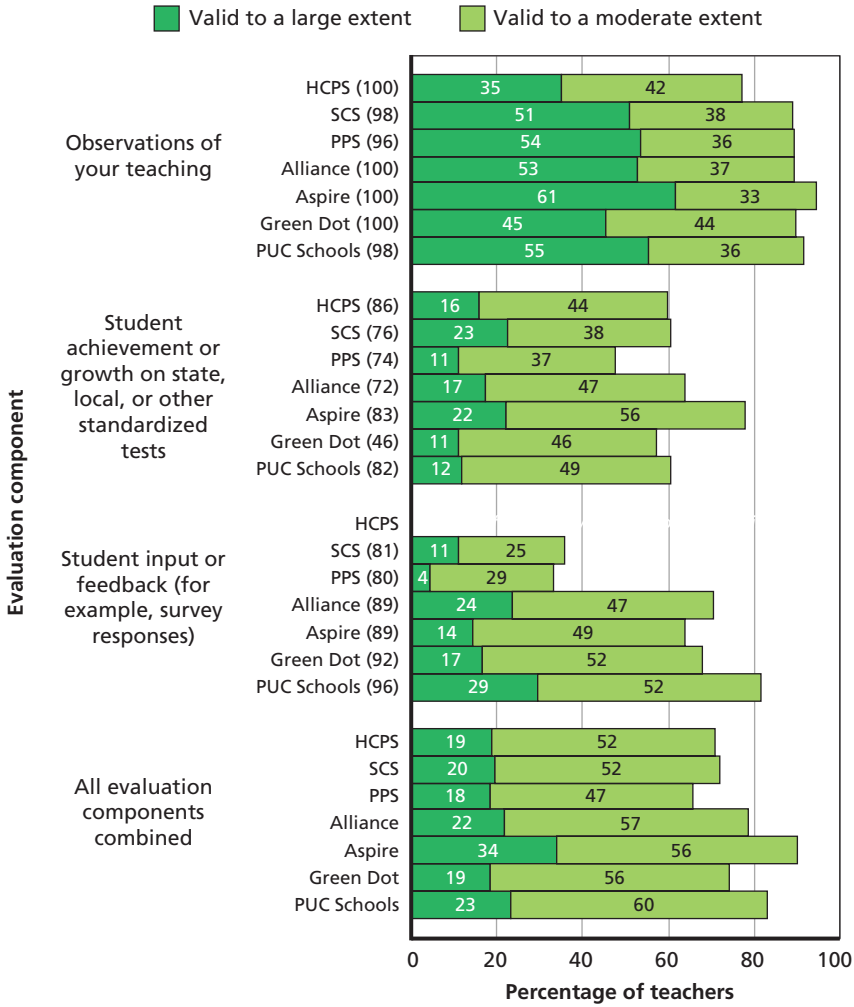
- "I have a clear sense of what kinds of things the observers are looking for when they observe my teaching" (83 percent to 94 percent, depending on the site and the year).
- "The people who observe my teaching are well qualified to evaluate it" (63 percent to 89 percent; without HCPS, 76 percent to 89 percent).
- "After my teaching is observed, I receive useful and actionable feedback" (65 percent to 87 percent).
- "I have a clear understanding of the rubric that observers are using to evaluate my teaching" (82 percent to 92 percent).
- "I have made changes in the way I teach as a result of feedback I have received from observers" (79 percent to 95 percent).

However, teachers were somewhat less likely to agree that the observation rubric was well-suited for measuring many different forms or styles of good teaching (48 percent to 75 percent) and that the observations were long enough and of sufficient frequency to provide an accurate view of their teaching (52 percent to 90 percent; without SCS, 52 percent to 80 percent).[4] Teachers in HCPS also tended to have markedly lower levels of agreement than teachers in the other six sites had about observers being well qualified and about receiving useful and actionable feedback.

In interviews with school-level staff, most teachers across sites agreed that they generally received feedback that was helpful for improving their teaching, but they also raised questions about the accuracy and validity of their observation scores. In 2013 and 2014, many interviewed teachers expressed concerns about observer subjectivity and lack of consistency in how different observers assigned ratings, and these views were particularly prevalent among teachers whom multiple evaluators had observed. A few teachers also noted that the

[4] It is worth noting that some sites (e.g., PPS and SCS) reduced the length or number of observations as the reform matured, in large part to reduce the burden on principals.

Figure 2.3
Percentage of Teachers Reporting That Evaluation Components Are Valid to a Large or Moderate Extent, School Year 2013–2014



NOTE: The numbers in parentheses next to the site names are the percentages of teachers (among those who reported being evaluated) who indicated that the component was part of their evaluation. All of the Intensive Partnership sites except HCPS include student input as a component in teachers' effectiveness ratings.

RAND *RR1295-2.3*

design of the observation rubrics required teachers to address every element during a single observed lesson, which was sometimes unrealistic. A teacher in Aspire, for instance, noted the value of the observation and feedback process but added,

> My only real complaint is that it's designed for you to see everything in one class period, especially in science. A lot of people do a dog and pony show, but I'm not going to do that and pretend that I do this every day. The way it's set up, if you have no evidence, it's a zero. I don't think that's a great way of setting it up.

Another teacher in HCPS shared,

> It still feels just like a snapshot to me. . . . I can put on a show for one day; something could go wrong for 20 minutes. I just don't feel like it gives a full representation even if I have five or six evaluations. To me, that's still a snapshot.

An additional HCPS teacher commented,

> I feel like the peer evaluators come in to only find things wrong with you. I feel like the district is training them—I'm going to be honest here and share my true feelings—to purposely only find things wrong so that, especially now, this year, we got raises, and our pay is tied to how we scored. . . . I think that's kind of a general consensus around here that everyone did so poorly compared to other years. I also feel like, when my team and I talk, after our formals and we're like, "well, what did she say to you?" we all kind of see the same things, so I find that kind of odd that, if one of us needs to work on higher-order questions, every one of us needs to work on turn and talk. Really, all five of us on my team screwed that up? I feel like they come in with like a certain thing and just come in to talk about that with everybody for the most part. That means you're just looking for certain things. You're not really judging the individual teacher on what she's actually doing.

Furthermore, several teachers also stated that their observers lacked classroom teaching experience in the subjects they were evaluat-

ing, which teachers believed limited the accuracy of the ratings and the utility of the feedback. As one Alliance teacher noted, "How do I feel confident about the advice they give me if they have never had to put it into practice themselves?" An HCPS teacher emphasized,

> I think it would be much more effective if our peers and mentors were teaching at a Title I school or came from a Title I school because some of the higher-order questions for us are regular questions in a school with a higher socioeconomic background, so some of our teachers get dinged. It's hard to align that with someone who has not been in your shoes.

Most School Leaders Reported That They Were Adequately Trained to Conduct Observations

In all seven sites, at least eight out of ten school leaders reported in 2014 that they themselves observed teachers' instruction as part of the teachers' evaluation. Most school leaders reported receiving training on how to observe classrooms (95 percent, on average across the sites), and, of those, most reported that the training provided opportunities to practice observing (88 percent) and covered how to assign scores based on their site's rubric (93 percent). Fewer reported that the training adequately covered "ways to deal with challenging situations" (66 percent) or helped them understand how to identify PD opportunities for teachers (75 percent).

Student Achievement Component

Many Teachers Questioned the Specific Methods and Measures Used to Incorporate Student Achievement into Their Evaluations

Teachers' reactions to the use of student achievement in their evaluations were more mixed. Across all seven Intensive Partnership sites, a majority of teachers surveyed indicated that they had a clear understanding of how student test scores were used to evaluate their performance, and the percentages increased from 2013 to 2014, suggesting that understanding might be on the rise. The percentages of teachers agreeing that "the student tests used in my evaluation measure important skills and knowledge" also rose from 2013 (62 percent) to 2014

(66 percent). Most teachers also agreed that the tests covered the right topics and were aligned with curriculum. In particular, among teachers who reported teaching a tested subject area and grade level (which, in 2014, ranged from 32 percent in Green Dot up to 75 percent in HCPS), we found the following:

- About half agreed that "scores on the student tests used in my evaluation are a good measure of how well students have learned what I've taught during the year," but fewer than 10 percent agreed strongly. Pittsburgh teachers tended to have the lowest levels of agreement.
- Sixty-six percent agreed that "the student tests used in my evaluation are well aligned with my curriculum." Again, Pittsburgh teachers had the lowest levels of agreement.

However, in both years, only 40 percent of teachers agreed that "the ways that student test scores are used to evaluate my performance appropriately adjust for student factors not under my control."

According to teacher survey respondents, tests are having an impact on teaching practice, although we cannot say whether this is due to their being part of TE measures or other accountability pressures. Among teachers who reported teaching a tested subject area and grade level, we found the following:

- About 75 percent indicated that they "devote more attention to subject-matter content that is tested than to content that is not tested"; 65 percent reported that they devote "significant class time to test-preparation activities." In the three districts, the percentages rose slightly from 2013 to 2014 on both items, but all four CMOs had substantial declines.
- About 80 percent of surveyed teachers reported that they had made changes in what (or how) they teach based on data from the student tests used in their evaluation. SCS teachers had a noticeably higher percentage of teachers reporting this, with 90 percent agreeing (compared with 78 percent across the other six sites) and

47 percent agreeing strongly (compared with about 25 percent across the other six sites).

In interviews with PPS teachers during the 2012–2013 and 2013–2014 school years, most teachers with whom we talked expressed numerous concerns about the district's curriculum-based assessments (CBAs), one of the tests used to calculate their VAM scores. In particular, teachers reported that the CBAs were not well aligned to the curriculum in the sense that they included material that had not yet been taught or that the CBAs contained numerous errors and were therefore invalid. Lack of trust in the CBAs was, according to most teachers, one of the key reasons they distrusted the district's value-added measures.

In the CMOs, although most teachers we interviewed appreciated the effort to measure student growth, some questioned the validity of the SGPs for teachers of nontested subjects who must rely on school-level scores and for teachers whose students lacked baseline scores in the subjects they taught. An interviewed Aspire teacher said,

> It doesn't mean much to me, to be honest, because it is world history scores in 10th grade being compared to English scores in 9th grade. So any gains are not based on building history skills or knowledge, necessarily. And the last time students might have taken world history is back in 7th grade. So I think the growth number is a little arbitrary in terms of how well I teach history.

Some teachers in HCPS also continued to have questions about how VAM is calculated. For example, one said,

> I would like to understand the value added and how they came up with that because I don't understand. . . . It's kindergarten; we don't have a standardized test in kindergarten, so we should have a different scale because it's not like the other grades.

As these quotes suggest, many teachers were not opposed to the general principle of measuring achievement growth but questioned the validity or appropriateness of the specific measures in use in their sites.

Student Feedback Component

Although Most Teachers Thought That Student Feedback Was Potentially Useful to Them, They Did Not Think That It Was an Appropriate Measure of Their Effectiveness

The use of student input in teacher evaluation has been a source of concern to many teachers.[5] Large majorities of surveyed teachers in every site indicated concerns that students do not understand the questions they are asked about their teacher or class or that too many students do not take the feedback opportunity seriously (especially in middle and high schools). Similarly, lower percentages of teachers—fewer than half in PPS and SCS—thought that students are good judges of a teacher's effectiveness and agreed that they trusted students to provide honest, accurate feedback about their teaching. At the same time, however, majorities of teachers—and high majorities in the CMOs—reported that they would consider making changes to their teaching based on feedback from their students. Many teachers—about 75 percent in HCPS and the CMOs and about half in SCS and PPS—also reported that the student feedback results help them understand their strengths and weaknesses as a teacher; novice teachers were more likely to say this than experienced teachers were.

Similarly, during our interviews, most teachers reported that they found their student survey results helpful for understanding students' perceptions of their class and for getting a sense of whether their students were learning anything, although some teachers added that these results were not specific enough to be actionable. As one PPS teacher said,

> Yes, [Tripod results are] helpful. . . . I'm interested to know what my students' perceptions are of my class. . . . Any answers that surprise you, it's in my nature to think, "What can I do differently?" I can't specifically pinpoint in my Tripod [where this happened]. I had a lower number in "Care" than I expected, and I . . . looked at the question breakdown, but [I] don't know how to improve in that area [based on Tripod data alone].

---

[5] All of the Intensive Partnership sites except HCPS include student input as a component in teachers' effectiveness ratings.

However, as the survey results suggest, many teachers we interviewed felt strongly that student feedback should not be part of their formal evaluations, in part because of concerns about inaccuracy or bias or because the process unfairly favored teachers with certain personality traits rather than those who were more effective. It appears that teachers see potential value in student feedback but do not trust it to be used as part of a formal accountability system.

Teachers in PUC Schools were somewhat more sanguine than teachers in the other sites about the use of student feedback in their evaluations. They were more likely to say that they trust their students to provide honest feedback and that students are good judges of teaching effectiveness than they were to express worries that students do not understand the questions or take the feedback opportunity seriously.

Uses of Evaluation

Both Teachers and School Leaders Were More Likely to Report That Evaluation Results Were Used for Instructional Improvement—Feedback, Professional Development, and Support—Than for Punitive, Remunerative, or Other Purposes

In all seven Intensive Partnership sites, large majorities of teachers expected that their evaluation results would be used to provide them with feedback they could use to improve their instruction. Slightly smaller majorities expected that the results would be used to identify areas for PD and to determine whether they needed instructional support (such as from an instructional coach). Nearly all school leaders reported that teacher-evaluation results would be used for these purposes.

Teachers and school leaders were less likely to indicate that teacher-evaluation results would be used for the kinds of "carrot and stick" purposes that tend to receive attention in the mainstream press and high-level policy dialogue. Just under half of the teachers reported that evaluation results would be used to determine whether they entered into some type of probationary status, and just over half agreed that the results would be used to determine whether they were qualified to continue teaching. Majorities of teachers in the CMOs and in HCPS

(but not SCS or PPS)[6] reported that their evaluation results would be used to determine whether they received monetary bonuses, and a large majority of teachers in Aspire reported that evaluation results would be used to determine salary increases or promotions and career-ladder placement.[7] Fewer than one in four teachers (on average across the sites) expected that their evaluation results would be used to make decisions about school placements or about teaching assignments within their current school, although the latter was cited more commonly by school leaders, especially in HCPS and SCS.

Few teachers reported that their evaluation results would be used to provide information to parents or the general public about the quality of their teaching, but, because, to our knowledge, none of the sites makes individual teachers' evaluation ratings publicly available in this way, it is perhaps surprising that *any* teachers—about 20 or 30 percent, depending on the site—reported this expectation. About another 20 percent reported that they did not know whether evaluation results would be used for this purpose.

Teachers Expressed Concerns About the Consequences Tied to Evaluation Results

Despite the relatively low percentages of teachers who reported the use of evaluation results for high-stakes purposes, such as termination or dismissal decisions, salary increases, and school placements, there was still widespread concern among teachers about the consequences of evaluation. In the three districts, fewer than 40 percent of teachers agreed that "the consequences tied to teachers' evaluation results are reasonable, fair, and appropriate"; the percentages agreeing were higher in the CMOs (about 60 percent), but, in six of the seven sites (all except Aspire), fewer than 10 percent of teachers agreed *strongly*. Not surprisingly, teachers who had received high effectiveness ratings were more likely to agree than lower-rated teachers. School leaders were also much more likely than teachers to agree. Interviews suggest that

---

[6] In PPS, this finding might stem from the fact that only teachers hired after July 2010 were eligible for merit-based salary increases.

[7] In 2014, Aspire moved to a pay-for-performance salary schedule based on TE levels.

some of these perceptions reflected inaccurate impressions of how the TE data could be used. Several PPS teachers, for instance, expressed fear that they could lose their jobs or fail to receive a pay increase as a result of one poor evaluation, despite the fact that neither of these consequences would have been imposed on the basis of a single evaluation. These teachers also indicated a lack of trust in district administration, which might have contributed to their fears.

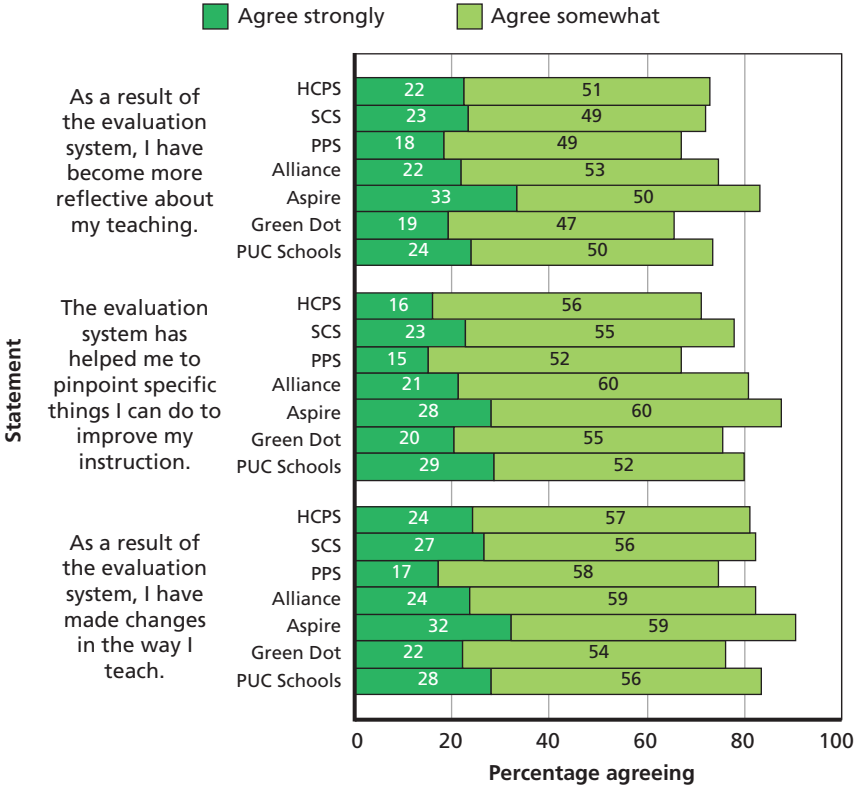
Most Teachers Indicated That Their Site's Evaluation System Made Them More Reflective About Their Teaching and Helped Them Identify Specific Improvements

As shown in Figure 2.4, majorities of teachers—70 to 80 percent—in all seven sites indicated that the evaluation system had influenced their instruction. (Two of the three items in Figure 2.4 also appeared on the 2013 survey, but there were no overall differences in responses from 2013 to 2014.) Many of the teachers we interviewed told us that the dimensions of practice included in their sites' observation rubrics had become the dominant language for talking about instruction with peers and school leaders and that having a common language to discuss instructional strengths and challenges was beneficial. So even though Figure 2.4 refers to the evaluation system as a whole, our interviews suggest that most of the perceived positive effects on instruction stem from the use of the observation rubrics.

In the Three Districts, Fewer Than Half of Teachers Thought That the Evaluation System Would Benefit Students in the Long Run, and, in All Seven Sites, the Percentages of School Leaders Agreeing Strongly About Long-Term Student Benefits Have Declined Markedly over the Past Three Years

Compared with the percentages of teachers who reported making changes to their instruction, much lower percentages, especially in the three districts, reported that they thought that students would benefit "in the long run" (see Figure 2.5). School leaders were much more likely than teachers to agree that students would benefit in the long run, but the percentages agreeing strongly declined from 2012 to 2014 in all seven sites (see Figure 2.6). It is hard to explain these results. They could be due in part to teachers' growing frustration with the

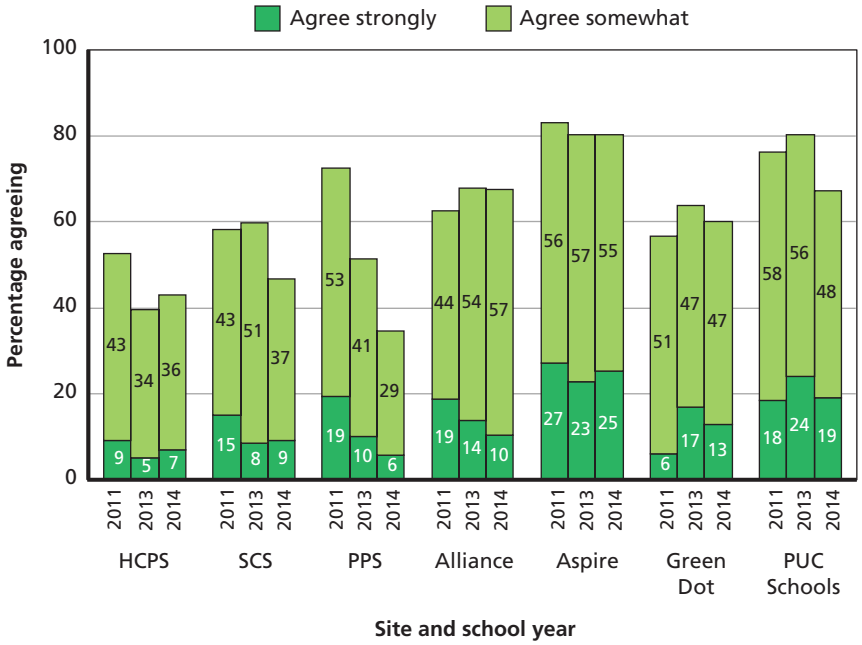
Figure 2.4
Percentage of Teachers Agreeing with Statements About the Effects of Evaluation on Their Teaching, School Year 2013–2014



RAND *RR1295-2.4*

evaluation system. In interviews, teachers expressed concerns about the observation component—that observations are too time-consuming, require too much preparation, and require them to teach to a checklist that is not best for students. On the 2014 survey, novice teachers responded more positively to this question than experienced teachers did, which might support the notion that teachers are growing frustrated with the evaluation system over time. The proportion of novice teachers is higher in the CMOs, which might explain the higher percentages of positive responses in these organizations. The strength of

Figure 2.5
Percentage of Teachers Agreeing That, "in the Long Run, Students Will Benefit from the Teacher-Evaluation System," Springs of 2011–2014



NOTE: The survey data were collected in the spring of each year, and they reflect opinions during that school year, which began the previous fall. So *2011* refers to the 2010–2011 school year, for example. Each bar with a date reflects a different school year.

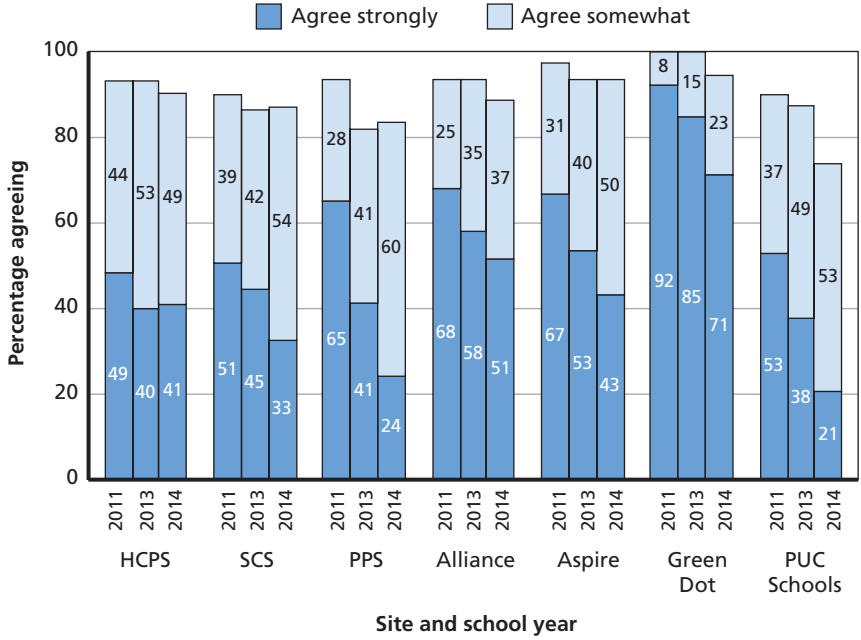
RAND *RR1295-2.5*

school leaders' opinions declined, although they still agreed that the system would benefit students in the long run.

Uses of Individual and Composite Measures for Personnel Decisions Are Influenced in Part by the Availability of Data at the Times When Decisions Needed to Be Made

The time required to generate each individual measure varies, with student achievement measures requiring the most time as a result of the need to obtain test scores and calculate the VAM or SGP. Sites took different approaches to addressing the timing. In PPS, for example, a teacher's VAM score in a given year excludes data from that school

Figure 2.6
Percentage of School Leaders Agreeing That, "in the Long Run, Students Will Benefit from the Teacher-Evaluation System," Springs of 2012–2014



NOTE: All survey results were collected in the spring of the stated year, so they are associated with practices in the school year that started the previous fall. Each bar with a date reflects a different school year.

RAND *RR1295-2.6*

year; for example, a teacher's school year 2013–2014 individual VAM score would include data from the 2012–2013, 2011–2012, and 2010–2011 school years. In SCS, teachers receive two reports—one at the end of the school year that includes observations and Tripod scores from the current school year and VAM and achievement measures from the prior year. Each receives an updated report in the fall that includes current VAM and achievement scores. In the CMOs, teachers' composite effectiveness scores are not available until November or December of the following year because SGP calculation relies on the assessment scores from the Los Angeles Unified School District (LAUSD) as a comparison group, and these scores are not available until the fall fol-

lowing the end of the school year. However, teachers have immediate online access to their observation scores. Student and family survey scores are typically reviewed at the end of the school year.

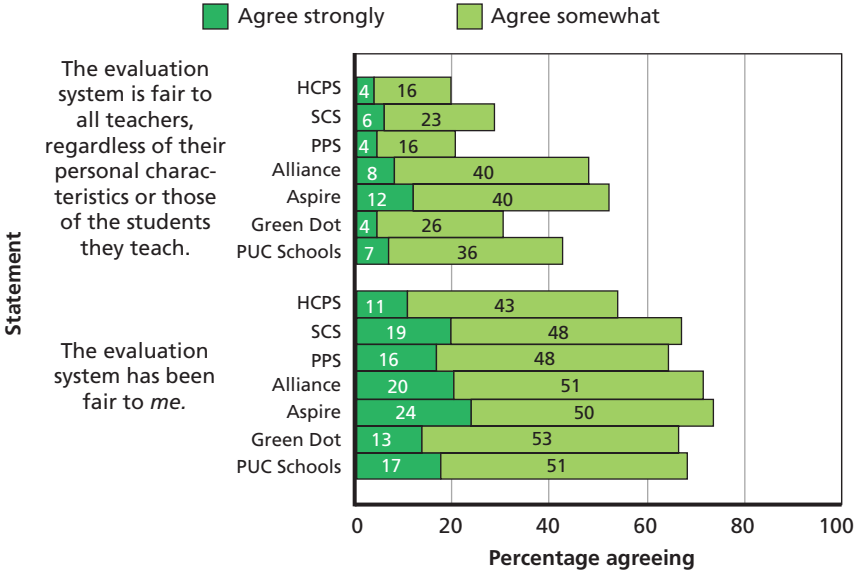
The availability of the measures influences the ways in which evaluation scores are used for personnel decisions. In SCS, personnel decisions are generally made based on the data available at the time and, in some cases, are revised when complete data are available. For example, a teacher's Teacher Effectiveness Measure (TEM) score at the end of the school year determines the number of observations he or she will receive, i.e., the observation "track" on which the teacher will be placed for the following year. This track is adjusted in October if the teacher's final TEM score, which would include a current VAM score, is different. SCS relies entirely on observation scores rather than on the cumulative TEM score to identify teachers who are in need of additional support during the current school year. In PPS, where the composite score that the teacher receives at the end of the school year is final, that end-of-year score is used to determine whether a teacher is rewarded, placed on an improvement plan (called "intensive support") for the next school year, or, lacking improvement over two years, dismissed.

Perceptions of the Fairness and Accuracy of the Evaluation Results

Most Teachers Reported That the Evaluation System Was Fair to Them Personally, but Fewer Reported That It Was Fair to Teachers Overall

In most of the sites, 50 percent or fewer teachers agreed that the evaluation system was fair to "all teachers," but higher percentages indicated that the system had been fair to them personally (Figure 2.7). This might reflect the fact that most teachers receive high ratings but are aware that other teachers do not. Thus, although they are satisfied with their ratings now, they worry that they could receive low ratings in the future. Also, although teachers endorse the evaluation system overall, they often have concerns about individual components (e.g., student surveys), which could influence their judgment about fairness

Figure 2.7
Percentage of Teachers Agreeing with Statements About the Fairness of the Evaluation System, School Year 2013–2014



NOTE: The second question was asked for the first time in 2014, so we cannot examine changes from prior years.

RAND *RR1295-2.7*

overall. Novice teachers were more likely than experienced teachers to agree with the statement about overall fairness in three of the sites and with the statements about personal fairness in two of the sites. Sixty to 80 percent of all teachers thought that "teachers who teach students who came into their class already performing at high levels have an advantage" in the system. In the three districts, high percentages of teachers—85 percent in HCPS—agreed that, "[e]ven if there are many highly effective teachers in a school, there is pressure to only rate a small number of them as very highly effective." However, there was less agreement with this statement among teachers in the CMOs, particularly in Aspire (40 percent).

During the school visits, most teachers described the evaluation system as reasonably fair to them. For example, one HCPS teacher shared this opinion: "I feel like it's as fair as it can be. Up to this point,

I feel like it's been very fair, even with some of the low scores that I've gotten."

In addition to concerns about the fairness of the system, many teachers also seemed to have concerns about the accuracy of the system. Only about half of teachers in the three districts, Green Dot, and PUC Schools (higher in the other two sites) agreed that "the way my teaching is being evaluated accurately reflects the quality of my teaching," and everywhere except Aspire, 10 percent or fewer agreed strongly. Similarly, just over 70 percent of teachers in HCPS and PPS and about 63 percent in the CMOs and SCS agreed that "the evaluation ignores important aspects of my performance as a teacher." Of those agreeing, almost half agreed strongly.

Higher-Rated Teachers Were More Likely Than Lower-Rated Teachers to Think That Their Rating Was Accurate

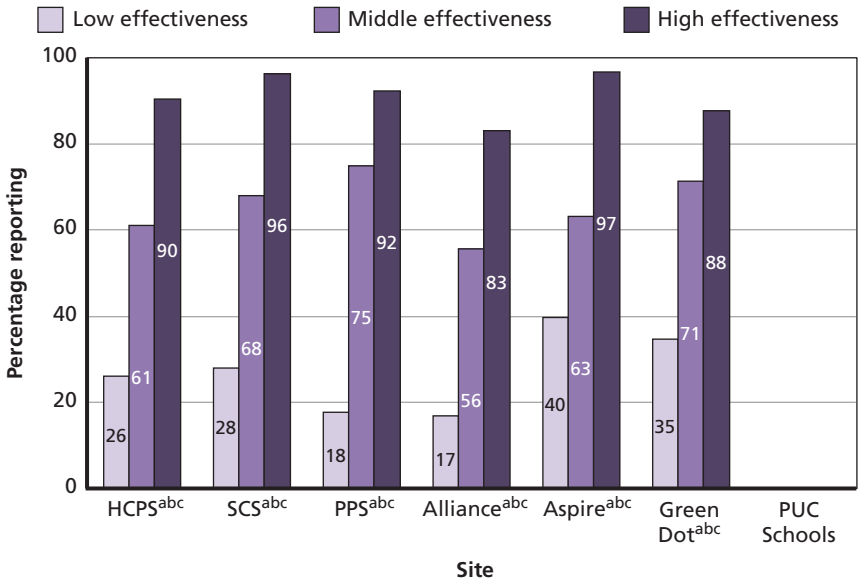
Despite the general concerns about the fairness and accuracy of the system, majorities of teachers who received an overall evaluation rating for the previous school year (2012–2013) thought that the rating was at least moderately accurate. Not surprisingly, the higher the rating a teacher received, the more likely the teacher was to say that the rating was accurate (see Figure 2.8). But even among teachers who received the lowest ratings—of whom there were not many—some thought that their rating was moderately or highly accurate.

Cost of Teacher Evaluation

In Chapter One, we presented estimates of the overall cost of the Intensive Partnership initiative in each of the sites from inception to 2014. Here we focus on just the cost of developing the new teacher-evaluation systems. Most of this effort occurred between 2009 and 2012, at which point the evaluation systems were operational. Thus, this section focuses on that time period and should be thought of as representing the "start-up" and early operational costs.[8] To understand

[8] In a future report, we will estimate the ongoing annual cost of maintaining the initiative.

Figure 2.8
Percentage of Teachers Reporting That Their Prior Year's Overall Evaluation Ratings Were Moderately or Highly Accurate, by Effectiveness Rating, School Year 2013–2014



NOTE: Evaluation ratings not available for PUC Schools.
[a] The difference between teachers in the low category and those in the middle category is statistically significant at $p < 0.05$.
[b] The difference between teachers in the low category and those in the high category is statistically significant at $p < 0.05$.
[c] The difference between teachers in the middle category and those in the high category is statistically significant at $p < 0.05$.
RAND *RR1295-2.8*

the magnitude of effort required to design and implement new teacher-evaluation systems, we conducted three case studies in the larger sites, HCPS, SCS,[9] and PPS.[10] We used fiscal data from each Intensive Part-

[9] The case study for SCS was based on information gathered in MCS prior to the merger.

[10] The three Intensive Partnership sites were at different points in the implementation of the new evaluation system because of the way their local systems are structured, the existing capacity, and the strategies that the districts selected. By school year 2010–2011, HCPS and PPS had rolled out the new teacher-observation system, and SCS implemented it for all classroom teachers in the following year. In HCPS, the first value-added calculations were

nership site as the primary source from which to estimate the expenditures allocated to the new evaluation system during the start-up period. We also conducted interviews with central-office staff to better understand the implementation process and the investments made to create the teacher-evaluation systems. Lastly, we calculated cost estimates of the additional time that school leaders and teachers spent on evaluation activities using time-allocation data from the surveys and compensation data that the sites provided. We developed separate estimates of the expenditures required to implement three components of the evaluation system: (1) classroom observations, (2) VAMs, and (3) student surveys.

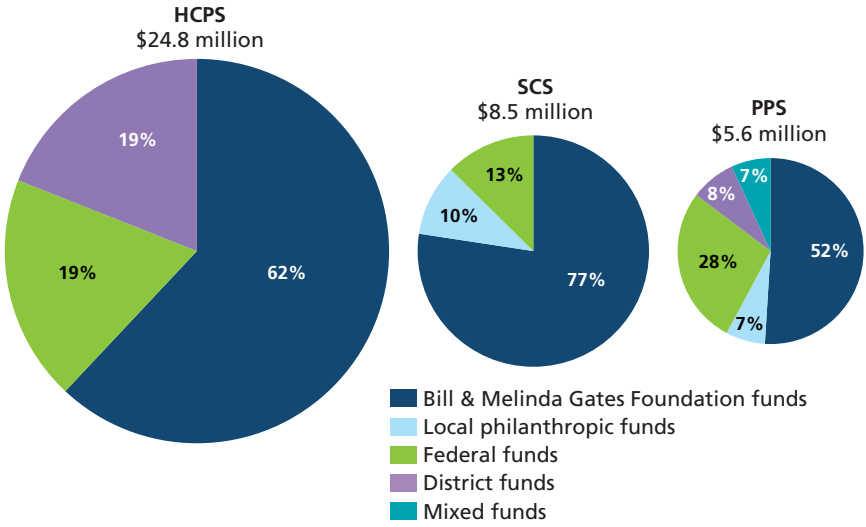
The Total Estimated Teacher-Evaluation System Implementation Expenditures in the Three Intensive Partnership Districts Is $38.9 Million; the Bill & Melinda Gates Foundation Grants Funded a Majority of These Expenditures (52 to 84 Percent)

The total estimated evaluation system expenditures from November 2009 to June 2012 amounted to $24.8 million in HCPS, $8.5 million in SCS, and $5.6 million in PPS (see Figure 2.9). In HCPS, almost two-thirds of the expenditures for the teacher-evaluation system were paid out of the Bill & Melinda Gates Foundation funds, and the other third was paid out of federal funds (i.e., RTT) and reallocated district funds. In SCS, foundation grant funds represented a larger share of the expenditures, 77 percent, and federal funding from Title I funds made up the second-largest share. PPS had the lowest proportion of foundation grant funds allocated to the implementation of the teacher-evaluation system, 52 percent, and the remainder of funds came from a combination of federal funding (28 percent) and the reallocation of district money.

---

released in the fall of 2011, whereas, in PPS, value-added scores were calculated for school year 2011–2012 for almost 40 percent of teachers and in SCS only for teachers in core subjects. Student surveys were incorporated into the teacher-evaluation systems in PPS and SCS in school year 2011–2012.

Figure 2.9
Funding Sources for Implementing the Teacher-Evaluation Systems in Hillsborough County Public Schools, Shelby County Schools, and Pittsburgh Public Schools, November 2009 to June 2012



SOURCE: Adapted from Exhibit B in Chambers, Brodziak de los Reyes, and O'Neil, 2013.
NOTE: *Mixed funds* refers to a combination of TIF, Title I, and district funds.
RAND *RR1295-2.9*

The Teacher-Observation Component Was the Most Expensive Component of the Teacher-Evaluation System to Implement in the Districts

The three Intensive Partnership districts spent more on activities related to the teacher observations than on activities associated with developing and implementing the VAM or the student survey components (see Table 2.1). Specifically, HCPS and SCS spent more than 80 percent of the evaluation system expenditures during this period on their classroom observations, while PPS spent about 48 percent.

Most of the funds that HCPS devoted to supporting teacher observations were used to employ teachers as full-time observers. SCS recruited instructional facilitators to help administrators carry out teacher observations. In contrast, PPS used only principals and assistant principals to conduct the observations as part of their regular job

Table 2.1
Overview of Expenditures on the Teacher-Evaluation Systems in Hillsborough County Public Schools, Shelby County Schools, and Pittsburgh Public Schools, November 2009 to June 2012

Expenditure	HCPS	SCS	PPS
Total evaluation system expenditures, in millions of dollars	24.8	8.5	5.6
Component			
Teacher observations			
Amount, in millions of dollars	21.6	7	2.7
Percentage of total expenditures	87	82	48
VAM			
Amount, in millions of dollars	3.2	0.1	2.5
Percentage of total expenditures	13	1	44
Student surveys			
Amount, in millions of dollars	Not applicable	1.4	0.4
Percentage of total expenditures	Not applicable	17	8

SOURCE: Adapted from Exhibit B in Chambers, Brodziak de los Reyes, and O'Neil, 2013.

responsibilities, so no new expenditures were incurred. (The cost of reallocated time is examined below.) Each district spent substantial amounts of money on software infrastructure to create in-house solutions to support the classroom-observation component.

Regarding the VAM measures, SCS expenditures are substantially low in comparison to the other two Intensive Partnership districts because SCS did not have to develop new statistical models or acquire new data systems. SCS used value-added estimates provided by TVAAS, the Tennessee state system, which has been in place for several years.

The Expenditures to Implement the Teacher-Evaluation Systems Represented a Relatively Small Percentage of Total District Spending, but They Increased Substantially in the Three Districts During This Period

In school year 2009–2010, evaluation system expenditures in the three districts were, on average, between 0.1 and 0.2 percent of total district expenditures. By school year 2011–2012, these percentages had increased to between 0.4 and 0.5 percent of total district spending (see Table 2.2). In addition, specifically, in school year 2009–2010, expenditures on the teacher-evaluation system across the districts ranged from 19 percent of total Intensive Partnership expenditures in SCS to 34 percent of total expenditures in HCPS. By school year 2011–2012, these percentages had increased to more than 40 percent in HCPS and PPS and to 29 percent in SCS.

The Estimated Value of the Additional Time That School Leaders Spent on Teacher-Evaluation Activities in School Year 2011–2012 Compared with Spending in School Year 2010–2011 Is $43 per Pupil; the Comparable Amount for Teachers Is $159 per Pupil

With the implementation of the new teacher-evaluation system, school leaders reallocated their time to devote more effort to teacher evaluation: spending more time on observing classroom instruction and providing feedback to teachers. The expenditures presented in Table 2.2 do not capture the value of the time that school leaders spent observing and evaluating teachers, the value of the additional time that teachers devoted to the evaluations, or the time devoted to targeted PD. We drew on time-allocation data from the teacher and school-leader surveys to estimate the value of the additional time that school leaders and teachers devoted to the new evaluation system.[11]

---

[11] Specifically, for school leaders, our estimates include time spent attending training to conduct teacher evaluations, observing classroom instruction, preparing and providing feedback to teachers as part of their evaluations, other activities related to evaluating teachers, and time spent evaluating nonteaching staff. For teachers, our estimates include the time spent attending training to conduct observations as part of a teacher's evaluation, preparing for classroom observations as part of a teacher's evaluation, observing classroom instruction for the purposes of evaluating teachers, preparing and providing feedback to teachers, and participating in other activities related to formally observing or evaluating teachers

On average, principals increased the proportion of their time devoted to evaluation from 14 to 28 percent of their weekly working hours, and the main shift occurred between school years 2010–2011 and 2011–2012.

The increase in time that school leaders spent on teacher evaluation is equivalent to about $43 per pupil ($85 per pupil estimated for school year 2011–2012 minus $42 per pupil estimated for school year 2010–2011). Figure 2.10 shows increases for the individual Intensive Partnership sites associated with school-leader evaluation time equaling $8 per pupil for HCPS ($33 per pupil for 2011–2012 minus $25 per pupil for 2010–2011), $46 per pupil for SCS, and $74 per pupil for PPS (for more detail, refer to Appendix E, available online only [Stecher, Garet, Hamilton, et al., forthcoming]). This large variation reflects the different approaches each site took to implement the initiative, as well as differences in the sizes of the districts. For example, in PPS, school leaders bore the primary responsibility to conduct the teacher observations and provide feedback, whereas, in HCPS, teacher observations were primarily conducted by peer evaluators.

There was also a small increase between school years 2010–2011 and 2011–2012[12] in the amount of time that teachers spent on mentoring and evaluation activities (from less than 1 percent of their time to 5 percent of their time). This included more time spent on providing feedback to teachers as a formal or informal mentor and conducting their own evaluations. We estimated the value of this increase in mentoring and evaluation activities to be $159 per pupil on average across the three Intensive Partnership sites; specific estimates were $119

(e.g., record keeping). Teacher estimates do not include time spent in PD as a result of the evaluations.

[12] We used the time-allocation data from the teacher and school-leader surveys to estimate the difference in the percentage of time they devoted to evaluation-related activities in 2011 and 2012 and then applied the average compensation rate provided by the Intensive Partnership sites to this difference. Given that the teacher surveys were administered only in the spring of 2013, we have imputed the values for 2012 based on the 2013 survey. This decision seems justified because school-leader time-allocation patterns remained relatively unchanged between 2012 and 2013, and our conversations with central-office staff confirmed that no major changes in teacher-evaluation responsibilities occurred between 2012 and 2013.

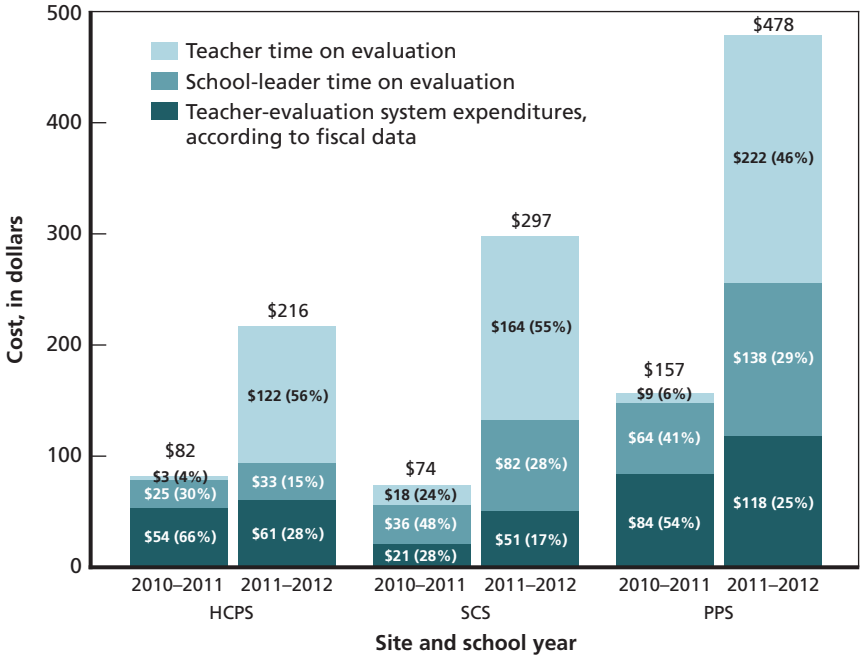
Table 2.2
Teacher-Evaluation System and Overall Intensive Partnership Initiative Expenditures, per Pupil and Percentages

	HCPS			SCS			PPS		
Expenditure	**2010**	**2011**	**2012**	**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Per-pupil expenditures, in dollars									
Teacher-evaluation system per-pupil expenditures (fiscal data)	13	54	61	8	21	51	50	87	118
Intensive Partnership initiative total expenditures per pupil	38	188	147	42	92	175	189	290	257
Overall district per-pupil expenditures	11,980	11,683	11,791	12,032	12,465	12,508	23,008	25,022	23,663
Overall teacher compensation per-pupil expenditures	5,275	5,144	5,192	4,806	4,979	4,996	8,680	9,440	8,928
As a percentage of total expenditures									
Teacher-evaluation expenditures as a percentage of total Intensive Partnership initiative expenditures	34.2	28.7	41.5	19.0	22.8	29.1	26.5	30.0	45.9
Teacher-evaluation expenditures as a percentage of total district expenditures	0.1	0.5	0.5	0.1	0.2	0.4	0.2	0.3	0.5
Teacher-evaluation expenditures as a percentage of overall teacher compensation	0.2	1.0	1.2	0.2	0.4	1.0	0.6	0.9	1.3

SOURCE: Exhibit C in Chambers, Brodziak de los Reyes, and O'Neil, 2013.

NOTE: The year indicated is the spring of the school year (e.g., *2010* = spring 2010, 2009–2010 school year).

Figure 2.10
Estimated Evaluation System Total Cost per Pupil for School Years 2010–2011 and 2011–2012



SOURCES: Chambers, Brodziak de los Reyes, and O'Neil, 2013; teacher and school-leader surveys, 2011, 2012, and 2013.
NOTE: Each bar with a date reflects a different school year.
RAND *RR1295-2.10*

for HCPS, $146 for SCS, and $213 for PPS (for more detail, refer to Appendix E, available online only [Stecher, Garet, Hamilton, et al., forthcoming]).

Figure 2.10 presents the total estimated costs of evaluation activities in each of the three Intensive Partnership sites, including direct expenditures (derived from Table 2.2 above) and the estimated value of the increased time that school leaders and teachers spent. The estimated total cost of the teacher-evaluation system in school year 2011–2012, including teacher and school-leader time, ranged from $216 per pupil to $478 per pupil.

Summary

A new teacher-evaluation process forms the foundation for the other components of the Intensive Partnership initiative. The evidence suggests that the sites were successful in implementing new measures of TE and incorporating them into an evaluation system that most teachers and school leaders initially supported. Although the percentage of teachers reporting that the system would benefit students in the long run has declined somewhat in the past year, the intended elements have been implemented, and the practices appear to be fairly stable in all the sites. In the next chapter, we look at evidence about the implementation of new staffing practices, many of which use the teacher-evaluation results as a key factor in determining a teacher's career trajectory.

CHAPTER THREE

Staffing

The designers of the Intensive Partnership initiative placed a major emphasis on using new information on TE to improve staffing decisions. The Intensive Partnership theory of action rests on the premise that improved information can be used to improve initial hiring, promote a more-equitable distribution of effective teachers across high- and low-need schools, and improve decisions about retention and dismissal. This chapter describes the reforms the sites adopted in hiring, incentives to work in high-need schools, and tenure and dismissal, focusing on the timing of implementation, as well as teacher and school-leader perspectives on these reforms.

Staffing Lever Implementation

Teacher hiring is one of three reforms we examine within the staffing lever. Teacher hiring is not directly informed by the TE measure (because there is no way to compute the measure on prospective candidates unless they have such information from a district where they worked previously), but the hiring, orientation, and initial training process can be informed by the dimensions of effectiveness that are incorporated into the evaluation system. Thus, the staffing lever includes training staff who review candidates (both in HR departments and in schools) to look for characteristics and skills that align with the new TE measures. This lever also includes practices to improve efficiency in seeking out candidates and making hiring decisions.

In addition, the lever includes policies and practices concerning how teachers are placed in schools, including measures to encourage equitable distribution of effective teachers across high- and low-need schools, and policies to retain effective teachers and dismiss those who are not. The TE measure is expected to play a direct role in both of these. These are the specific policies and practices included under the staffing lever:

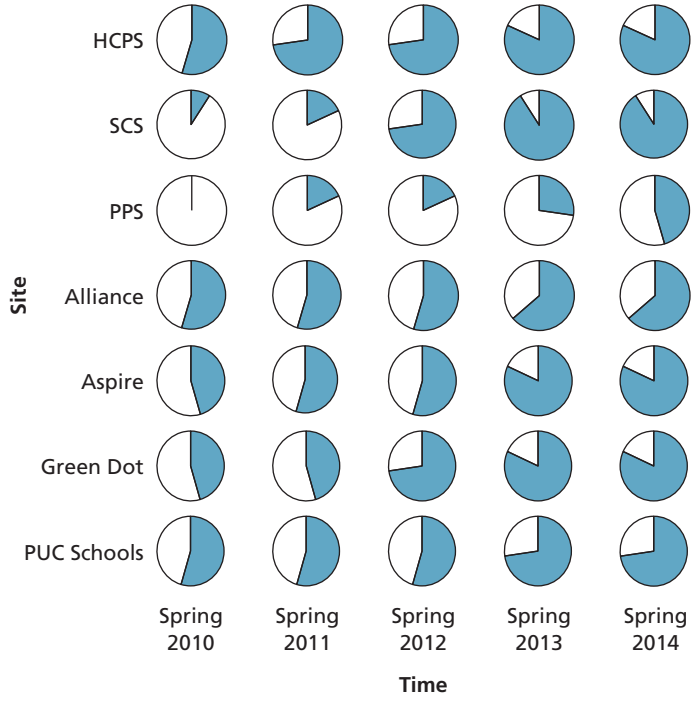
- early or expedited recruiting or hiring for high-need positions
- early hiring for all vacancies
- schools making the final hiring decision
- administrators trained to make good hiring decisions (e.g., in interviewing and team-building)
- new applicant screening model based on TE rubric
- incentives offered to work in high-need schools and classrooms
- transfers and furloughs not heavily influenced by seniority
- school leaders making the final decision about which teachers are placed in their schools
- tenure and retention linked to effectiveness ratings
- effectiveness rating used as a basis for dismissal
- schools making final decision about teacher retention and dismissal.

Hillsborough County Public Schools and the Charter Management Organizations Were Already Implementing Many of the Staffing Practices at the Beginning of the Reform

Figure 3.1 shows the status of the staffing lever over time across the Intensive Partnership sites.

Prior to the start of the initiative, several staffing practices were in place at all of the CMOs and in HCPS. In the CMOs, schools had final hiring and firing authority, transfers and furloughs were not heavily influenced by seniority, and none of the CMOs had tenure; teachers were at-will employees, rehired annually. Typically, teacher recruitment began in the CMOs in March. An attempt at starting earlier was unsuccessful because principals could not identify vacancies until teachers submitted their letters of intent to remain.

Figure 3.1
Proportion of the Staffing Lever Implemented, Spring 2010 to Spring 2014



RAND *RR1295-3.1*

In HCPS, schools had made the final hiring and retention or dismissal decisions for teachers prior to the Intensive Partnership initiative, and the previous evaluation system included performance ratings that the district used to dismiss low-performing teachers (although the district's approach to this changed under the EET initiative). HCPS also had long offered incentives to work in high-need schools.

The Three Districts Adopted Different Staffing Policies and Practices, Reflecting Different Local Conditions and Relationships with Teacher Organizations

Although all of the sites revised their approaches to recruitment, hiring, placement, and the other staffing functions, there was no common pattern to these changes. HCPS initially proposed but ultimately chose

not to collaborate with any outside organizations to support teacher recruiting for high-need schools. SCS established a relationship with TNTP (an independent teacher-training organization) prior to receiving the Intensive Partnership grant, and then used some of its Intensive Partnership funds to expand this relationship so that TNTP functioned to some degree as the district's HR office, taking responsibility for implementing most of the staffing levers. Policy changes at the state level, particularly those enacted as a result of Tennessee's RTT grant, facilitated SCS's staffing reforms. In PPS, a variety of factors constrained implementation of the staffing lever. Budget shortfalls and declining enrollment limited hiring of new teachers. The district contract with the local teachers' union also influenced staffing policies; the contract links teacher seniority to transfer and placement policies, limiting school leaders' ability to make decisions about which teachers work in their school. In the spring of 2012, PPS implemented some career-ladder roles, which are positions for teachers with extra responsibilities and extra pay, in certain high-need schools, thus serving as an incentive for teachers to teach in these schools. Even so, by 2014, PPS had adopted fewer of the staffing practices than the other sites, in large part because of these constraints.

Teacher and School-Leader Perspectives on the Staffing Lever

Sites enacted new policies to improve staffing, including new hiring procedures, new policies regarding teacher placement and retention, and new teacher-tenure and dismissal policies. These policies were aligned with or directly linked to the new teacher-evaluation systems. In the next sections, we describe school-leader and teacher reactions to the policies and their impact on the effectiveness of the teaching staff.

We begin with new hiring policies. Nearly all of the schools in the seven Intensive Partnership sites hired at least one new teacher in the 2013–2014 school year, according to principals, and most schools hired at least three. So, most principals had experience with teachers hired through their sites' reformed practices.

Perceptions of Teacher Hiring

Most School Leaders Were Satisfied with the Performance of New Teachers Regardless of the Types of Preparation Programs from Which They Came

Most school leaders indicated that their schools had hired teachers from traditional teacher-preparation programs. In 2014, large majorities of these school leaders reported that they were satisfied with the performance of teachers from such programs, particularly in HCPS (91 percent) and Green Dot (94 percent).

In all of the sites except PPS, majorities of school leaders (ranging from 55 percent in Aspire up to 72 percent in HCPS) also reported having hired teachers from alternative teacher-preparation programs. Most of these school leaders (on average, about 80 percent) were satisfied with the performance of these teachers as well.

School leaders in five of the sites—SCS and the four CMOs—reported that their site works "with external organizations to hire new, high-quality teachers (for example, TNTP or Teach for America)." In Alliance, Green Dot, and PUC Schools, 65 percent or more of school leaders reported that their school had benefited from such relationships, and the percentages of school leaders reporting benefits generally increased from 2011 to 2014, although three of the five sites had declines from 2013 to 2014.[1] Aspire did not begin recruiting with an external organization until school year 2013–2014, and a lower percentage of Aspire school leaders reported benefits from such an arrangement than leaders in the other sites did.

Consistently over time, school leaders in HCPS and in Alliance have been more likely to report satisfaction with the performance of new teachers (regardless of the source) than school leaders in the other sites have. In the three districts in 2014, leaders at schools with lower

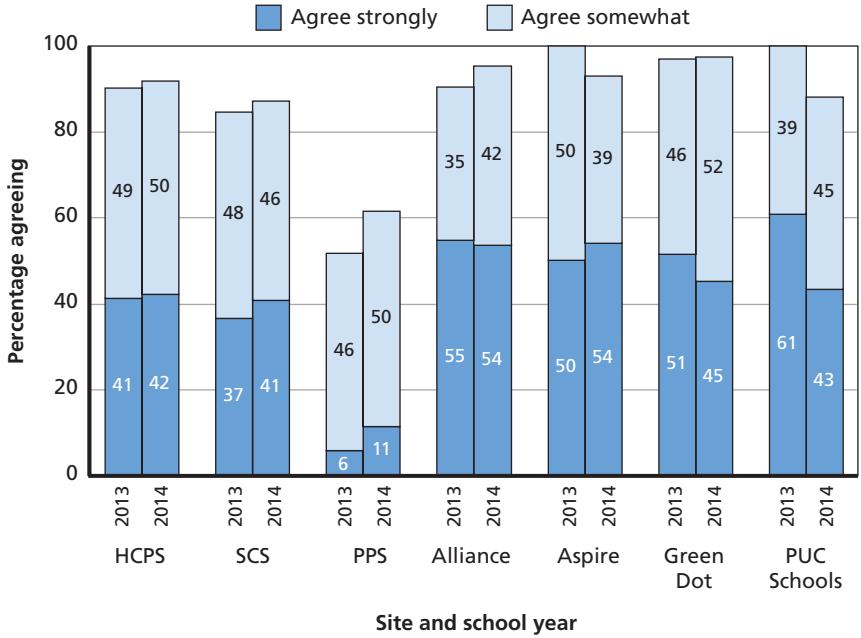
[1] Very few school leaders reported that their school had been hurt (from 0 percent to 12 percent in 2014, depending on the site), but "neither hurt nor benefited" was a fairly common response.

proportions of LIM students were more likely to indicate satisfaction (87 percent, on average) than those in higher-LIM schools (75 percent).[2]

Most School Leaders Were Satisfied with Teacher-Hiring Processes and Their Level of Control in Those Processes

With the exception of PPS, large majorities of school leaders in each site reported that the process for hiring teachers worked well (see Figure 3.2). Similarly, other than in PPS, where hiring of new teach-

Figure 3.2
Percentage of School Leaders Agreeing That "the Processes by Which Teachers Are Hired to My School Work Well," Springs of 2013 and 2014



NOTE: The survey data were collected in the spring of each year, and they reflect opinions during that school year, which began the previous fall. So *2013* refers to the 2012–2013 school year, for example. Each bar with a date reflects a different school year.

RAND *RR1295-3.2*

[2] Higher-LIM schools are the schools in the top half of the LIM distribution (within site); lower-LIM schools are the schools in the bottom half of the distribution.

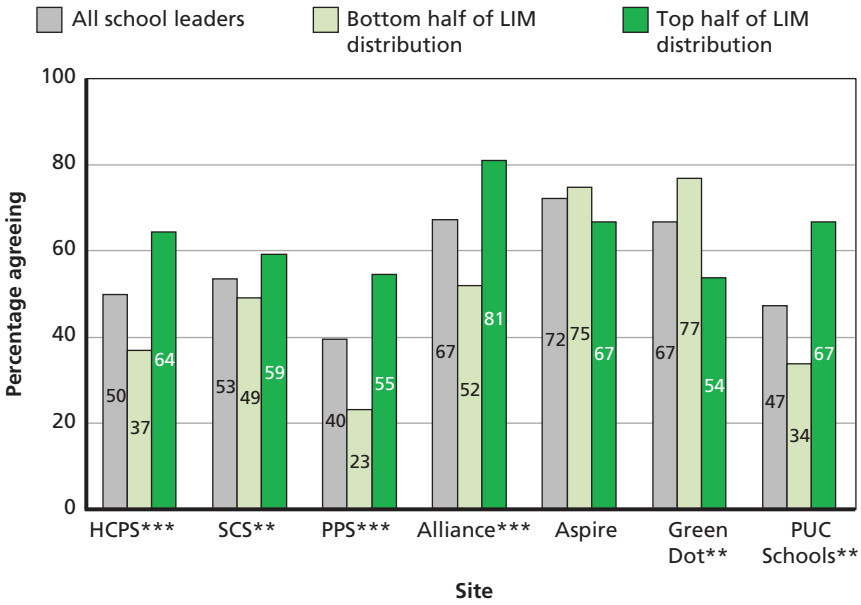
ers was very limited because of funding constraints, large majorities of school leaders agreed that they had "a sufficient amount of control over who comes to teach" in their schools. Almost all school leaders in the CMOs agreed, which is consistent with the CMOs' decentralized hiring procedures. Although levels of agreement with this sentiment were lower in the districts, they increased from 2012 to 2014. As one SCS principal described the situation, "Right now, it's a positive hire process. They say I need 15 teachers, and I can choose 15 from those who apply, and quite a few have applied. . . . I'm getting candidates who are new to the profession and those [who] wish to transfer." Another school leader in HCPS echoed satisfaction with the hiring process: "There is support. If you need a new teacher, need to hire a teacher, the district is there with background information [it] can provide, and [it has] the certification part already taken care of, what [the candidate qualifies] for."

Perceptions of Teacher Mobility and Placement

The Charter Management Organizations Struggle with the Loss of Good Teachers to Better Opportunities Elsewhere, While the Districts Struggle with Policies Governing the Movement of Teachers from School to School

Losing good teachers to better opportunities appears to be a growing problem for schools in the Intensive Partnership sites. In 2013, 20 percent of school leaders (on average across the sites) agreed strongly that, "more often than is good for my school, good teachers leave my staff because they perceive better opportunities elsewhere." In 2014, 25 percent strongly agreed; most of the increases were in the CMOs, where the percentage strongly agreeing rose from 16 percent in 2013 to 30 percent in 2014. In five sites—all but Aspire and Green Dot—leaders in schools with higher proportions of LIM students were significantly more likely to agree than leaders in schools with lower proportions (see Figure 3.3). One contextual factor relevant to the CMOs on this point is that, in 2013, LAUSD began hiring for the first time in several years, which affected the number and quality of teacher applicants to CMO schools and probably resulted in loss of teachers to LAUSD. Several of the CMOs reported difficulty in hiring effective teachers who also

Figure 3.3
Percentage of School Leaders Agreeing That, "More Often Than Is Good for My School, Good Teachers Leave My Staff Because They Perceive Better Opportunities Elsewhere," 2014



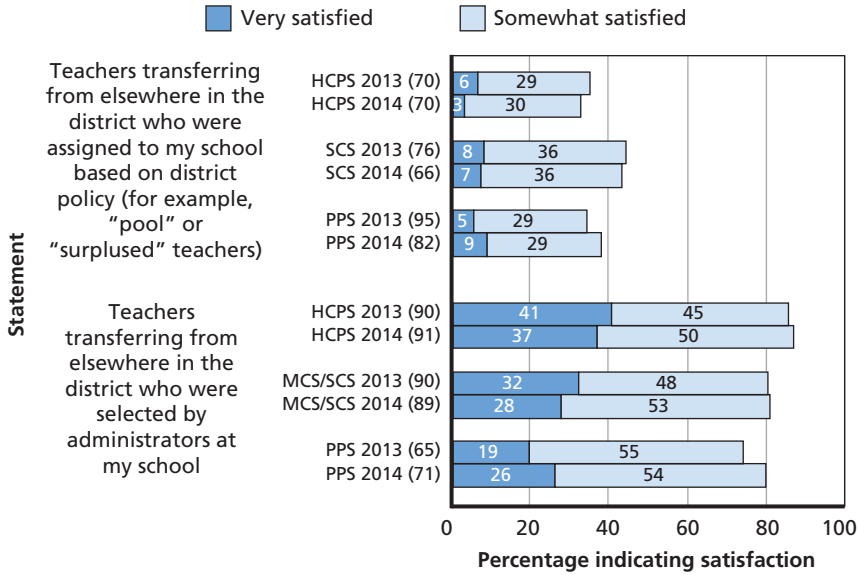
NOTE: ** = difference significant at $p < 0.01$. *** = difference significant at $p < 0.001$.
RAND *RR1295-3.3*

embraced the CMO culture of serving students in low-income, underperforming areas.

School leaders in the districts, meanwhile, continue to grapple with a different problem: good teachers being "bumped" to other schools because of seniority or other policies, and less-effective teachers transferring in. School leaders in PPS were the most likely to express concerns of this nature. In 2014, 61 percent of PPS school leaders agreed that, too often, good teachers were forced to leave the staff, and 85 percent agreed that "district procedures sometimes require my school to take on a teacher who is not a good fit for the school." (However, both of these percentages are substantially lower than they were in 2013, suggesting that principals' experience with transfers might be improving.) In the other two districts, only 20 to 40 percent of

school leaders agreed that good teachers were forced to leave, but 70 to 80 percent reported that they were sometimes required to take on teachers who were not a good fit. (On the latter, however, SCS showed a notable decrease from 2012 to 2014.[3]) Relatedly, school leaders were more likely to express satisfaction with transferring teachers they had selected than with teachers who were assigned to the school based on district policy (see Figure 3.4).

Figure 3.4
Percentage of School Leaders Reporting Satisfaction with the Performance of Teachers Who Transferred to Their Schools, Springs of 2013 and 2014



NOTE: The numbers in parentheses next to the site names and years are the percentages of school leaders who indicated applicability (i.e., teachers hired from that source). The percentages in the bars are among the school leaders who indicated applicability. The survey data were collected in the spring of each year, and they reflect opinions during that school year, which began the previous fall. So *2011* refers to the 2010–2011 school year, for example. Each bar with a date reflects a different school year.

RAND *RR1295-3.4*

[3] This decrease in SCS might be due to adoption of a mutual-consent policy, in which the district generally did not place teachers but instead the teacher and principal had to agree, or mutually consent, to fill a vacancy.

Perceptions of Teacher Tenure and Dismissal

In the Three Districts, the Meaning of Tenure Has Become Murkier in the Past Couple of Years; However, Most School Leaders Agreed That It Is Harder for Teachers to Earn Tenure Than It Used to Be

Official teacher tenure applies to only two of the three districts.[4] All three still have at least some form of tenure, though, with meanings that differ from those prior to the Intensive Partnership reforms.[5] There appears to be some ambiguity among school leaders in the three districts about the current (as of the spring of 2014) status and meaning of tenure. Small majorities of school leaders in HCPS and SCS indicated that their sites still had tenure but that the nature of tenure had changed, but sizable minorities (10 to 20 percent) indicated either that tenure had been abolished or, oppositely, that tenure had not recently changed. In PPS, meanwhile, nearly all school leaders indicated that tenure still existed, but they were about evenly split as to whether the nature of tenure had changed in recent years. Thus there appears to be some confusion (or at least disagreement) among school leaders in all three sites about whether teacher tenure has changed recently, and even (in HCPS and SCS) whether teacher tenure still exists.[6]

Among the school leaders who indicated that tenure is awarded (either with or without changes in recent years), most agreed at least somewhat that they "have a clear understanding of the current criteria used in my district to determine whether teachers receive tenure." There have been large increases in the past four years (from 2011 to 2014) in the percentages of school leaders agreeing with the statement, "Over the past two years, it has become more difficult for teachers to earn

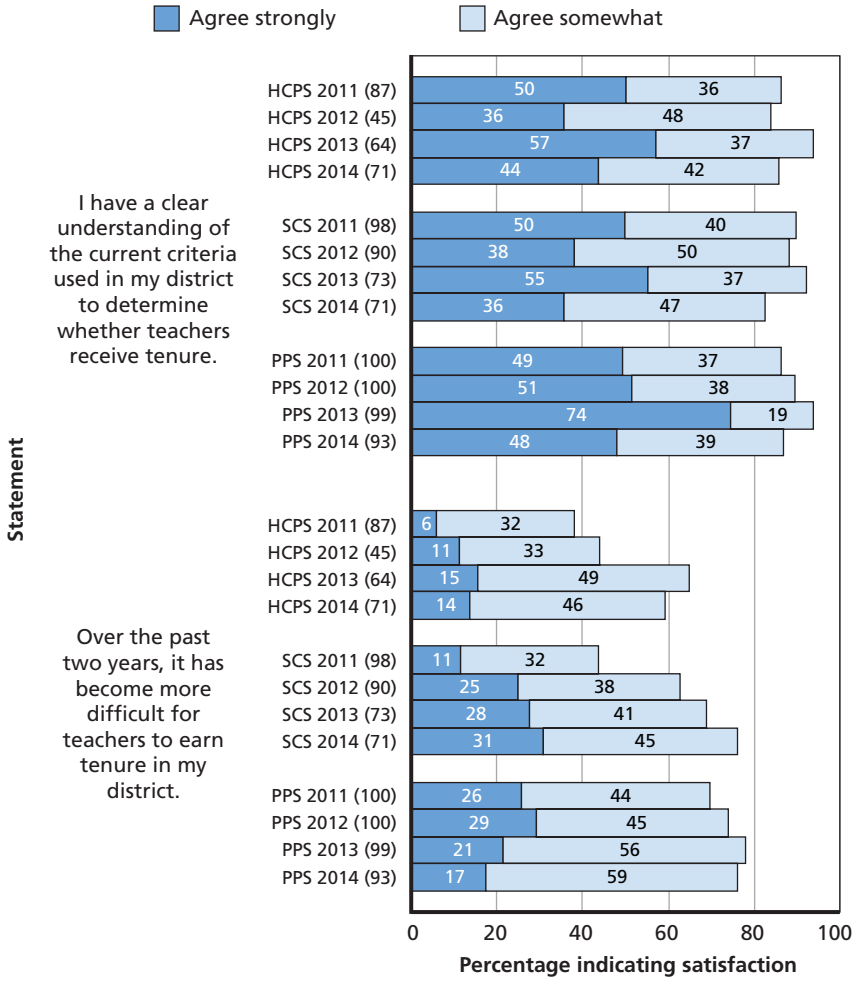
---

[4] Because of changes in Florida state law, the districts can no longer offer tenure. HCPS has retained a distinction, however, between probationary and nonprobationary teachers and is continuing its process of using evaluation results to dismiss low-performing teachers.

[5] In SCS, tenure law changed at the state level in 2012, and the new policy applies only to teachers hired after summer 2012. Districts may still grant tenure, but the requirements are different (award of tenure is now contingent on five years of good performance), and maintaining tenure is now contingent on continued good performance on the TEM.

[6] It might also be that some of the differences are essentially semantic, in how people think of and define *tenure* for themselves.

Figure 3.5
Percentage of School Leaders Agreeing with Statements About Site Tenure Policies, Springs of 2011–2014



NOTE: The numbers in parentheses next to the site names and years are the percentages of school leaders who indicated that their district grants tenure. The percentages in the bars are among the school leaders who indicated that their district grants tenure. Each bar with a date reflects a different school year.

RAND *RR1295-3.5*

tenure in my district" (see bottom half of Figure 3.5). The increases are particularly prominent in SCS.

School-leader responses to other items suggest that growing numbers have doubts about tenure's impact on the teacher workforce, and many think that tenure should be linked to evaluation results in some fashion. Tenure and dismissal policies are among the more-contentious aspects of the Intensive Partnership initiative, and many people hoped that policies that retained the idea of tenure but linked it to TE might be acceptable to teacher organizations, as well as district leaders. Surveys show that school-leader responses are changing over time and are not always consistent; this might reflect school-leader concerns about the validity of the effectiveness measure for high-stakes decisions and their sense that there are fewer ineffective teachers:

- Majorities of school leaders (60 to 80 percent, depending on the site and year) agreed that, "as currently implemented in my district, tenure protects bad or ineffective teachers." However, the percentages agreeing strongly declined in all three districts, most notably SCS (from 45 percent in 2011, down to 18 percent in 2014).
- The percentage of school leaders agreeing that "tenure should be linked to teachers' evaluation results" declined from between 87 and 100 percent in 2011 to between 73 and 100 percent in 2014, particularly in PPS (from 89 to 73 percent).
- In 2014, school leaders were more likely to agree (94 to 100 percent)—and especially to agree strongly (62 to 100 percent)—that "tenure should be granted only to teachers who have proven their ability to be effective with students" than that tenure should be linked to evaluation results.
- Fewer than half of school leaders (in any of the three districts in any year) agreed that "tenure should be abolished altogether."

Most Teachers in the Intensive Partnership Sites Were Not Very Worried About Being Dismissed

Perhaps surprisingly given the school-leader responses shown above, fear of being dismissed does not appear to be widespread among teach-

ers in the Intensive Partnership sites. In both 2011 and 2013,[7] few teachers in any of the seven sites agreed that "I am really worried about being dismissed." Yet, the percentage of teachers agreeing increased from 2011 to 2013 in all of the sites except HCPS and Alliance, so perhaps teachers were just becoming aware of new policies that were being implemented. Similarly, across all of the sites except PPS, in all years from 2011 to 2014, few school leaders agreed that "many teachers at my school are worried about being dismissed."[8] The largest increase was in PPS, and the spring of 2014 was the first year in which evaluation ratings had consequences for PPS teachers in terms of tenure, improvement planning, or dismissal. In interviews, teachers did express fear that low evaluation ratings would lead to job loss. One teacher told us, "Teachers are under serious stress and anxiety; they're afraid of losing their jobs because of this new evaluation system." Our interviews also suggest that at least some PPS teachers were unaware of the fact that no teacher would be dismissed as a result of a single low rating.

School Leaders in the Three Districts Indicated That Burdensome Procedures Created Obstacles to the Dismissal of Low-Performing Teachers

One possible explanation for teachers' lack of concern about dismissal is that remaining policies still make it difficult to dismiss a teacher. In the three districts, most school leaders indicated that there were continuing obstacles to the dismissal of low-performing teachers. In all four years (school years 2010–2011 to 2013–2014), most school leaders agreed that "[t]he termination/dismissal procedures in my district/CMO are so burdensome that most school administrators try to avoid using them." The percentages agreeing have, however, dropped in all three districts (for instance, in HCPS, from 72 percent in 2011 to 60 percent in 2014), suggesting that perhaps the procedures are becoming less burdensome. In the four CMOs, few school leaders agreed with the statement in any year, but, in contrast to the pattern observed in

[7] There are no results for 2014 because the shorter 2014 teacher survey did not include this question.

[8] In PPS, half of the school leaders agreed with the statement in 2013 and 2014, in contrast to agreement rates of 30 percent or lower in the two prior years.

the districts, the percentages agreeing increased from 2011 to 2014—for example, from 7 percent to 16 percent in Aspire. The low percentages in the CMOs probably reflect the fact that the process for dismissing teachers tends to be simpler than in the districts; principals hire and dismiss teachers, who serve on an at-will basis. The exception is Green Dot, which has a union and requires dismissals to adhere to a specific procedure outlined in the union contract.

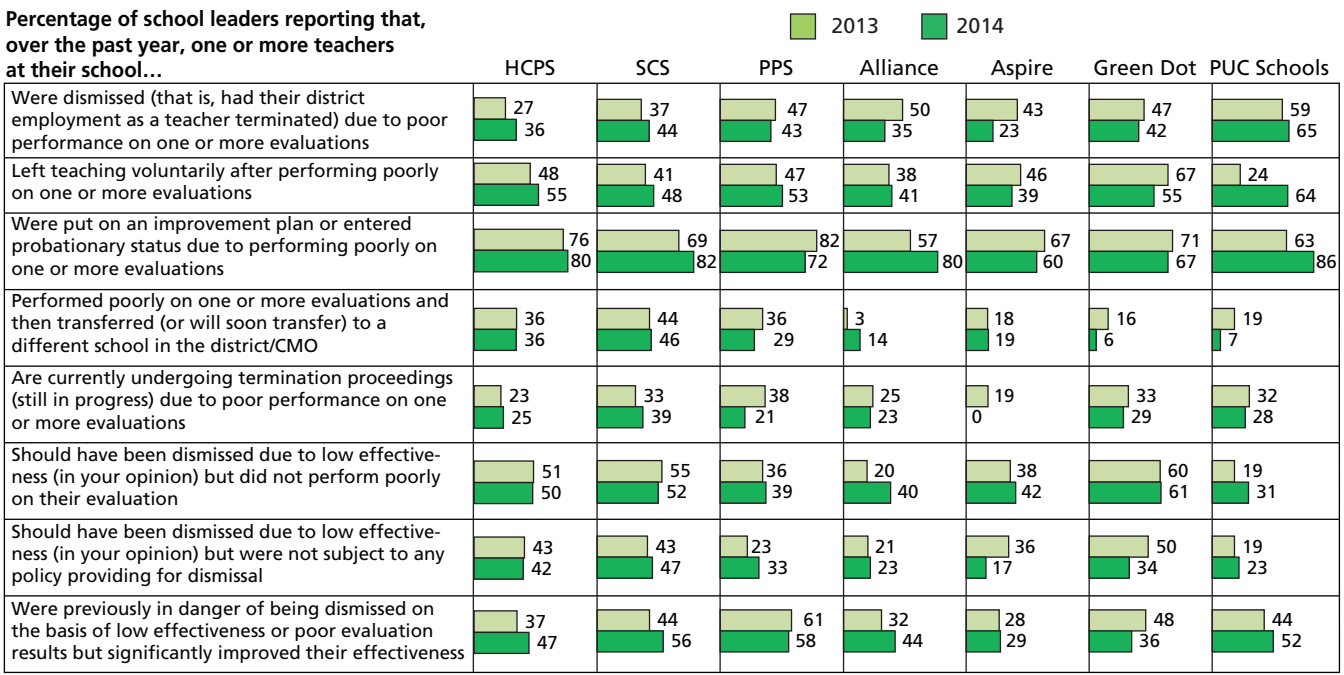
Presented with a list of 12 possible "barriers to the dismissal of poor-performing or incompetent teachers," leaders in the districts and in Green Dot were more likely to indicate that the barriers were present to a moderate or large extent than leaders in Alliance, Aspire, and PUC Schools were. Generally speaking, the factor most likely to be rated as a barrier to dismissal was "effort required for documentation." There have, however, been declines over time, in some of the sites, in the percentages of school leaders marking this factor (which is consistent with the results for the question about burdensome procedures, discussed above). Most notably, in SCS, the percentage of school leaders marking "large extent" on the "effort required" factor dropped steadily from 64 percent in 2011 to 49 percent in 2014. HCPS and Green Dot also showed steady declines.

According to School Leaders, Low-Performing Teachers Are More Likely to Be Put on Improvement Plans—and Then to Improve or Resign—Than to Be Dismissed

School leaders were also asked (in 2013 and 2014) what had happened to (or with) teachers considered low-performing or ineffective. Given a list of seven possible outcomes for such teachers—including dismissal or termination—respondents were asked to indicate how many teachers at their school experienced each outcome in the past year. (Results are shown in Figure 3.6.) Of the seven outcomes presented, the one marked by the most school leaders as applying to at least one teacher was "put on an improvement plan or entered probationary status due to performing poorly on one or more evaluations." Other outcomes marked somewhat commonly were

- "left teaching voluntarily after performing poorly on one or more evaluations"

Figure 3.6
Percentage of School Leaders Reporting That One or More Teachers Experienced Various Outcomes, Springs of 2013 and 2014



Percentage of school leaders reporting that, over the past year, one or more teachers at their school…	HCPS	SCS	PPS	Alliance	Aspire	Green Dot	PUC Schools
Were dismissed (that is, had their district employment as a teacher terminated) due to poor performance on one or more evaluations	27 / 36	37 / 44	47 / 43	50 / 35	43 / 23	47 / 42	59 / 65
Left teaching voluntarily after performing poorly on one or more evaluations	48 / 55	41 / 48	47 / 53	38 / 41	46 / 39	67 / 55	24 / 64
Were put on an improvement plan or entered probationary status due to performing poorly on one or more evaluations	76 / 80	69 / 82	82 / 72	57 / 80	67 / 60	71 / 67	63 / 86
Performed poorly on one or more evaluations and then transferred (or will soon transfer) to a different school in the district/CMO	36 / 36	44 / 46	36 / 29	3 / 14	18 / 19	16 / 6	19 / 7
Are currently undergoing termination proceedings (still in progress) due to poor performance on one or more evaluations	23 / 25	33 / 39	38 / 21	25 / 23	19 / 0	33 / 29	32 / 28
Should have been dismissed due to low effectiveness (in your opinion) but did not perform poorly on their evaluation	51 / 50	55 / 52	36 / 39	20 / 40	38 / 42	60 / 61	19 / 31
Should have been dismissed due to low effectiveness (in your opinion) but were not subject to any policy providing for dismissal	43 / 42	43 / 47	23 / 33	21 / 23	36 / 17	50 / 34	19 / 23
Were previously in danger of being dismissed on the basis of low effectiveness or poor evaluation results but significantly improved their effectiveness	37 / 47	44 / 56	61 / 58	32 / 44	28 / 29	48 / 36	44 / 52

NOTE: Each bar with a date reflects a different school year.

RAND *RR1295-3.6*

- "previously in danger of being dismissed on the basis of low effectiveness or poor evaluation results but significantly improved their effectiveness"
- "should have been dismissed due to low effectiveness (in your opinion) but did not perform poorly on their evaluation."

Thus, it appears that most schools in the Intensive Partnership sites generally have at least one teacher identified as needing improvement, and, at many of the schools, such teachers are then either leaving voluntarily or improving their effectiveness. However, it also appears that nearly half of school leaders (ranging from 31 percent in PUC Schools up to 61 percent in Green Dot) believe that the evaluation system is not flagging for improvement some teachers who ought to be dismissed for ineffectiveness.[9] Even so, on average, across the seven Intensive Partnership sites, about 40 percent of school leaders in both 2013 and 2014 reported that one or more teachers had, over the past year, been "dismissed (that is, had their district employment terminated) due to poor performance on one or more evaluations." In four of the sites (HCPS, PPS, Aspire, and Green Dot), leaders in higher-LIM schools were significantly more likely than their lower-LIM counterparts to report that at least one teacher had been dismissed because of poor performance.

School leaders in the districts were more likely than school leaders in the CMOs to say that at least one teacher performed poorly on one or more evaluations and then transferred to a different school in the district or CMO.[10] Moreover, in the three districts, leaders from higher-LIM schools were significantly more likely than their lower-LIM counterparts to report this.[11] On a similar question—"Currently, ineffective teachers are more likely to get moved to a different school within the

[9] According to Green Dot central-office staff, about 20 teachers should be on improvement plans but have overall evaluation scores just above the union-negotiated threshold for improvement plans.

[10] The three districts averaged 37 percent, while the four CMOs averaged 11 percent.

[11] The three districts averaged 43 percent of leaders in higher-LIM schools versus 31 percent in lower-LIM schools.

district than to have their employment terminated"—school leaders in the districts were again much more likely than CMO school leaders to agree. These differences between district and CMO responses might be due to the fact that CMO teachers are not directly transferred from one school to another but instead have to apply for open positions along with other candidates.

In HCPS and SCS, however, the percentages of leaders agreeing that ineffective teachers were more likely to be moved than terminated decreased steadily from 2011 to 2014. The decrease is particularly noticeable in SCS, which had a steady decline from 2011 to 2014; the percentage of school leaders in SCS "agreeing strongly" declined from 50 percent in 2011 to 14 percent in 2014. This decline might be attributable to an SCS policy starting in 2012 that provided principals with greater freedom to remove low-performing teachers from their schools; these teachers were not guaranteed placement in other schools, even though they had not been formally dismissed from the district. In 2013, a district administrator in SCS confirmed to us that more teachers were being terminated for a variety of reasons, including poor teaching performance.

Summary

Each of the sites implemented reforms in hiring, incentives to work in high-need schools, and changed tenure and dismissal policies, although they differed considerably in their approaches to these policies. School leaders were generally satisfied with teacher-hiring practices but were less pleased with transfer and dismissal practices. School leaders reported that low-performing teachers were unlikely to be dismissed directly, and most teachers were not very worried about being dismissed. Some of these concerns might also stem from a lack of availability of high-quality PD and other supports to help struggling teachers improve. The next chapter presents evidence about changes in PD, with a focus on PD that is linked to performance on the evaluation measures.

CHAPTER FOUR

Professional Development

Teacher support and PD play an important role in the Intensive Partnership theory of action. The designers of the initiative believe that the measures of effectiveness will reveal not only how strong teaching is but also where it is weak, and that this information can be used to customize efforts to improve teaching. This chapter describes the sites' efforts to implement customized PD and teachers' and school leaders' experiences with PD. It is important to recognize that sites' efforts to customize PD do not need to come at the expense of systemwide PD that is offered to all teachers or to groups of teachers based on criteria other than demonstrated needs, such as PD focused on Common Core implementation. Our analyses focus on PD tied to the evaluation system because that form of PD is central to the Intensive Partnership theory of action, but all sites continued to offer other forms of PD as well.

Professional-Development Lever Implementation

The PD lever includes using evaluation data to identify teachers' individual development needs and then offering PD, feedback, coaching, or mentoring targeted to teachers' needs and designed to help teachers improve on those components of the evaluation. This lever also includes supports for new teachers and systematic supervisor oversight of teachers' participation in PD, a key component if TE data are to be linked systematically to targeted PD opportunities throughout the site. Finally, this lever includes an electronic system for PD data collection,

in which sites would record which teachers accessed what resources, and which would enable empirical exploration of the degree to which the site's PD was effective. The specific practices included under this lever are as follows:

- Use evaluation data to identify teacher development needs.
- Offer PD designed to improve specific teaching skills measured in the evaluation.
- Link coaching and mentoring feedback to evaluation components.
- Provide induction, mentoring, coaching, or academies for new teachers.
- Have supervisors systematically oversee teachers' PD participation.
- Create an electronic system for PD data collection.

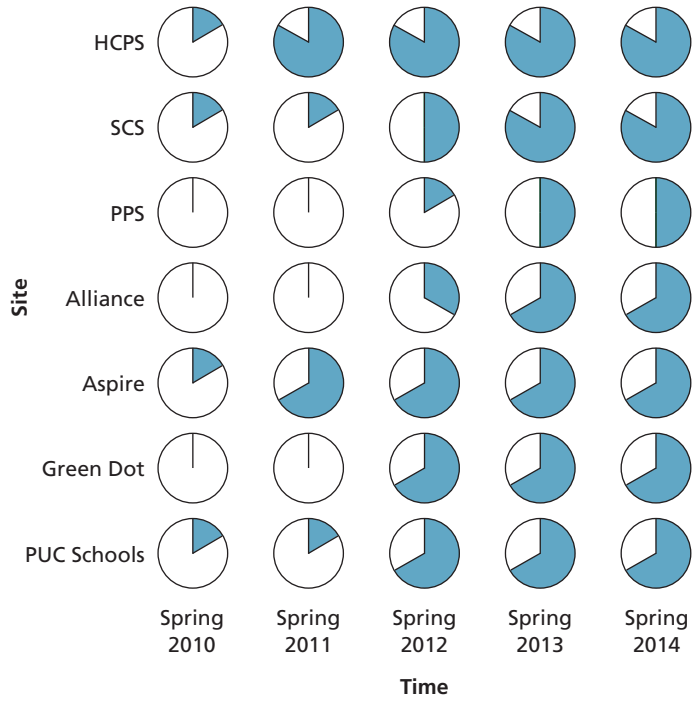
Figure 4.1 depicts the status of these practices over time across the Intensive Partnership sites.

What we see in the data that underlie Figure 4.1 is that sites did not begin to change policies related to PD until they had their measure of effectiveness in place (roughly 2012), and they still have not implemented the full slate of possible policies.

Once the Teaching-Effectiveness Measures Were in Place, Sites Began Linking Professional Development to Scores on These Measures

As noted in Chapter Two, in all sites, the observation rubrics became the definition of good instructional practice and provided a common language for teachers and observers to have detailed discussions about instruction. As such, the observations are the TE measure most commonly used to identify teachers' PD needs; the other measures are used less frequently. In HCPS, for example, a teacher with a final evaluation score in the "needs improvement" or "unsatisfactory" category must have an assistance plan built around suggestions from observers (although ultimately teachers decide how and whether to participate in the PD recommended as part of these plans). SCS began using

Figure 4.1
Proportion of the Professional-Development Lever Implemented, Spring 2010 to Spring 2014



RAND *RR1295-4.1*

observation data to identify teacher development needs in the same year as its effectiveness measure was adopted (school year 2011–2012). Teachers who received low scores (i.e., 1 or 2 out of 5) on rubric components were encouraged to seek PD designed to help them improve in those areas. To facilitate this, SCS developed a handbook called the Resource Book (Whitney et al., 2011), an online and printed listing of PD resources (e.g., videos, articles, lesson plans, in-person PD sessions) and a crosswalk so that teachers could easily identify which resources were relevant to which rubric components. In school year 2013–2014, the district's coaching model was redesigned to ensure that struggling teachers received some coaching support. Teachers who received scores of 1 or 2 on more than two rubric indicators in a given observation

were required to work with a coach for about six weeks, after which they would be observed again. SCS thus linked coaching and mentoring feedback to the observations and ensured some level of supervisor support for PD participation.

At the beginning of the initiative, PPS did not use TE data to inform PD options or recommendations for teachers. Although informal coaching and feedback had been part of the RISE process since 2010, use of the RISE data to identify development needs was not systematic or consistent across the district until the spring of 2012. Starting in June 2013, PPS began to link effectiveness ratings to PD opportunities more systematically, but, as of the spring of 2014, the district had not consistently achieved high levels of customization for most teachers. PPS's approach to PD emphasized allowing teachers to create their own PD plans rather than requiring teachers to participate in specific PD. The TE data are supposed to inform teachers' PD planning, but the district does not monitor this link in any formal way.

The CMOs do not offer a set of PD sessions focused on different topics. Instead, PD is delivered in weekly school site sessions and occasional CMO-wide PD days. CMO school leaders plan their weekly PD sessions based in part on indicators from the observation rubric on which most teachers need to improve. Teachers also select several rubric indicators as individual "growth goals," for improvement based on their previous year's observation results.

Although Four of the Sites Began Implementing the Professional-Development Reforms Within the First Two Years of the Initiative, the Reforms Were Not Operating Widely Until the Fourth Year

By the spring of 2014, all sites had reached substantial, though not complete, implementation. Two reforms—supervisor oversight of PD participation and an electronic system for PD data collection—were not implemented in most sites as of the spring of 2014. (HCPS and SCS are the exceptions; both implemented an electronic system for PD data collection in 2014.) Although we do not expect all sites to implement all the practices related to each lever, supervisor oversight of PD participation is important for ensuring effective and systematic implementation of PD, and collecting PD participation data (among

other information) in an electronic format is critical for understanding whether the PD the site is offering can be linked to improvements in TE.

Since the Initiative Began, Teachers and School Leaders Have Increased Their Time Allocated to Professional Development

Both school leaders and teachers spent more time on PD activities in school year 2012–2013 than in school year 2010–2011 across all seven Intensive Partnership sites. The school leaders increased the percentage of their weekly working hours spent on providing PD from 15 percent to 24 percent. In addition, they spent more time on interschool collaboration, attending external courses, and providing training to individual or groups of teachers and nonteaching staff. Teachers increased the percentage of time they spent in PD from 4 percent to 14 percent. The increase consisted of additional training and related activities that the district sponsored, as well as taking courses and engaging in informal, self-directed learning.

Customizing Professional Development to Address Individual Needs Has Proven to Be Logistically Challenging, and Sites Are Just Beginning to Implement Such Practices

Our central-office interviews suggested that customizing PD based on evaluation data has been difficult because of limited resources, including observer time, which have hindered observers' ability to provide customized feedback and PD suggestions to every observed teacher. Central-office leaders also noted a lack of high-quality PD opportunities to address specific needs identified through the evaluation system.

Despite these challenges, sites are making progress in this area. For example, HCPS's proposal outlined a prescriptive system that would link student achievement data and teachers' PD based on identified areas of weakness for teachers on assistance plans. HCPS has not yet integrated the PD tracking system with TE data because of delays in beginning this work. However, the new evaluation system provides data (classroom-observation data and VAM scores) that have helped the district to better target PD resources based on teacher need. These resources are also now associated with evaluation rubric components,

so teachers can more easily select PD resources that are intended to directly address their practice based on their individual needs.

SCS and PPS have also struggled to customize PD on a broad scale. Limited resources for coaching staff and limited observer time have led both districts to focus most of their attention and resources on the lowest-performing teachers. In PPS, each teacher who performs at the lowest level is required to develop a plan for professional growth and support (called an intensive support plan) and to have the school's principal approve that plan. The lowest-performing SCS teachers are required to develop similar plans, and teachers who receive low observation ratings throughout the year receive coaching support. At the school level, PPS principals often try to maximize their impact by offering small-group PD sessions that focus on areas of the rubric with which several teachers in the building are struggling.

Initially, the CMOs began developing online resources to customize PD. In 2012, two CMOs began creating online PD resources available to all teachers and linked to specific indicators on the observation rubric. Aspire began creating video clips of the teaching practices of effective teachers directly linked to specific indicators on the observation rubric, and PUC Schools began creating instructional guides also directly linked to rubric indicators. These efforts were shared with the other CMOs, which continue to add to the video collection and to create their own versions of the instructional guides. When teachers receive the results of their observations, these online resources are identified for the teacher in conjunction with each rubric indicator. However, no data are available to indicate the extent to which school leaders recommend these resources to improve teacher instructional practice or teachers make use of them. Teacher interviews suggest that they are not often directed to these resources, nor are they technically easy to access, but they are deemed to be worthwhile once they are accessed, especially for new teachers. Currently, most of the individualized PD for first- and second-year teachers in the CMOs occurs through one-to-one coaching. All of the CMOs have increased their coaching staff and have created peer-coaching positions at the school sites.

Teacher and School-Leader Perspectives on the Professional-Development Lever

This section examines teacher and school-leader reports regarding the availability of PD and other supports to improve teaching. It also documents survey and interview findings on the perceived utility of PD.

Perceptions of the Availability of Supports Tied to Their Evaluation Results

Most Teachers Reported That Supports Aligned to Their Evaluation Results Are Available, Appropriate, and Helpful but Not Well Cataloged

Except in PPS, large majorities of school leaders and at least half of teachers in all sites reported that "needs identified as part of teachers' formal evaluations" had a moderate or large influence on what PD they participated in during the 2013–2014 school year.[1]

Similarly, large majorities of school leaders and majorities of teachers in all sites reported that teachers' PD experiences in school year 2013–2014 were "designed to address needs revealed by analysis of student data" and "aligned with or focused on specific elements of the district teacher observation rubric." In 2014, approximately 50 percent of teachers in HCPS and PPS and approximately 60 to 70 percent of teachers in the other sites agreed that "supports aligned to evaluation results are appropriate and helpful." Agreement was typically higher for teachers in the CMOs than for teachers in the districts.

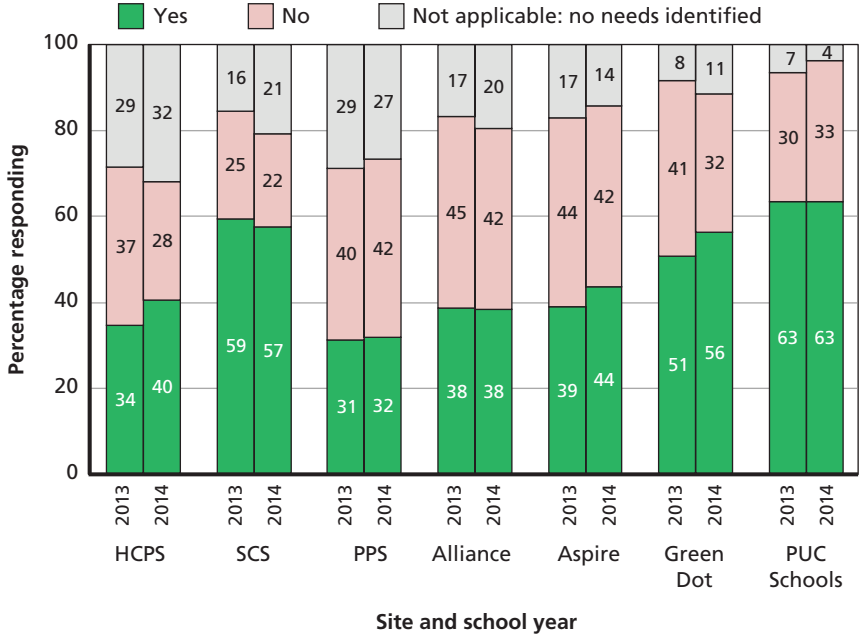
Sites have begun to provide teachers with support (e.g., coaching, PD) to address the needs identified in their annual performance evaluations. Figure 4.2 presents the percentage of teachers who indicated

---

[1] In PPS, which has adopted a decentralized approach to supporting teachers, many teachers receive reports in which their TE data are accompanied by fairly general guidance rather than guidance that is targeted toward their identified needs. One PPS teacher described this situation:

> I didn't expect [the action steps at the end of the educator-effectiveness report] to speak to me personally, and they were very general and good feedback overall. I prefer the feedback I get from my administrator, who provides better guidance for next steps. [The educator-effectiveness report action steps] tend to be pretty broad and general, not specific. There's also no follow-up.

Figure 4.2
Percentage of Teachers Responding to the Question, "Has Support (Coaching, Professional Development, etc.) Been Made Available to You to Address the Needs Identified by Your Evaluation Results?" School Years 2012–2013 and 2013–2014



NOTE: Each bar with a date reflects a different school year. Numbers might not sum to 100 because of rounding.

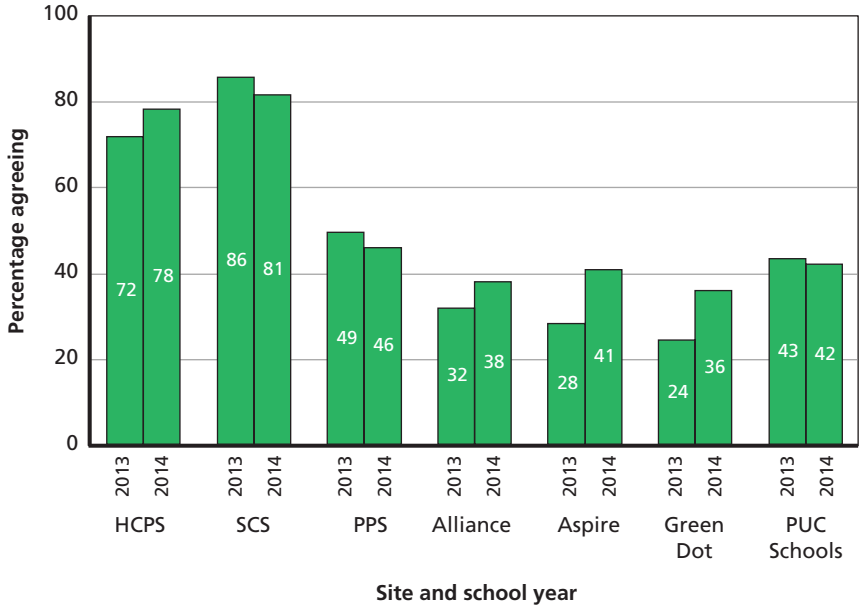
RAND *RR1295-4.2*

that support has "been made available to you to address the needs identified by your evaluation results" in 2013 and 2014. Majorities of teachers in SCS, Green Dot, and PUC Schools reported that such support had been made available.

For teachers who indicated that support was made available to address needs identified by their evaluation results, a majority of teachers in all sites except PPS reported that this support helped them address the identified needs. That is, teachers generally found the support linked to their evaluation results useful for improving their instructional practice.

Nonetheless, with the exception of teachers in HCPS and SCS, most teachers reported that they do not have "easy access to a catalog of professional development opportunities aligned with [the] district/CMO teacher observation rubric" (see Figure 4.3). However, in Alliance, Aspire, and Green Dot, more teachers agreed that they had access to such a PD catalog in 2014 than in 2013. In 2013 and 2014, a majority of school leaders in all sites reported that teachers had access to such a catalog, suggesting a difference in perceptions between the two respondent groups.

Figure 4.3
Percentage of Teachers Agreeing That "I Have Had Easy Access to a Catalog of Professional Development Opportunities Aligned with My District/CMO Teacher Observation Rubric," School Years 2012–2013 and 2013–2014



NOTE: Each bar with a date reflects a different school year.

RAND *RR1295-4.3*

Perceptions of Coaching and Collaboration

The Percentages of Teachers Who Reported Receiving Coaching Varied Across Sites

Coaching is one approach that sites have used to provide teachers with targeted PD. Across all sites, many teachers reported receiving some form of coaching, but teachers in the CMOs were more likely than teachers in the districts to report having received coaching or mentoring (either one on one or as part of a group). For example, although 45 percent of teachers in PPS indicated that they had received coaching in school year 2013–2014, 87 percent of teachers in PUC Schools reported that they had received coaching in the same year.

Novice teachers, teachers in core subject areas, and teachers in tested grades and subjects were more likely than their counterparts to indicate that they had received coaching. For example, in SCS, coaching is an important part of customized PD for struggling teachers and novice teachers. Coaching plays a less important role in customized PD in PPS. In the CMOs, most coaching staff focused their time on first- and second-year teachers, and school leaders also provided a great deal of coaching. In interviews, coaching and collaborating with other teachers was the most frequently cited preference for PD in the CMOs. An Alliance teacher stated, "When I meet with my math coach is the best professional development. It is fantastic."

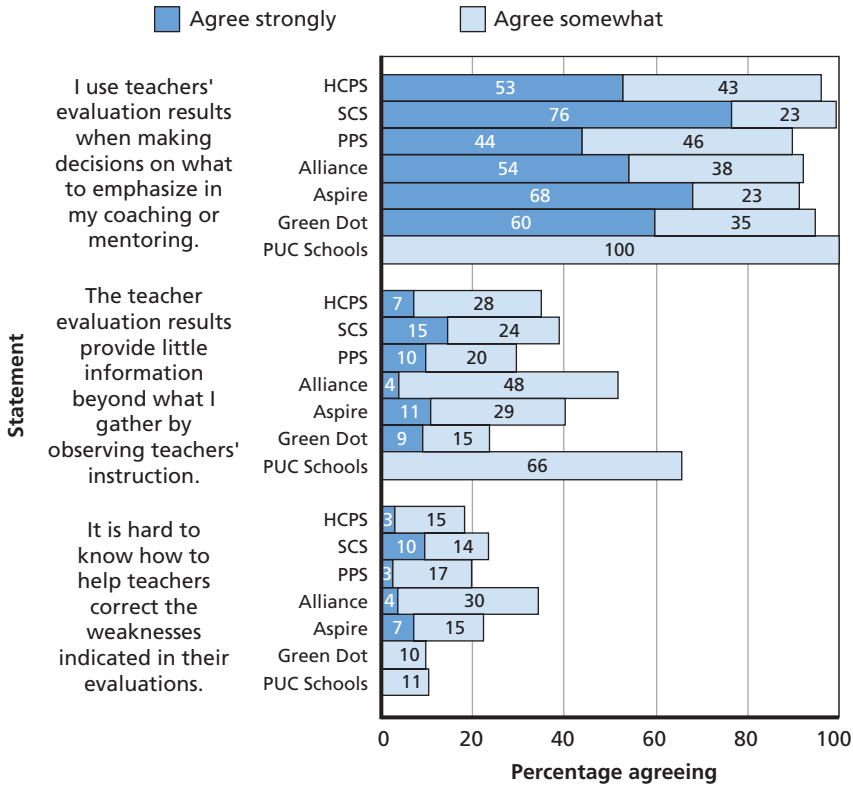
Among teachers who received coaching, a majority in all sites reported that the coaching was moderately or very useful and that they received a sufficient amount of it. Novice teachers were typically more likely than experienced teachers to report that the coaching they received was useful, although they did not necessarily report that they had received a sufficient amount of coaching.

School Leaders Reported Using Teachers' Evaluation Results to Provide Individualized Coaching or Mentoring as a Form of Professional Development

School leaders indicated that they themselves often provide coaching or mentoring to their teaching staff, and most of them reported that they use teachers' evaluation results when making decisions on what to emphasize in their coaching and mentoring. Additionally, most

school leaders indicated that the evaluation results provide information beyond what they can gather by simply observing a teacher's instruction and that they know how to help teachers correct weaknesses indicated by their evaluations (see Figure 4.4). On the other hand, in interviews, many school leaders in both the districts and the CMOs indicated that they tend to rely on observation results when deciding what to emphasize when they coach teachers. Some school leaders say that the student survey scores inform their decisions as well, but the

Figure 4.4
Percentage of School Leaders Agreeing with Statements About Coaching and Mentoring the Whole Staff, School Year 2013–2014



RAND *RR1295-4.4*

majority of those interviewed relied almost entirely on observations, rather than the evaluation as a whole.

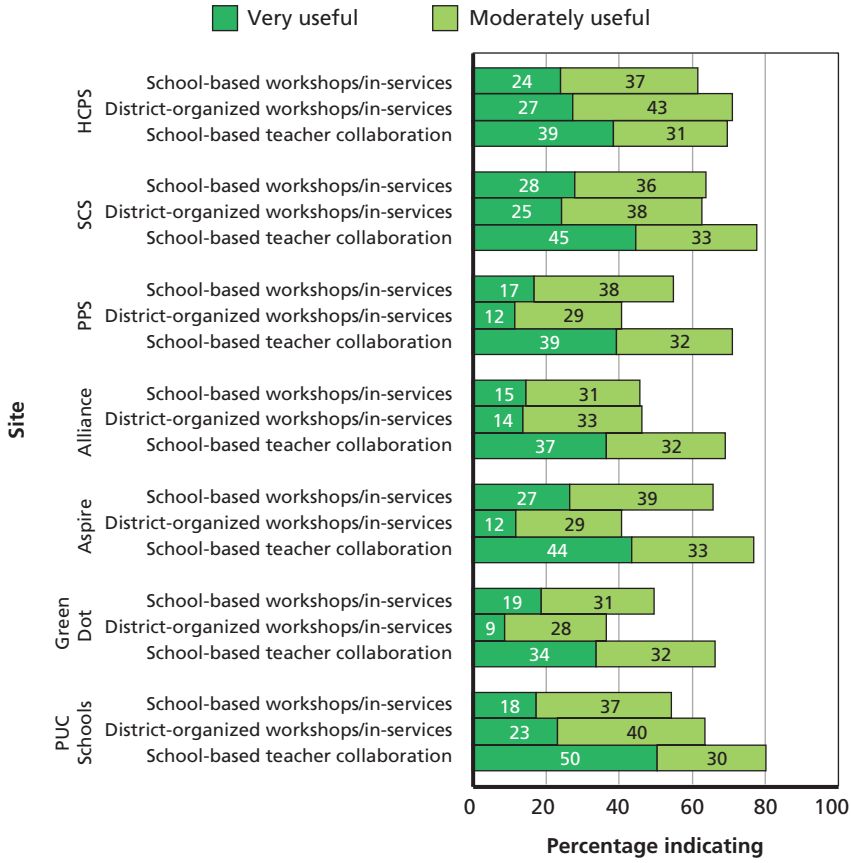
Most Teachers Indicated That School-Based Teacher Collaboration Was More Useful Than School- or Site-Based Workshops or In-Services

In addition to coaching, teachers' PD included school- and district-based workshops, as well as school-based teacher collaboration (among other forms of PD). Across these three types of PD, teachers reported that collaboration was the most useful (see Figure 4.5). This perception might stem in part from the ways in which collaboration can be tailored to a teacher's specific circumstances. A PUC Schools teacher pointed out that, "working with colleagues, we both teach biology; when we have time to work with each other or share resources, it's more useful than a lot of the professional-development time we spend waiting for someone to finish talking." Similarly, school leaders indicated that school-based workshops are more useful for teachers than workshops that the district or CMO organizes. On the other hand, school leaders did not generally report that school-based teacher collaboration was necessarily more useful than school-based workshops; school leaders rated all school-based forms of PD high in terms of its utility for teachers.

Summary

It has taken sites longer to implement PD reforms than either evaluation or hiring reforms. In some ways, this is to be expected: PD reforms required the evaluation measures to be in place long enough to assess how well they were working, to develop strategies for linking that information to PD recommendations, and to identify or develop PD opportunities that would address teachers' needs. Sites are finding it difficult to link individual PD to needs identified by the teacher-evaluation process, a challenge that stems from incomplete information in the evaluation measures, from a lack of PD options that are tailored to areas of weakness that those measures identify, and from insufficient time for principals to offer tailored feedback to each

Figure 4.5
Percentage of Teachers Indicating the Usefulness of School-Based Workshops and In-Services, District- or Charter Management Organization–Based Workshops and In-Services, and School-Based Teacher Collaboration, 2014



RAND *RR1295-4.5*

teacher. Nevertheless, school leaders reported that the evaluation information is helpful in focusing mentoring and support, and majorities of teachers in some sites reported having access to coaching or other PD that addresses their needs. The sites are starting to use more mentors or coaches, which could overcome some of the challenges associated with using formal PD systems to try to meet individual teacher needs.

The next chapter provides evidence about changing teacher compensation and career ladders designed to formalize and professionalize this mentor role.

CHAPTER FIVE

Compensation and Career Ladders

Compensation and career-ladder policies that offer monetary rewards for effective performance and create new teaching positions with added responsibilities (and added pay) are a key component of the Intensive Partnership theory of action. The goals of such policies are to retain effective teachers at higher rates and to give teachers opportunities to use their skills to support improvement in other ways. In the Intensive Partnership sites, these policies tend to overlap in that specialized teaching roles are often associated with additional compensation. For this reason, we cluster them together under the compensation and career-ladder lever.

Compensation and Career-Ladder Lever Implementation

The compensation portion of this lever includes monetary rewards for effective teachers (e.g., a bonus or a permanent salary increase for achieving a certain effectiveness score), as well as financial incentives for teaching in a high-need position, such as a hard-to-staff subject area or grade level.[1] The lever also reflects whether the site has stopped

[1] In this report, *bonus* indicates a temporary financial reward, awarded each year, and it might or might not be repeated in subsequent years. A teacher can receive such bonuses more than once.

We classify incentives that are offered to attract teachers to work in any position in a high-need school as part of the staffing lever, which is discussed in Chapter Three. We acknowledge that some sites do not make a clear distinction between an incentive for working in a high-need school and an incentive for filling a specific high-need position, so there could be

exclusively using a step-based salary schedule, in which raises are given based on years of experience and advanced degrees, and bases some of a teacher's salary on effective performance. The career-ladder portion of the lever includes creating specialized roles for teachers that offer rewards for taking extra responsibilities and demonstrating greater leadership.[2] This lever relies heavily on the TE measure, which is often used as the eligibility criterion for performance-based incentives, either year-by-year bonuses or permanent salary increases, or teaching positions with extra responsibilities.[3] These are the specific policies and practices included under the compensation and career-ladder lever:

- year-by-year bonuses or stipends or permanent salary increases awarded based on a teacher's individual effectiveness measure
- traditional step-based salary schedule not used exclusively
- year-by-year bonuses or permanent salary increases given for high-need positions
- financial incentives given for desired teacher behavior (e.g., low absenteeism)
- positions created for effective teachers with different responsibilities.

Figure 5.1 shows the status of the compensation and career-ladder lever over time across the Intensive Partnership sites.

---

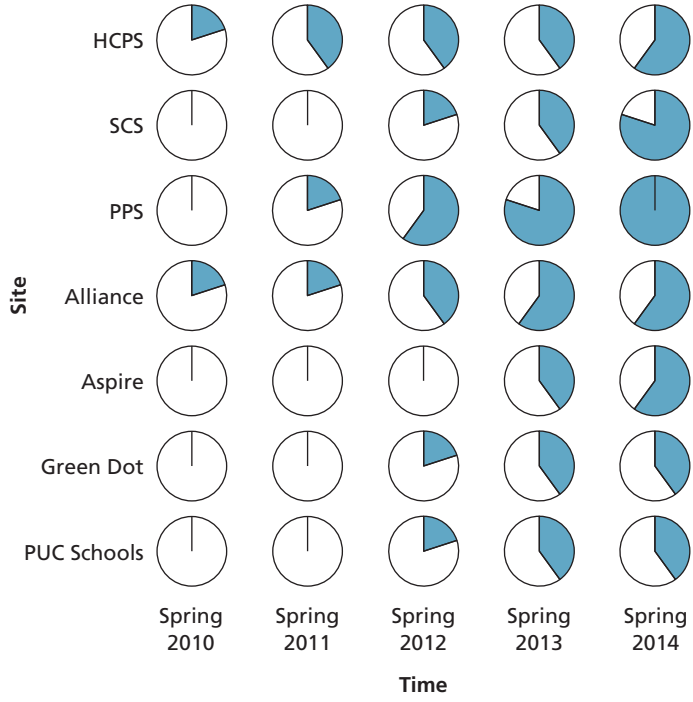
a small degree of overlap between the staffing lever and the compensation and career-ladder lever. Teachers' reactions to incentives of either type are discussed in this chapter.

[2] The survey offered the following definition:

> In a career ladder, teachers may be promoted and are given additional pay to take on new or different responsibilities, such as mentoring other teachers, typically without having to give up teaching. The positions on a career ladder may vary, but the higher-level positions typically have titles like "teacher leader" or "master teacher."

[3] We use the term *bonus* or *stipend* to refer to a financial reward that is given for a single year and must be earned again in subsequent years. *Salary increase* refers to a permanent increment to a teacher's base salary.

Figure 5.1
Proportion of Compensation and Career-Ladder Lever Implemented, Spring 2010 to Spring 2014



RAND *RR1295-5.1*

Sites Have Adopted New Compensation Policies (e.g., Bonuses for Effective Teaching) as Well as Specialized Teaching Roles (Career-Ladder Positions)

All of the sites offer some type of bonus or salary increase for demonstrating effective teaching, although the eligibility criteria and incentive structure differ across the sites. For example, in HCPS, all nonprobationary school-based personnel who are evaluated under the new evaluation system (which was implemented in school year 2010–2011) are now automatically considered for performance-based incentives. Similarly, SCS began offering effectiveness-based bonuses in 2012. PPS developed a merit-based salary schedule that went into effect for all teachers hired after July 2010. However, PPS does not simply award

bonuses to individual teachers for high effectiveness scores; instead, PPS awards group-based bonuses at the teacher team and school levels. In the CMOs, teachers began receiving bonuses in the fall of 2013 based on their effectiveness scores from school years 2011–2012 and 2012–2013. Aspire eliminated bonuses for the 2013–2014 school year, switching instead to a salary schedule based on effectiveness scores and years of teaching experience.

Similarly, all the sites have created some specialized teacher roles that entail additional responsibilities or greater leadership coupled with additional pay. In this report, we use the term *career ladder* to refer to these positions, but it is important to note that some educators use the term more expansively to refer to a progression of specialized roles with additional responsibilities and compensation that is available to the most-effective teachers as an alternative to going into traditional administrative roles (e.g., school principal). This broader notion of a coherent series of positions for career advancement within teaching is sometimes referred to as a *career pathway*. In fact, in the Bill & Melinda Gates Foundation request for proposal and in the Intensive Partnership site applications, there were many references to this broader notion of career pathways for teachers that would provide a sequence of opportunities for advancement while also allowing teachers to continue to spend time in the classroom, providing instruction. The Intensive Partnership sites planned to link these career pathways to a range of compensation policies (e.g., increased base salary, opportunities for bonuses) that would provide an incentive for teachers to apply for the positions.

As of the spring of 2014, most sites had developed career ladders in the form of specialized teaching positions that received annual bonuses or stipends, but they had not yet created more-expansive career pathways. For example, HCPS launched peer-observer and mentor positions in the fall of 2011 and recently began piloting a new career role—teacher leader—at 15 high-need schools. Teacher leaders allocate half of each day to carrying out regular classroom instructional duties and the other half to serving as instructional coaches for other teachers at their schools. Similarly, by 2013, all of the CMOs established at least a few positions offering effective teachers more responsibili-

ties and additional stipends. In 2012, SCS implemented two positions in which teachers could assume additional responsibilities and receive stipends. PPS also developed several differentiated teaching positions, which come with increased compensation.

Compensation Policies and Career-Ladder Positions Were Generally Adopted Later Than Other Intensive Partnership Reforms

Sites generally spent a few years planning for and then creating new specialized career-ladder positions and compensation approaches, in part because the TE measures needed to be in place to serve as a basis for these policies. Moreover, in some districts, these reforms required fundamental changes in negotiated teacher contracts, so more time was needed to develop acceptable policies. PPS was an exception to this pattern. The district was able to develop several specialized teaching positions and ratify significant compensation reform in 2010, and the district began to implement these policies in 2011. PPS was able to implement quickly because new compensation and career-ladder programs were part of the district's plan at the beginning of the Intensive Partnership initiative and were ratified in 2010 as part of the new contract with the teachers' union. Implementation of these policies was gradual and phased, and all the planned levers were fully implemented by the spring of 2014. HCPS also began implementing some new compensation policies in the first two years of the initiative, but SCS and the CMOs did not begin modifying their compensation policies until the third or fourth year of the initiative. The CMOs postponed development of career-ladder positions until TE data began to be available at the end of the 2011–2012 school year. Alliance had a system in place at the beginning of the initiative for compensating teachers based on high student attendance, but, once the TE measures were operational, bonuses based on TE ratings replaced attendance bonuses.

SCS awarded individual performance-based bonuses for teachers with high effectiveness scores starting in 2012, but, as of the spring of 2014, it had not yet implemented any career-ladder positions. Lack of central-office staff and high turnover among the staff working on this aspect of the reform have been challenges for SCS and contributed to the slow implementation of this lever.

As of School Year 2013–2014, More of the Compensation Policies Were Implemented in the Districts Than in the Charter Management Organizations

Four of the five policies included in this lever relate to compensation, and the three districts implemented more of these policies than the CMOs did. The districts were able to use outside funding to pay for some of the incentive payments, including TIF and RTT grants. PPS is the only site to have implemented all the compensation practices included in this lever; in particular, it is the only site that implemented incentive policies designed to shape other teacher behavior outside of instruction—in this case, to reduce teacher absenteeism. In schools eligible for bonuses based on high student achievement growth, individual bonus amounts are based on the number of days a staff member has worked at the school during the year. Although this creates an incentive to avoid absences, interviews with central-office staff and teachers suggest that the policy has not been widely publicized and is not well known in the district.

In HCPS, all nonprobationary, school-based personnel who are evaluated under the new evaluation system (implemented in school year 2010–2011) will automatically be considered for performance-based bonuses. The district awards bonuses beginning with the highest evaluation score and continues to award bonuses until all funds are exhausted. The bonus is equal to 5 percent of a teacher's salary, prorated on the basis of time worked each pay period, and is paid during the subsequent school year.

Aspire was the only CMO to eliminate the traditional step-and-column–based salary structure, whereas all three districts had eliminated such a salary structure by the spring of 2014. There was a downturn in state funding for education in California beginning in school year 2009–2010, and the CMOs feared that they would not be able to sustain a pay-for-performance salary structure with reduced per-pupil funding.

Many of the sites also offer bonuses or salary increments for high-need positions, such as special education, middle school mathematics, or science teachers. For example, PPS offers bonuses, in the form of higher step placement on the salary schedule, for newly hired teach-

ers in high-need positions; this policy has been in place since 2013. Administrators in SCS reported that they consider all positions in high-need schools to be high-need positions, and they offer bonuses for serving in these positions. (We discuss incentives to work in high-need schools in the discussion of the staffing lever in Chapter Three.) HCPS has long offered incentives to work in high-need schools (as we note in Chapter Three). Teachers at district schools with at least 90 percent of students qualifying for free or reduced-price lunch receive one-time bonuses and can qualify for salary differentials depending on the achievement gains among the lowest-performing students. Because all of the CMOs serve primarily LIM populations, teachers who choose to work at these CMOs know that they will be working with high-need students. Only Aspire offers an incentive for teaching at the lowest-performing schools, and that incentive is only for newly hired teachers who were Aspire teacher residents. Alliance encourages principals to offers stipends or bonuses for high-need positions, but few principals have sufficient resources to do this.

Teacher and School-Leader Perspectives on the Compensation and Career-Ladder Lever

In this section, we draw on the surveys and interviews to describe teacher and school-leader opinions about the ways in which their sites award compensation and whether they offer career-ladder positions (with associated financial rewards).

Perceptions of Compensation Policies

Most District Teachers Thought That Base Pay Should Be Based on Seniority, and a Majority of All Teachers Thought That Teachers Should Receive Additional Compensation for Demonstrating Outstanding Teaching Skills and for Working in Low-Performing Schools

As discussed above, all of the Intensive Partnership sites have tied some compensation to teacher-evaluation results, either through monetary bonuses, salary increases, or both. The percentage of teachers agree-

ing that their "district/CMO's compensation system rewards teachers based on their effectiveness" tends to be higher in the CMOs (ranging from 47 percent in Green Dot to 88 percent in Aspire) than in the districts (ranging from 30 percent in SCS to 57 percent in HCPS). Given the widespread use of financial rewards in many of the Intensive Partnership sites, we might expect overall agreement to be higher.[4]

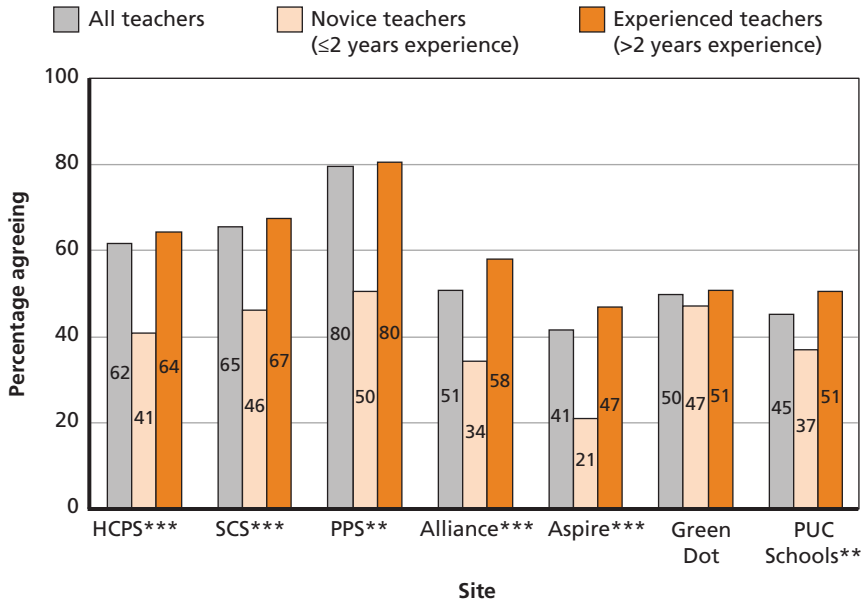
Teachers' beliefs about how teacher compensation systems should be structured are complex. A majority of teachers in all sites agreed that, "generally speaking, teachers should receive additional compensation" for demonstrating outstanding teaching skills, if their students show outstanding achievement gains, and for working in low-performing schools or with high-need students. This was true in all three survey years (2011, 2013, and 2014) and was generally true for both novice and experienced teachers. At the same time, a majority of teachers in the districts also agreed that "a teacher's base pay should be based on seniority," although fewer agreed in the CMOs. As Figure 5.2 shows, experienced teachers were more likely than novice teachers to agree with the statement. These patterns were present in 2011 and 2013 as well.

Teachers' Attitudes Were Mixed Regarding the Fairness and Incentive Effects of Their Site's Compensation Policies, with Teachers in Charter Management Organizations Slightly More Positive Than Teachers in the Districts

To provide context to teachers' perceptions about the fairness of the compensation systems in their sites, we examined the extent to which teachers reported that the total amount of compensation they receive allows them to live "reasonably well." In 2014, 41 to 61 percent of teachers agreed that "the amount of compensation I receive as a teacher allows me to live reasonably well" in all sites except PPS. The percentage of agreement in PPS—81 percent—was substantially higher than in all of the other sites.

[4] However, eligibility rules for financial rewards are such that some teachers, particularly those who are ineligible for the rewards, might be unaware that their district's or CMO's compensation system rewards teachers based on their effectiveness.

Figure 5.2
Percentage of Teachers Agreeing That "a Teacher's Base Pay Should Be Based on Seniority," 2014

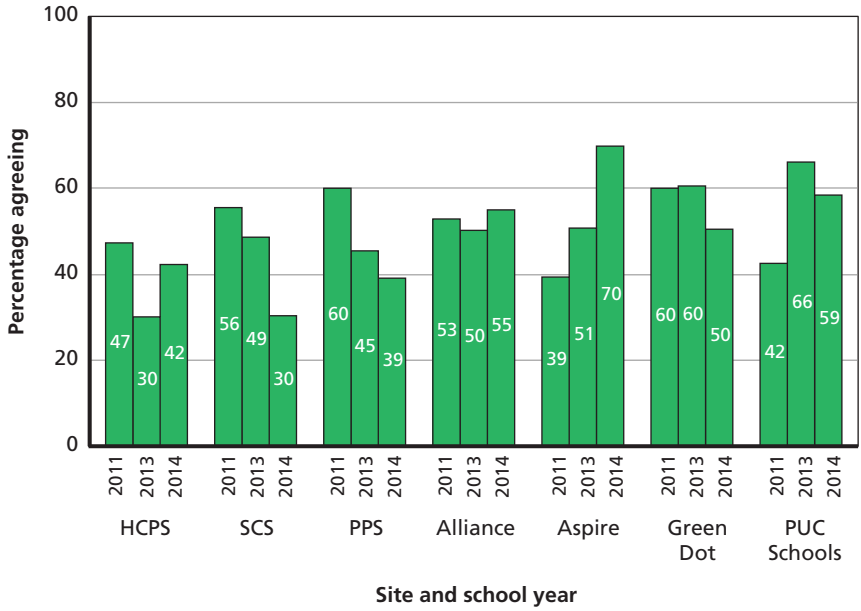


NOTE: ** = difference significant at $p < 0.01$. *** = difference significant at $p < 0.001$.
RAND *RR1295-5.2*

More than half of the teachers in the CMOs (but not the districts) reported that the way compensation decisions are made in their site is fair. Figure 5.3 shows the percentage of teachers agreeing that "the way compensation decisions are made in my district/CMO is fair to most teachers." Over time, there was increase in the percentage of teachers agreeing in PUC Schools and Aspire but a decrease in the percentage of teachers agreeing in SCS, PPS, and Green Dot.

Comments from teacher interviews illustrate mixed opinions about the fairness of compensation policies. For example, one PUC Schools teacher praised the use of performance-based pay: "I think it's a good system overall compensating for your performance; I think it's fair." However, an Alliance teacher in a nontested grade or subject expressed concerns about the fairness of one measure in particular:

Figure 5.3
Percentage of Teachers Agreeing That "the Way Compensation Decisions Are Made in My District/CMO Is Fair to Most Teachers," Springs of 2011, 2013, and 2014



NOTE: Each bar with a date reflects a different school year: 2010–2011, 2012–2013, and 2013–2014.

RAND *RR1295-5.3*

> I don't like the [student assessment] part . . . . I think [compensation] should be driven by you as a person if that kind of money is at stake. I mean, decisions to give bonuses should be based on parts of the evaluation that you control as a teacher.

Other teachers expressed concern about the uncertainty introduced by performance-based compensation. A PPS teacher described her concerns this way:

> Before my contract, teachers who were on their 10th step, who always moved up, you knew what you were going to make. There's no clear path for me . . . . if nothing changes, I could top off at $60,000. That's a pretty low salary here if you don't ever move

> across the pay scale. I'm not sure if things will change. I understand they're doing this to hold teachers accountable, but it's still a little scary when every administrator isn't on the same page and is doing things differently.

During school visits, teachers in HCPS tended to talk favorably about the new compensation system. For example, two HCPS teachers shared the following, respectively:

> It sounded pretty good to me. It was like a win-win situation. There was nothing shocking about this idea for us. It seemed pretty laid out. It seemed too good to be true, which is why most people waited too long to opt in because it was like, "well, our district has tricked us on things in the past; when's that going to come out? When are we going to find out that we did something that we should not have done?"

> I think the structure is great. If you're not doing the job, perhaps there's another job in an unrelated field that you can seek. If we're not up to the job, train us, and, if that still doesn't work, "maybe you should go into another field." I think seniority and not being able to be fired is wrong. I paid a secretary 18 years ago more than I get paid. People are paid more than me but have no business in the classroom and have been there 30 years too long. I'm all for this thing. The moment I heard of it I said, "finally."

In 2014, in most of the sites, a minority of teachers agreed that their district's or CMO's "compensation system motivates me to improve my teaching"; the exceptions to this trend were Aspire and PUC Schools, where 61 percent and 50 percent of teachers agreed, respectively. In four sites—the three districts and PUC Schools—teachers who work in high-LIM schools were significantly more likely than teachers who work in schools with lower percentages of LIM students to agree that their site's compensation system motivated them to improve their teaching.

Our interviews suggest that disagreement with the statement that the compensation system was motivating does not necessarily indicate

lack of support for the system; instead, it generally seems to reflect teachers' beliefs that other factors motivate them. An Aspire teacher stated, "I received a bonus, but it didn't change my motivation or practices. I'm pretty intrinsically motivated. Financial incentives would have to be significantly higher to retain people who weren't intrinsically motivated." Similarly, a PUC Schools teacher described the bonus as nice but not highly influential:

> [Bonuses are] really great; who's going to complain? But we're not in the job because it's a lucrative field. We do it because we love what we do. It's nice to have that fiscal validation. But it's not the deal-breaker for motivating whether I'm going to stay as a teacher or not.

Of course, such self-reports might not accurately reflect the behavioral effects of performance-based compensation, but they do provide some context for interpreting the survey results.

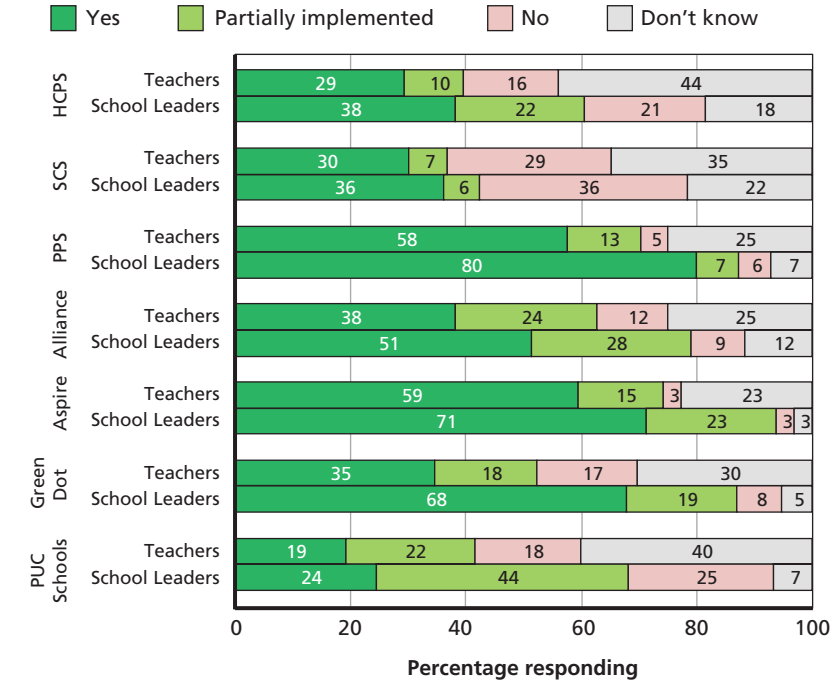
Perceptions of Career Ladders

Teachers and School Leaders Reported That Sites Have Fully or Partially Implemented Career-Ladder or Specialized Positions for Teachers in the 2013–2014 School Year

Career ladders can be useful for motivating teachers and encouraging retention only if teachers are aware that they exist and think they are motivating.[5] Figure 5.4 presents the distributions of teachers' and

[5] When completing the survey, HCPS respondents might not have considered the district's peer and mentor evaluators to be career-ladder positions because the survey defined such positions as "specialized instructional positions that teachers may take on if they are considered qualified." Indeed, as shown in Figure 5.4, relatively low percentages of teachers and school leaders in HCPS responded "yes" or "partially implemented" to the question, and, among those who did, only about half indicated that there were teachers in their schools who held higher-level career-ladder or specialized instructional positions. For these reasons, we have excluded HCPS results from the remainder of this section. On the other hand, although relatively few teachers and school leaders in SCS responded "yes" or "partially implemented" to the initial question, among those who did, 87 percent of teachers and 94 percent of leaders indicated that some teachers in their schools held higher-level career-ladder or specialized instructional positions. This suggests that, although only some SCS teachers and leaders were aware of their district's career-ladder system, unlike in HCPS, those who were aware

Figure 5.4
Percentage of Teachers and School Leaders Responding to "This Year, Does Your District/CMO Have in Place a 'Career Ladder' for Teachers, or Specialized Instructional Positions That Teachers May Take on If They Are Considered Qualified?"



RAND *RR1295-5.4*

school leaders' responses to the question, "This year, does your district/CMO have in place a 'career ladder' for teachers, or specialized instructional positions that teachers may take on if they are considered qualified?" In 2014, most teachers and school leaders (except in HCPS and SCS) reported that their site had fully or partially implemented career-ladder or specialized instructional positions for teachers as of the 2013–2014 school year. Results presented in the remainder of this section are

---

of it seemed to have fairly direct experience with teachers in career-ladder roles within their schools. Moreover, the career-ladder definition used in SCS aligns with the one used in the survey. For these reasons, we include SCS results in the remainder of this section.

restricted to the teachers and school leaders who responded "yes" or "partially implemented/being phased in."[6]

Among Teachers Reporting That a Career Ladder Was Present, Majorities in Most Sites Reported That the Process for Selecting Teachers for Career-Ladder Positions Is Fair, That They "Aspire to a Position on the Career Ladder," and That "the Opportunity to Take a Career Ladder Position" Motivates Them to Improve Their Instruction and Increases the Chances They Will Remain in Teaching

Majorities of teachers in most sites agreed with the following statements (see Figure 5.5):

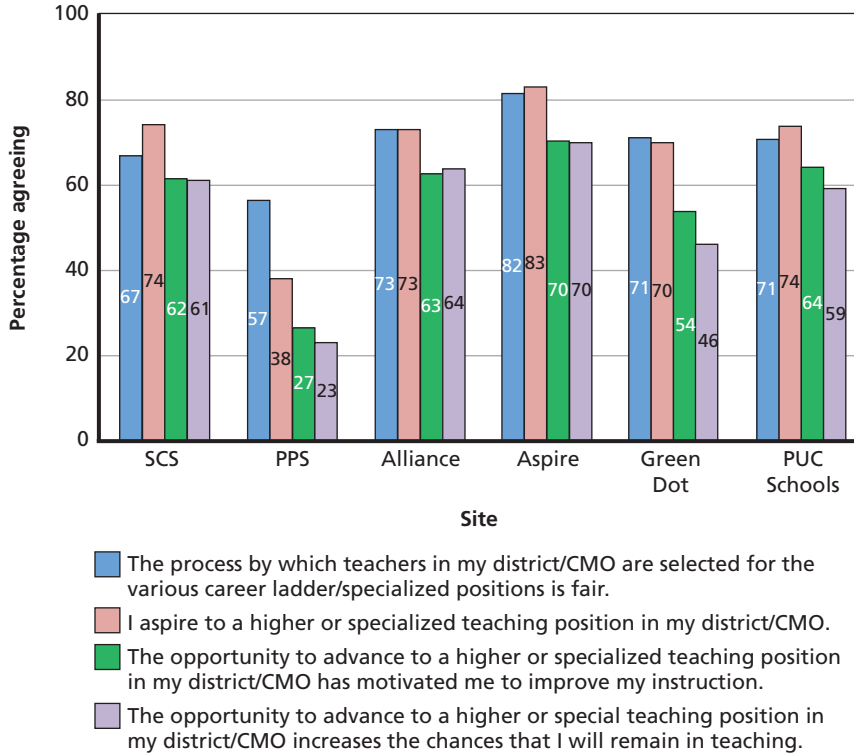
- "The process by which teachers in my district/CMO are selected for the various career ladder/specialized positions is fair."
- "I aspire to a higher or specialized teaching position in my district/CMO."
- "The opportunity to advance to a higher or specialized teaching position in my district/CMO has motivated me to improve my instruction."
- "The opportunity to advance to a higher or special teaching position in my district/CMO increases the chances that I will remain in teaching."

The fact that PPS teachers reported lower levels of agreement than teachers in other sites did might be attributable to the fact that PPS had relatively few openings for such positions in the 2013–2014 school year, but we know from interviews that teachers were also concerned about the short-term nature of the positions, the increased workload, and the sense that the positions had not been very effective in the past. PPS central-office staff we interviewed echoed these concerns; as one central-office staff member said,

> I think [the number of applications for career-ladder positions was] low because of the sustainability concerns that were known

[6] The survey explained "partially implemented/being phased in" as, "for example, some positions are currently available while others are still being developed." Other examples could be positions being piloted or available to teachers only at certain schools or grade levels.

Figure 5.5
Percentage of Teachers Agreeing with Statements About Career-Ladder Positions, 2014



RAND *RR1295-5.5*

> by all. If a teacher knows the position may go away, then he is less inclined to get involved. Also, if teachers do not see success, then they will not get involved. Also, some of the roles are a lot of work. Some of the most-distinguished teachers are already doing a lot of work. For teachers, there is a balance between "continue doing what you are doing" and asking the teacher to also help another teacher move up with them.

In addition, some teachers we interviewed in PPS expressed a lack of enthusiasm for these positions because of the extra work and the need to move to a new, usually high-need, school, as well as the uncer-

tainty regarding whether their current job would be available to them once the career-ladder position ended.

Most Teachers and School Leaders Reported Positive Perceptions of Teachers in Career-Ladder Roles

Many teachers and a majority of school leaders in all six sites indicated that, in their perception, teachers who hold higher-level positions at their school are effective educators, have helped improve student achievement at the school, and deserve the additional compensation they receive.[7] One Green Dot teacher who held such a position in the past described some of the benefits that teachers in specialized positions provide:

> I think the mentor teachers are incredibly helpful at each site, and the data fellows, understanding data is important at every site. One hundred percent, I think these career-ladder positions that Green Dot has created to have teachers help other teachers to get better are important, needed, and effective.

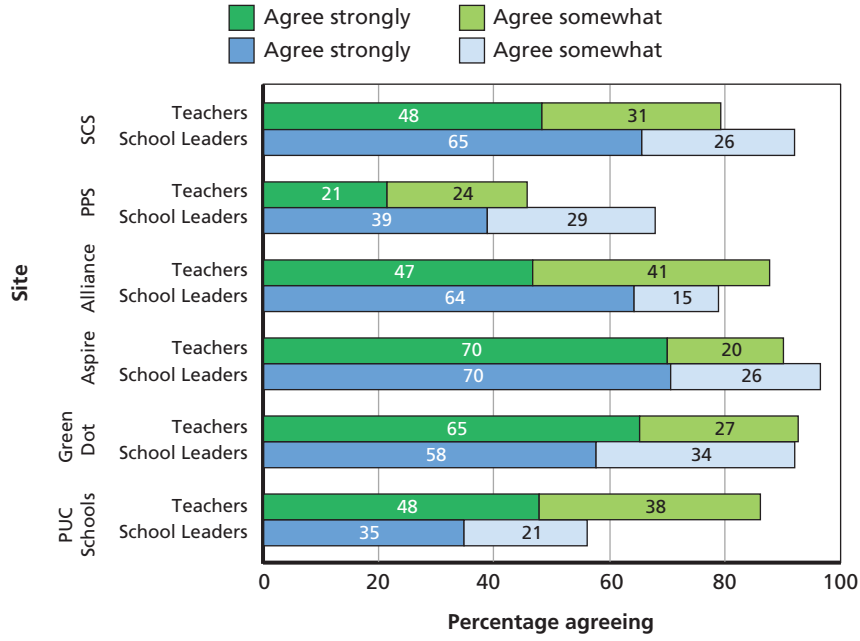
Figure 5.6 presents the percentage of teachers and school leaders who agreed that career-ladder teachers deserve the additional compensation they receive. In general, teachers and school leaders within any given site tended to have similar perceptions on this question, as did novice and experienced teachers.

Summary

Teachers' responses were mixed regarding the fairness and incentive effects of their site's compensation system, with teachers in the CMOs slightly more positive than teachers in the districts. Most district teachers thought that base pay should be based on seniority and a majority of all teachers also thought that teachers should receive additional compensation for demonstrating outstanding teaching skills and for teaching in low-performing schools. Many teachers do not object to such

[7] Novice teachers were more likely than experienced teachers to agree.

Figure 5.6
Percentage of Teachers and School Leaders Agreeing That "the Teachers Who Hold Higher-Level Positions at My School Deserve the Additional Compensation (Bonuses or Higher Salaries) They Are Receiving," 2014



RAND *RR1295-5.6*

incentives when they are offered as part of a system that also rewards experience and degrees. Teachers' attitudes were mixed when it came to the fairness and incentive effects of their site's compensation system, and most teachers in the districts favored a system in which base pay was determined, at least in part, by seniority. The sites are working to implement career ladders that offer effective teachers new roles as mentors or coaches in addition to their teaching duties; however, these new personnel policies are just beginning in many of the sites. Most teachers and school leaders expressed positive opinions about the career-ladder positions, indicating that teachers in those positions had contributed to improved student achievement and deserved the additional compensation they received.

CHAPTER SIX

Summary and Conclusions

The goal of the Intensive Partnership initiative is to dramatically improve student outcomes by increasing TE on a systemwide scale. The Bill & Melinda Gates Foundation believes that this increase can be achieved when high-quality measures of effectiveness are used to develop the teacher workforce, from selection and placement to support and PD to compensation and career differentiation. The Intensive Partnership initiative is designed to test this theory, and the results of the initiative are particularly relevant now that performance-based teacher evaluation is increasingly being adopted at the state and local levels. For example, as of the 2014–2015 school year, the majority of states (42 states and the District of Columbia Public Schools) had passed legislation requiring the incorporation of "objective" measures of student achievement and growth into teacher and principal evaluations (Doherty and Jacobs, 2013; AIR, undated; National Council on Teacher Quality, undated; Students First, 2013). The Intensive Partnership initiative, which started ahead of these recent policy reforms, could offer insight into how these policies can be implemented, the reactions of teachers and school leaders, and the challenges that have to be overcome.

Summary

Despite the fact that the seven Intensive Partnership sites differed in many ways (e.g., size, student demographics, and state education policy context), there were similarities in the way they designed and

implemented their TE reforms. For example, the process of developing and enacting the TE measures took about two years in each of the sites. After that effort was launched, changes were made, in roughly sequential order, to staffing policies, PD activities, and compensation and career-ladder policies. Given the differences in the sites, there were exceptions to these patterns; for example, the CMOs had some of the staffing policies in place prior to the start of the initiative, and PPS had developed an observation-based TE measure just before the initiative began. Excluding such preexisting differences, the sites went through the process in roughly the same order and on roughly the same schedule. (Of course, there were exceptions; for example, PPS adopted performance-based compensation and specialized teacher positions as early as 2011.) The sites were also broadly similar in terms of the resources invested to support implementation. Although the Intensive Partnership grants were important catalysts in focusing the sites on TE reforms, the total cost of the changes were relatively modest as a percentage of each site's overall budget (1 percent of total budget or less), with the largest portion of cost being associated with changes in how principals and teachers spent their time.

Teaching-Effectiveness Lever

It took each of the sites about two years to design and implement its TE measure, including engaging stakeholders, defining the component measures, training observers to rate classroom practice reliably, determining weights for the composite measure, and producing the effectiveness scores. All sites included structured classroom observations and student achievement growth in their TE measures and gave these two factors the greatest weight.

Overall, our surveys and interviews suggest that teachers and school leaders thought that the effectiveness measures had positive effects. Both teachers and school leaders reported that the information they received was useful for improving instruction, particularly the information from classroom observations. At the same time, teachers expressed some concerns about the validity of the components that go into the overall effectiveness measure, from the achievement tests and the student growth measures to the expertise of the observers who rate

teacher practice to students' judgments about teachers. Perceptions of validity tend to be higher for classroom observations than for student achievement growth and student survey measures. Yet, overall, teachers reported that the ratings had merit and were fair, particularly when asked about their own performance. Not surprisingly, teachers with high ratings were more likely than teachers with lower ratings to report that the ratings were accurate.

Most teachers reported that the sites emphasized using measures for improvement far more than for dismissal or termination, though some teachers expressed concerns that the measures might eventually lead to job loss or other undesirable outcomes. Large majorities of teachers in each site received ratings equivalent to "effective" or "highly effective," and the percentages in these categories have increased over time, which suggests that any negative consequences are likely to be limited to a relatively small proportion of teachers unless sites make significant changes to their measures, their cut scores, or their methods for combining scores from these measures into overall ratings.

Staffing Lever

Sites made changes to their procedures for recruitment, hiring, placement, tenure, and dismissal of teachers to try to improve the overall effectiveness of their teacher workforce. These changes included earlier identification of vacancies, more-aggressive recruitment, better screening of candidates in terms of effectiveness, more-strategic referrals to high-need schools, interview training for principals, and more-effective orientation and initial training for new hires. Sites varied in staffing changes because the CMOs and HCPS had many of these staffing practices in place at the start of the initiative and the new staffing policies that were enacted had to be consistent with local laws and contractual agreements. By 2013 or 2014, most school leaders reported that they were satisfied with the quality of new teachers coming to their schools and with the hiring process. However, teacher mobility remains an issue—the CMOs struggle to retain effective teachers, and some principals in the districts expressed dissatisfaction with teachers who were assigned to them without their agreement. Some of the results suggested a positive relationship between school leaders' per-

ceptions of autonomy and their satisfaction with staffing policies and teacher quality, as illustrated by the increases in satisfaction among SCS school leaders as they gained autonomy over decisions about staffing in their schools.

Professional-Development Lever

Sites did not make many changes to their PD practices for the first couple of years of the initiative, in part because they lacked comprehensive measures of teaching quality on which to base PD decisions. However, once the TE measures were operational, the sites began to explore strategies for supporting teachers based on their identified needs in an effort to improve instruction. The classroom-observation rubrics became the common language for discussing instruction, and both teachers and principals noted the advantages of having shared definitions about effective teaching. In addition, teachers often viewed the feedback received during and immediately after their observations as a way to improve their practices (and thus as a form of PD). During the first few years of the initiative, teachers in all sites reported receiving more PD, and clearly the sites were emphasizing instructional improvement. Recently, many of the sites began to place greater emphasis on supporting teachers through coaching or mentoring rather than formal workshops, courses, and the like. The coaches were usually teachers who had been recognized as effective, although, in some cases, the mentoring might occur as part of a teacher learning community within the school. Teachers reported that they preferred local coaching to more-formal PD.

However, sites found it challenging to customize PD to the needs of individual teachers for a variety of reasons. First, much of the formal PD that sites provided had been traditionally delivered in a group format, and individualizing training delivered in this manner is difficult. After three years, much of the PD remained group-oriented (e.g., workshops and institutes). Second, the new observation rubrics identified elements of effective teaching that were not always aligned with a site's existing collection of PD offerings. Sites had to develop, acquire, or identify training that addressed different aspects of teaching. Third, sites did not previously have information systems that would allow

sites to match support to individual teacher needs. One approach that some sites followed was to try to catalog or index their PD options to the dimensions of effective practice in the observation rubrics; but the interface that would link teachers' needs to PD opportunities was not always effective. Fourth, some sites tried to expand their PD offerings to include videos, online training, individual readings, and other methods, but not everyone was aware of these options. Fifth, it took more time for principals, coaches, and teachers to develop individualized development plans.

Individualized PD also presented challenges in terms of recordkeeping. Some sites put the onus on the teachers to seek out the PD they need; in other sites, the principals were responsible for overseeing the development of individual growth plans to meet teachers' needs. Neither approach made documentation easy. It was also very difficult to maintain good records of individualized PD that coaches or mentors delivered; some of these encounters occur on a planned schedule, but many occur informally.

Compensation and Career-Ladder Lever

Compensation reforms and differentiated career positions were implemented later than the other levers in most of the sites. By school year 2013–2014, all of the sites had adopted some form of year-by-year effectiveness bonus, i.e., awarding extra compensation for a year to teachers who receive the highest effectiveness ratings. In addition, all the sites are in the process of developing or implementing some career-ladder positions in which effective teachers take on specialized roles, such as coaching or mentoring, and receive a year-by-year bonus or a permanent salary increase or a stipend for the new responsibilities. Some of these positions are full time, but most are "hybrids" that expect teachers to continue spending some of their time in classroom instruction.

District teachers reported that they preferred a salary schedule that has a base pay determined by seniority with bonuses for effectiveness. Teachers gave more-mixed responses to the idea of career ladders. Some teachers indicated that they preferred extra compensation that is tied to increased responsibilities, as it is in the career-ladder positions;

others liked having extra compensation linked to performance on the TE measures; and some thought both were fine.

Variation Across Sites

In the report, we focus primarily on describing common patterns in implementing the levers across the seven Intensive Partnership sites rather than trying to explain variation among the sites. However, we noted some cross-site differences that warrant some discussion. At this time, we cannot establish causal connections between site characteristics, policy choices, and principal and teacher responses, but we can suggest some hypotheses that seem like reasonable explanations for some of the observed variation. Many of the observed actions and responses seem to be related to prior site experience, governance structures, state policies, and changes in leadership. We briefly describe these here, and we will explore between-site variation in greater detail in future reports.

The sites differed in the extent to which some elements were already in place at the start of the grant period. Those with some ingredients in place appear to have implemented related levers more rapidly (or more efficiently) than other sites. For example, HCPS already had student achievement tests in all subjects and grades; it was able to build on this past work to implement student growth measures for more teachers, whereas other sites had to determine other ways to compute student growth measures for teachers in subjects and grades not tested by their state. Similarly, Aspire and PUC Schools had their own observation rubrics in place prior to the Intensive Partnership initiative; thus, it was relatively easy for them to implement the rubric that the CMOs adopted.

Some fundamental differences in governance between districts and the CMOs influenced implementation and attitudes. Perhaps the most-significant difference related to relationships with teachers. There were no tenure provisions in any of the CMOs, and, with the exception of Green Dot, they did not have to negotiate work rules with an organized teacher bargaining unit. This meant that it was far easier for the leadership to implement many of the Intensive Partnership levers. Although the CMOs went to great lengths to involve staff and incor-

porate their input, in the end, they could implement bonuses and performance pay more easily than the districts could.

Finally, changes in state laws and regulations affected almost every site, and every site but one had a change in site leadership during the past five years. For the most part, the changes in state regulations were significant. For example, California stopped statewide testing for two years while redesigning its accountability system, and the CMOs were left without a test they could use to compute student growth measures. Other significant state changes included legislation in Tennessee leading to the merging of MCS with SCS, legislation to support Tennessee's RTT proposal that mandated teacher evaluation while eliminating tenure and seniority-based staffing, Pennsylvania's Act 82 that mandated teacher evaluation based on multiple measures, and the elimination of teacher tenure in Florida. Cuts in funding were also significant in PPS and the CMOs. On the other hand, changes in site leadership have not had dramatic effects on the implementation of the Intensive Partnership initiative. For the most part, new leaders have continued to support the reforms with minor or no modifications.

Discussion

It is premature to draw overall conclusions about the implementation of the Intensive Partnership initiative—the sites' implementation grants include school year 2015–2016, and the sites are still modifying policies and procedures as they gain more experience. There are many questions for which we have, at best, partial answers, and we want to highlight a few of them here because they are likely to reflect challenges with which other sites will struggle if they decide to adopt similar reforms. We expect to address some of them in the remaining two years of the evaluation, but we know that the experiences of Intensive Partnership sites alone will not answer others.

How Long Should It Take to Implement Such Human-Capital Reforms as the Intensive Partnership Initiative?

In developing the Intensive Partnership initiative, the foundation recognized that change takes time, particularly change designed to modify core district policies and practices. As a result, they funded the sites for six years and encouraged them to plan strategically, engage stakeholders, take the time to review and revise, and so on. How much time should another state or district expect to devote to such reforms? It appears to us that the sites benefited from the time they devoted to initial planning and early engagement (e.g., involving stakeholders in the design and development of key components of the teacher-evaluation systems, particularly the observation rubrics and the weighting of elements that went into the overall effectiveness score). Many of the sites also took time to operate their systems on a low-stakes or pilot basis first before making them official. These initial processes took a couple of years, on average, and the time spent appears to have helped the sites fully engage teachers and other stakeholders and develop policies that had their general support. Similarly, it appears to have been an effective strategy to focus on the creation of the teacher-evaluation metric first and then address the components of the reform that use the metric for other functions—customized PD, compensation and career ladders, and retention and dismissal. It remains to be seen whether another site could enact similar changes more quickly, but the Intensive Partnership experience suggests that there were advantages to taking adequate time to develop and implement the effectiveness measures. In addition, a newly adopting site could probably learn useful lessons from the experiences of the Intensive Partnership sites in many areas, including training observers, defining terms in the observation rubric, combining multiple measures into a single effectiveness score, creating specialized career-ladder positions, and developing PD to support identified needs.

How Should Sites Interpret Improvements in the Distributions of Teacher-Evaluation Ratings over Time?

Across sites, a substantial majority of teachers received ratings equivalent to "effective" or "highly effective." The percentage of teachers per-

forming as effective or higher increased from 2012 to 2014 in most sites. The growing percentages of teachers performing at the highest levels on the teacher-evaluation measures, along with the shrinking percentages scoring at the lowest levels, could be interpreted as evidence that the reform is "working" as a means of improving teaching quality in the Intensive Partnership sites. However, other factors could explain these shifts. For instance, observers might be applying the classroom-observation rubrics in a more generous way out of concern for the increasing stakes attached to teachers' performance. Before concluding that the quality of teaching has risen, it will be important to examine and rule out other possible explanations for the changes.

How Does a Site Keep Such Reforms on Track in the Face of Resistance and Changing Local Conditions?

To sustain the reform, every site had to figure out how to cope with unanticipated changes in state policy; most had to address complaints or criticisms from various stakeholders, and almost all had to deal with unanticipated changes in site leadership. How did they manage to stay on course in the face of these challenges? There is no simple answer or formula (and some sites are veering further from the initial vision than others), but a range of factors probably helped them to stay the course. One factor is their initial effort to develop long-term strategic plans for the implementation of the reforms. The foundation program officers played a key role in this effort. The program officers helped the sites develop strategies to manage change, including communicating about the reforms, engaging stakeholders, and being responsive to their concerns.

Another factor in sustaining the policy changes was the endorsement of the reform from the larger local community, including local student advocacy groups and philanthropic organizations. A third factor was taking the time in the beginning to communicate about the reform and try to ensure that all relevant constituents understood what was intended and had opportunities to give input. The goodwill developed in the initial years was potentially helpful in later years as challenges arose. There were probably many other factors that are not easily replicable, including a history of effective change; the presence of

strong, committed leadership; and the ability of all parties to agree on a common direction for change—these were parts of the foundation's initial selection process. Furthermore, at least one of the sites seems to have encountered growing resistance to some features of the reform, and teachers in some sites are voicing some discontent with the evaluation systems, particularly once stakes are attached. Nevertheless, there are elements in the Intensive Partnership implementation process that others could emulate.

The changing local context also created challenges for the foundation, and one factor in maintaining the forward momentum of the initiative has been the foundation's willingness to adapt its vision for the initiative in response to the sites' needs. For example, early expectations were that student growth measures would be the dominant factor in the effectiveness measure, but those expectations changed in response to concerns from teachers in the sites about the validity of the VAM metrics and in response to evidence from the MET study that a more equal weighting was preferable. Many at the foundation also expected that the most important mechanism for improvement would be the termination of ineffective teachers and the retention of highly effective teachers. But informed by the experiences of the Intensive Partnership sites, foundation staff have embraced the idea that improving the effectiveness of the largest group of teachers, who fall in the middle of the distribution, could be the more potent option.

How Important Was the Role of the Bill & Melinda Gates Foundation in the Implementation of the Initiative?

From our perspective, the foundation appeared to be a key player in the implementation, although the site leaders who participated in our interviews did not attribute quite as much influence to the foundation as our other data suggest. In addition to providing funding for the initiative, the foundation played other roles that we see as key: sustaining the vision, convening the sites for learning and dialogue, connecting the sites with experts to meet their needs (including consultants in communication and instructional technology), and acting as critical friend. This last role included providing an engaged program officer who could reflect on the initiative from an outsider's perspec-

tive, not responsible for other duties, and not situated within the local organization.

How Does a Site Secure and Maintain Buy-In from Teachers and School Leaders for Effectiveness-Based Reforms, Particularly as the Stakes Are Increased?

As noted already, the Intensive Partnership sites have all taken teacher buy-in seriously from the inception of the reforms and, accordingly, have taken steps to secure it—namely, by promoting the reforms as mechanisms to help teachers improve their practice. Even as of the spring of 2014, most teachers in the Intensive Partnership sites still perceive that the primary purpose of their site's evaluation system is to provide feedback to help them improve their instruction, and many report that they have reflected more about their teaching because of the reforms.

Yet, although teachers see value in the effectiveness reforms for helping them to improve their instruction, the introduction of stakes tied to teacher performance—for example, consequences related to compensation, tenure, and employment itself—could come at the cost of teacher buy-in, even among teachers who perform relatively well. Some of the sites are beginning to experience reduced support from teachers as stakes are introduced and intensified. In at least a couple of the sites, teachers have begun to report that the observations are stressful and punitive, rather than helpful and informative. The stakes could turn out to be counterproductive, if teachers throw out the baby (improved teaching practice based on feedback and data) with the bathwater (stakes with the perceived potential to affect teachers' livelihoods and sense of professionalism).

How Does a Site Best Support Teachers in Improving Their Practice?

Improving teaching practice could be the major mechanism through which the Intensive Partnership initiative succeeds. The Intensive Partnership sites have implemented a range of strategies to help teachers improve their effectiveness, including centralized PD targeting common challenges, customized workshops to target the needs of groups of teachers, local coaching to help teachers address their own

individual challenges, and collaborative communities of practice in which teachers work with one another to improve practice. The sites have not yet determined which of these strategies is most effective, but teachers and school leaders do seem to perceive some of the strategies as more valuable than others. We hope to learn whether their perceptions reflect reality and whether the less-popular strategies should be deemphasized (or even abandoned).

Future Analyses

We will continue to evaluate the implementation and impact of the Intensive Partnership initiative through its six-year life; the foundation grants are scheduled to end after the 2015–2016 school year. In the subsequent two school years, the evaluation team will be providing updates on the implementation of the reform, including its cost, and we will look specifically at the steps the sites are taking to sustain the reforms after the foundation grants expire.

In addition, we will investigate other aspects of the initiative, including the quality of the TE measures, e.g., whether scores from the observation measures are reliable and whether the overall scores are valid indicators of effectiveness, whether the new measures are more sensitive than the old performance appraisal process, and whether newly established performance levels reflect higher standards than the old. We will also issue reports on the reform's effects on student outcomes, including (1) whether the initiative affected the level and distribution of TE; (2) whether it affected student achievement, graduation, and college enrollment; and (3) whether selected levers were effective in raising student outcomes. Finally, another component of the evaluation will look at the extent to which other districts replicate the policies and practices that the Intensive Partnership sites adopted. As we noted, many of these policies are becoming law in other states, and we will focus our attention on the extent to which others learn from the experience of the Intensive Partnership sites when implementing their mandated reforms.

APPENDIX A

Methods for Interview Data Collection and Analysis

Each fall, we conducted in-person interviews with the key central-office administrators in each Intensive Partnership site (see Table A.1) who were involved in developing, implementing, or reviewing the Intensive Partnership levers, as well as two or three selected local stakeholders (e.g., teachers' union officials, school board members). The interviews focused on the development and implementation of the Intensive Partnership reforms and policies, such as the use of TE ratings, development and implementation of targeted PD, challenges, implementation

Table A.1
Number of Central-Office Administrators and Stakeholders Interviewed

Year	HCPS	SCS	PPS	Alliance	Aspire	Green Dot	PUC Schools
Fall 2010	21	12	19	1	1	2	2
Fall 2011	21	10	28	3	3	4	3
Fall 2012	11	12	18	9	8	8	5
Fall 2013	14	15	18	9	5	7	5
Fall 2014	12	13	17	13	7	10	8

NOTE: We have also interviewed TCRP leaders who coordinate activities among the CMOs. We conducted interviews with five leaders in 2010, two in 2011, one in 2012, and one in 2014. The numbers of interviewees changed over time as a result of site input into which staff should be included. For example, in HCPS, during the initial years, we interviewed several staff who worked in finance or information technology; several of these staff were dropped from the sample in later years.

successes, local contextual factors, and interactions with the foundation and with other districts.

Each spring, we conducted in-person and telephone interviews with school staff at seven schools in each district and one to two schools in each of the four CMOs. We purposefully sampled the schools with feedback from staff in each site to ensure representation across grade-level configurations, geography, and achievement levels. We also considered site-specific implementation factors, such as piloting or implementation of policies or programs of interest in certain schools. In the first year of the project (the spring of 2011), we conducted in-person visits at all seven schools in each site and all seven schools in the CMOs. During these visits, we conducted individual interviews with three school leaders (including the principal) and three teachers, as well as a focus group of six to eight teachers. In the second year of the study (the spring of 2012), to minimize burden on the schools, we conducted in-person visits at half of the schools in each site, conducting interviews with school leaders and teachers as described above, and telephone interviews with the principals in the remaining schools. We randomly selected schools for each group and switched them in subsequent years (e.g., schools that received in-person visits in the spring of 2012 received telephone interviews in the spring of 2013).

In the third year of the study (the spring of 2013), we adjusted our participant sample with the goal of increasing the number of teachers interviewed. In the schools that received in-person visits, we reduced the number of school leaders sampled from three to two (i.e., the principal and another school leader), and we increased the number of teachers sampled for individual interviews from three to four. The teacher focus group was unchanged. In the schools that received telephone interviews, we sampled two teachers for telephone interviews in addition to interviewing the principals. In the fourth year of the study (the spring of 2014), we further refined our participant sample; in the schools that received telephone interviews, we reduced the number of teachers sampled from two to one to reduce the burden on the school. Sampling for the schools that received in-person visits was unchanged. In the three districts, teachers who participated in focus groups scheduled after school hours received a $25 gift card to thank them for their

time. Table A.2 shows the number of school-level staff interviewed in each site.

A member of the research team conducted each interview using a semistructured protocol to guide the questioning. We also used probe questions as needed to follow up. We informed all participants that their interview responses would be confidential and that any reporting would be done in the aggregate. We also informed participants that no responses or quotations would be reported in a way that would allow

Table A.2
Number of School-Level Staff Interviewed

Interview	HCPS	SCS	PPS	Alliance	Aspire	Green Dot	PUC Schools
Spring 2011 school visit							
Number of school leaders interviewed	18	20	21	4	1	4	2
Number of teachers interviewed (individual and focus group)	52	65	63	19	9	15	3
Spring 2012 school visit							
Number of school leaders interviewed	11	13	13	3	3	4	3
Number of teachers interviewed (individual and focus group)	31	34	42	12	9	9	9
Spring 2013 school visit							
Number of school leaders interviewed	12	8	8	3	2	2	2
Number of teachers interviewed (individual and focus group)	47	48	37	15	15	4	4
Spring 2014 school visit							
Number of school leaders interviewed	8	8	9	3	3	2	3
Number of teachers interviewed (individual and focus group)	33	40	45	11	9	2	13

them to be identified. School-based in-person and telephone individual interviews lasted approximately 45 minutes, and the in-person focus group lasted approximately one hour. We randomly sampled teachers for the individual interviews and focus groups to ensure variability across grades and subjects (tested and not tested), years of teaching experience, and level of involvement or holding special roles in the school (e.g., coaching or career-ladder roles). We requested the staff rosters used for sampling directly from the district central office or from the principals of the CMO schools, and we requested supplemental information (e.g., teachers serving in coaching or career-ladder roles) from the principals. Interviews with central-office staff lasted one hour.

The analysis of the interview data each year proceeded in several steps. First, we compared interview notes to the audio recording and cleaned them to serve as a near-transcript of the conversation. We then loaded the cleaned interview notes into the qualitative analysis software package NVivo 10 and auto-coded them by interview question (i.e., so that responses to specific interview questions were easily accessible). We also coded them using a thematic codebook that we developed. Once we finished the thematic coding, we conducted a second round of coding, analyzing the data according to research questions of interest (e.g., how do principals' opinions about the teacher-evaluation measures differ from teacher opinions). In this stage, we used an inductive coding process (i.e., we derived codes from the data rather than from a structured codebook) to develop responses to the question of interest. The codebook remained largely unchanged since the beginning of the study, with some minor revisions to eliminate redundancies or to capture new themes as they emerged. The consistency of the codebook and coding methodology over time allowed us to examine changes over time, as well as look at each year's interviews individually.

APPENDIX B

Methods for Coding Implementation Status

Our analysis of implementation draws on the central-office interviews and documentation from the sites to examine the status of each site's implementation policies and procedures and to track changes over time. We relied on two data sources: interviews with eight to 12 central-office staff in each site from the fall of 2010 through the fall of 2014 and site-produced documents, including annual stocktake reports for the Bill & Melinda Gates Foundation, as well as other Intensive Partnership reform status updates.

We use the term *lever* to refer to four broad groups of specific policies and practices that sites adopted: teacher evaluation, staffing, PD, and compensation and career ladders. Within these four levers, we described the specific policies and practices being implemented across the sites. We focus on these levers because they are areas included in the theory of action that guides the implementation of the Intensive Partnership reforms. Table B.1 lists the specific policies and practices included in each lever, along with definitions.

To describe the pattern of implementation over time, we classified each site as "implementing" or "not implementing" a practice at each of five time points,[1] spanning the period from the time the Intensive Partnership initiative funding was awarded, in the spring of 2010, through

[1] After we developed the implementation tables, we shared them with site leaders to confirm their accuracy. In a few cases, we made changes to our classifications in response to additional information provided by these site leaders.

Table B.1
Levers, Practices, and Definitions

Lever and Practice	Definition
Teacher-evaluation lever	
Observation by principals or other administrator included in formal evaluation	The formal evaluations for all teachers include observations by the principal or another administrator at the teacher's school.
Observation by an additional set of observers (e.g., other school leaders, content-area specialists, peers, central-office administrators, coaches) for at least some teachers included in formal evaluation	The formal evaluations for at least some teachers include observations by observers other than the principal or another administrator (e.g., peers, school leaders).
Student or parent surveys included in formal evaluation	The formal evaluations for all teachers include surveys of students or parents.
Other measures of TE (e.g., content knowledge, professionalism, peer survey) included in formal evaluation	The formal evaluations for all teachers include other measures of TE (e.g., content knowledge, professionalism, peer survey).
Individual VAM or SGP score for subjects and grades with state test included in formal evaluation	The formal evaluations for teachers who teach grades or subjects with state tests include individual measures of student outcomes (e.g., VAM or SGP).
Individual value-added or SGP score for subjects and grades with no state test or other alternative measures of student growth included in formal evaluation	The formal evaluations for teachers who teach grades or subjects without state tests include individual measures of student outcomes (e.g., VAM or SGP) or some other alternative measure of student growth (e.g., portfolio or rubric-based measure).
Multiple measures combined using weights	Multiple measures of TE are weighted and combined into a single score or rating (e.g., performance level) for all teachers.
Data warehouse established for TE data	The site has a data warehouse in which TE data are stored and via which they are accessed.

Table B.1—Continued

Lever and Practice	Definition
Staffing lever	
Early or expedited recruiting and hiring for high-need positions	Recruiting and hiring for high-need positions (e.g., math teachers) occurs early in the calendar year (e.g., February or March) or can be accomplished more quickly than is typical for other positions in the district.
Early hiring for all vacancies	Hiring for any vacant teaching position occurs early in the calendar year (e.g., February or March).
Schools make final hiring decision	Teacher-hiring decisions are routinely made at the school level (e.g., by the school leader) rather than at the district level.
Administrators trained to make good hiring decisions (e.g., in interviewing and team-building)	All administrators are offered training to help them make good decisions about which teachers to hire. Such training includes how to conduct informative interviews and to hire candidates who would contribute to the building's overall functioning and effectiveness.
New applicant screening model based on TE rubric	The site has a method for screening all teacher applicants that is based on the rubric used to measure TE.
Incentives offered to work in high-need schools and classrooms	The system offers incentives (e.g., signing bonuses) to work in schools or classrooms that serve high-need (i.e., predominantly low-income and minority) students.
Transfers and furloughs not heavily influenced by seniority	Seniority does not heavily influence decisions about teacher transfers and furloughs (e.g., the least-senior teachers are not necessarily the first to be transferred or furloughed).
School leaders make final decision about which teachers are placed in their schools	Decisions about how teachers are assigned to subjects and grades within schools are routinely made at the school level (e.g., by the school leader) rather than at the district level.
Tenure and retention linked to effectiveness ratings	Award of tenure and retention (i.e., continuing in a teaching position) is linked to effectiveness ratings.

Table B.1—Continued

Lever and Practice	Definition
Effectiveness rating used as basis for dismissal	Any teachers with low effectiveness ratings can be dismissed.
Schools make final decision about teacher retention and dismissal	Decisions about which teachers are dismissed or asked to stay are routinely made at the school level (e.g., by the school leader) rather than at the district level.
PD lever	
Use evaluation data to identify teacher development needs	Principals or other supervisors use evaluation data to determine what PD should be offered for each teacher.
Offer PD designed to improve specific teaching skills measured in the evaluation	All teachers have access to PD that is designed to improve specific teaching skills measured in the evaluation.
Link coaching and mentoring feedback to evaluation components	Coaching and mentoring feedback is linked to evaluation components (i.e., feedback is specific to teacher development needs as identified in the evaluation).
Provide induction, mentoring, coaching, or academies for new teachers	The site provides induction, mentoring, coaching, or academies for new teacher hires.
Supervisors systematically oversee teachers' PD participation	Supervisors (e.g., school administrators, teacher leaders) systematically monitor and provide oversight of teachers' participation in PD.
Electronic system for PD data collection	The site has an electronic system for collecting data about PD (e.g., participation, frequency, topic).
Compensation and career-ladder lever	
Bonuses, stipends, or salary increments awarded based on individual effectiveness measures	The site offers bonuses, stipends, or salary increments based on individual measures of effectiveness.
Traditional step-based salary schedule not used exclusively	The site does not compensate teachers exclusively according to a traditional step-based salary schedule, which links higher pay with greater experience.

Table B.1—Continued

Lever and Practice	Definition
Bonuses or salary increments given for high-need positions	The site provides bonuses or salary increments for high-need (i.e., positions for which there are typically few qualified candidates, such as special education or high school science) positions throughout the district.
Incentives given for desired teacher behavior (e.g., low absenteeism)	The site provides incentives (e.g., salary increments or bonuses) for desired teacher behavior (e.g., low absenteeism).
Positions created for effective teachers with different responsibilities	The site offers positions with different responsibilities (e.g., coaching) to effective teachers.

the spring of 2014.[2] We classified a practice as implementing if it was in effect for all intended staff[3] or if it was being piloted and was not yet in effect for all intended staff. Otherwise, we classified the practice as not implementing. We then assigned one point for practices that we classified as implementing and zero points for practices that we classified as not implementing. We summed point values for each of the four levers over each time period and then converted them to percentages. In Appendix D, we present the detailed lever tables for each site.[4] We intend these tables to provide a relatively simple way of summarizing the status of implementation in each site, but, of course, they do not capture the details of each site's many related efforts.

[2] *Spring of 2010* describes the practices the sites had in place at the beginning of the initiative (as described in their proposals); *spring of 2011* summarizes implementation as of April–May 2011, at the end of the first full school year after the initiative was launched. Our most-recent summary, from the spring of 2014, describes implementation status as of April–May 2014, the end of the fourth school year of the initiative.

[3] We consider an evaluation measure implementing when it is obtained, or calculated, for all intended teachers regardless of when consequences were attached to the measure.

[4] Appendix D is available online only (Stecher, Garet, Hamilton, forthcoming).

APPENDIX C

Methods for Survey Data Collection and Analysis

Target Population and Sampling

In each Intensive Partnership site, the survey sampling frame included all regular, public schools serving students in grades K–12.[1] Table C.1 presents the number of schools in each site in each year.

We surveyed all school leaders and a sample of teachers from every school within each site. We used a stratified random sampling procedure to select the teachers, taking into account subject area taught and years of teaching experience;[2] the number of teachers selected in each school varied by site and school level. School leaders included principals, assistant principals, and all other staff holding equivalent titles (e.g., director, instructional leader, dean). Table C.2 shows the total number of teachers and school leaders invited to participate in the survey during each administration.

[1] We excluded charter schools in the three districts. In 2014, we excluded schools in SCS that were with the district only temporarily (i.e., "legacy" Shelby County schools that were departing to municipalities following the 2013–2014 year).

[2] Specifically, we stratified based on core/noncore subject area, in order to ensure adequate representation from teachers of all types. We defined core teachers as general-education teachers of reading and ELA, mathematics, science, social studies, and (at middle and high school levels) foreign languages. We defined noncore teachers as teachers of other subject areas and special-education teachers. Our samples typically consisted of approximately 80 percent core teachers and 20 percent noncore teachers. In addition, we oversampled novice teachers in the districts (which have high proportions of experienced teachers) and experienced teachers in the CMOs (which have high proportions of novice teachers) to ensure adequate representation from each group.

Table C.1
Number of Schools Surveyed

Year	HCPS	MCS/SCS	PPS	Alliance	Aspire	Green Dot	PUC Schools
2011	239	191	62	18	30	16	12
2012[a]	228	188	60	20	34	18	13
2013	240	178	54	21	34	18	13
2014	240	186	54	20	37	16	13

[a] In 2012, we surveyed only school leaders. In HCPS, some small alternative schools lacked school leaders, so the 2012 number of schools is slightly smaller than that for the other years. Other year-to-year changes reflect growth or decline in the actual number of schools in each site.

Table C.2
Number of Teachers and School Leaders Surveyed

Year	Teachers	School Leaders
2011	4,311	1,174
2012[a]	N/A	1,209
2013	4,697	1,172
2014	4,838	1,287

[a] In 2012, we surveyed only school leaders.

Data Collection

Surveys were web-based and administered in the late spring. As of this report, teachers have been surveyed three times: the springs of 2011, 2013, and 2014. School leaders have been surveyed four times: the springs of 2011, 2012, 2013, and 2014. We designed both surveys to take 45 to 60 minutes to complete, except for the 2014 teacher survey, which was a short version designed to take 20 to 30 minutes to complete.

We contacted survey recipients at the email addresses that site central offices provided to the RAND team responsible for collecting site administrative data. We provided each recipient with a unique link to access the survey. The link included an embedded identification code by which we could track responses and merge them with administrative data, such as teachers' grade level taught and effectiveness rating, and school demographic characteristics. We contacted nonrespondents about once a week throughout the data-collection period, initially by email and later by phone. Every individual who completed the survey received a $25 gift card;[3] there were also occasional drawings for $50 gift cards and, at the end of each year, a final drawing for $500 school prizes from among schools with high response rates.

We calculated the survey response rate as the number of responding teachers and school leaders divided by the number of sampled teachers and school leaders.[4] Tables C.3 and C.4 show the response rates for teachers and school leaders, respectively, in each site in each year.

Data Analysis

- **weighting:** We calculated sampling weights for each teacher based on the sampling design. (School leaders had an implicit sampling weight of 1 because all school leaders were surveyed.) Following data collection, for both teachers and school leaders, we conducted nonresponse analyses to adjust the weights. The nonresponse analysis was conducted as a two-level hierarchical generalized linear model (individuals nested within schools) predicting the probability of response based on person-level characteristics, such as gender and years of experience, as well as school-

[3] All gift cards were iCards, allowing respondents to choose from among some 200 merchants. In 2014, for the short teacher survey, each teacher was offered a $10 gift card rather than a $25 gift card.

[4] To be included in the response-rate calculation, as well as in the analysis, a survey had to have at least one question answered in more than half of the major survey sections.

Table C.3
Teacher Response Rates, Surveys Completed, and Teachers Sampled

	HCPS			MCS/SCS		
Year	Rate (%)	Completed	Sampled	Rate (%)	Completed	Sampled
2011	84	1,168	1,393	82	1,052	1,282
2013	75	1,040	1,393	83	1,038	1,244
2014	79	1,109	1,397	84	1,087	1,298

	PPS			Alliance		
Year	Rate (%)	Completed	Sampled	Rate (%)	Completed	Sampled
2011	78	657	838	77	140	182
2013	75	586	783	77	313	407
2014	70	548	780	79	344	435

	Aspire			Green Dot		
Year	Rate (%)	Completed	Sampled	Rate (%)	Completed	Sampled
2011	86	261	303	65	132	203
2013	79	285	359	61	206	335
2014	80	300	375	68	231	341

	PUC Schools		
Year	Rate (%)	Completed	Sampled
2011	82	90	110
2013	76	134	176
2014	75	159	212

level characteristics, such as percentage of students who are LIM and school level (elementary, middle, or high). Accordingly, the reported survey percentages represent the full populations of teachers or school leaders in each site.

- **data analysis:** We conducted survey analyses in Stata, using Stata's survey estimation procedures. For both teachers and school

Table C.4
School-Leader Response Rates, Surveys Completed, and Leaders Sampled

	HCPS			MCS/SCS		
Year	Rate (%)	Completed	Sampled	Rate (%)	Completed	Sampled
2011	77	465	607	76	259	339
2012	81	493	610	82	277	337
2013	77	459	597	65	207	317
2014	68	433	637	66	254	386

	PPS			Alliance		
Year	Rate (%)	Completed	Sampled	Rate (%)	Completed	Sampled
2011	83	85	102	59	23	39
2012	80	78	97	67	33	49
2013	74	64	86	65	31	48
2014	71	58	82	78	43	55

	Aspire			Green Dot		
Year	Rate (%)	Completed	Sampled	Rate (%)	Completed	Sampled
2011	81	30	37	56	18	32
2012	72	38	53	66	25	38
2013	69	33	48	65	33	51
2014	62	32	52	71	37	52

	PUC Schools		
Year	Rate (%)	Completed	Sampled
2011	72	13	18
2012	76	19	25
2013	72	18	25
2014	70	16	23

leaders, we specified a two-stage design, with schools as the first stage and individuals as the second stage. At the first stage, we treated each site as a stratum, and we included a finite population correction for the number of schools in each site. At the second stage for teachers, we treated core and noncore teachers within each school as strata, with a finite population correction for the number of teachers (within school) in each stratum. At the second stage for school leaders, we specified principals and assistant principals within each school as strata, with a finite population correction for the number of leaders in each stratum.

References

American Institutes for Research, Center on Great Teachers and Leaders, "Databases on State Teacher and Principal Evaluation Policies," undated. As of May 22, 2015:
http://resource.tqsource.org/stateevaldb/

Bill & Melinda Gates Foundation, "Foundation Commits $335 Million to Promote Effective Teaching and Raise Student Achievement: Bill & Melinda Gates Foundation," press release, undated. As of October 5, 2015:
http://www.gatesfoundation.org/Media-Center/Press-Releases/2009/11/Foundation-Commits-$335-Million-to-Promote-Effective-Teaching-and-Raise-Student-Achievement

———, *Intensive Partnerships to Empower Effective Teachers: Invitation-Only Request for Proposal*, April 2009.

Chambers, Jay, Iliana Brodziak de los Reyes, and Caitlin O'Neil, *How Much Are Districts Spending to Implement Teacher Evaluation Systems? Case Studies of Hillsborough County Public Schools, Memphis City Schools, and Pittsburgh Public Schools*, Santa Monica, Calif.: RAND Corporation, WR-989-BMGF, 2013. As of September 30, 2015:
http://www.rand.org/pubs/working_papers/WR989.html

Chambers, Jay, Iliana Brodziak de los Reyes, Antonia Wang, and Caitlin O'Neil, *How Are School Leaders and Teachers Allocating Their Time Under the Partnership Sites to Empower Effective Teaching Initiative?* Santa Monica, Calif.: RAND Corporation, WR-1041-1-BMGF, 2014. As of September 30, 2015:
http://www.rand.org/pubs/working_papers/WR1041-1.html

Coburn, Cynthia E., "Shaping Teacher Sensemaking: School Leaders and the Enactment of Reading Policy," *Educational Policy*, Vol. 19, No. 3, July 2005, pp. 476–509.

Dee, Thomas, and James Wyckoff, *Incentives, Selection, and Teacher Performance: Evidence from IMPACT*, Cambridge, Mass.: National Bureau of Economic Research, Working Paper 19529, October 2013. As of July 30, 2015:
http://www.nber.org/papers/w19529

Doherty, Kathryn M., and Sandi Jacobs, *Connect the Dots: Using Evaluations of Teacher Effectiveness to Inform Policy and Practice*, Washington, D.C.: National Council on Teacher Quality, October 2013. As of April 8, 2015: http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

Fryer, Roland G., *Teacher Incentives and Student Achievement: Evidence from New York City Public Schools*, Cambridge, Mass.: National Bureau of Economic Research, Working Paper 16850, March 2011. As of March 21, 2011: http://www.nber.org/papers/w16850

Goldhaber, Dan, "Exploring the Potential of Value-Added Performance Measures to Affect the Quality of the Teacher Workforce," *Educational Researcher*, Vol. 44, No. 2, March 2015, pp. 87–95.

Goldring, Ellen, Jason A. Grissom, Mollie Rubin, Christine M. Neumerski, Marisa Cannata, Timothy Drake, and Patrick Schuermann, "Make Room Value Added: Principals' Human Capital Decisions and the Emergence of Teacher Observation Data," *Educational Researcher*, Vol. 44, No. 2, March 2015, pp. 96–104.

Hamilton, Laura S., Elizabeth D. Steiner, Deborah Holtzman, Eleanor S. Fulbeck, Abby Robyn, Jeffrey Poirier, and Caitlin O'Neil, *Using Teacher Evaluation Data to Inform Professional Development in the Intensive Partnership Sites*, Santa Monica, Calif.: RAND Corporation, WR-1033-BMGF, 2014. As of September 30, 2015: http://www.rand.org/pubs/working_papers/WR1033.html

Jiang, Jennie Y., Susan E. Sporte, and Stuart Luppescu, "Teacher Perspectives on Evaluation Reform: Chicago's REACH Students," *Educational Researcher*, Vol. 44, No. 2, March 2015, pp. 105–116.

Kane, Thomas J., and Douglas O. Staiger, *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*, Bill & Melinda Gates Foundation, January 2012. As of May 22, 2015: http://www.metproject.org/downloads/MET_Gathering_Feedback_Practioner_Brief.pdf

Marsh, Julie A., Matthew G. Springer, Daniel F. McCaffrey, Kun Yuan, Scott Epstein, Julia Koppich, Nidhi Kalra, Catherine DiMartino, and Art Peng, *A Big Apple for Educators: New York City's Experiment with Schoolwide Performance Bonuses—Final Evaluation Report*, Santa Monica, Calif.: RAND Corporation, MG-1114-FPS, 2011. As of September 30, 2015: http://www.rand.org/pubs/monographs/MG1114.html

Master, Benjamin, "Staffing for Success: Linking Teacher Evaluation and School Personnel Management in Practice," *Educational Evaluation and Policy Analysis*, Vol. 36, No. 2, 2014, pp. 207–227.

National Council on Teacher Quality, "NCTQ State Policy: Policy Issues," undated. As of August 17, 2015: http://www.nctq.org/statePolicy/2014/statePolicyIssues.do

Rotherham, Andrew J., and Ashley LiBetti Mitchel, *Genuine Progress, Greater Challenges: A Decade of Teacher Effectiveness Reforms*, Boston, Mass.: Bellwether Education Partners, May 2014. As of September 30, 2015: http://files.eric.ed.gov/fulltext/ED545140.pdf

Schmidt, Michèle, and Amanda Datnow, "Teachers' Sense-Making About Comprehensive School Reform: The Influence of Emotions," *Teaching and Teacher Education*, Vol. 21, No. 8, November 2005, pp. 949–965.

Spillane, James P., Brian J. Reiser, and Todd Reimer, "Policy Implementation and Cognition: Reframing and Refocusing Implementation Research," *Review of Educational Research*, Vol. 72, No. 3, Autumn 2002, pp. 387–431.

Springer, Matthew G., Dale Ballou, Laura S. Hamilton, Vi-Nhuan Le, J. R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher, *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*, Nashville, Tenn.: National Center on Performance Incentives, Vanderbilt University, 2010. As of September 30, 2015: http://www.rand.org/pubs/reprints/RP1416.html

Springer, Matthew G., John F. Pane, Vi-Nhuan Le, Daniel F. McCaffrey, Susan Freeman Burns, Laura S. Hamilton, and Brian Stecher, "Team Pay for Performance: Experimental Evidence from the Round Rock Pilot Project on Team Incentives," *Educational Evaluation and Policy Analysis*, Vol. 34, No. 4, December 2012, pp. 367–390.

Stecher, Brian M., and Michael Garet, *Introduction to the Evaluation of the Intensive Partnership for Effective Teaching (IP)*, Santa Monica, Calif.: RAND Corporation, WR-1034-BMGF, 2014. As of September 30, 2015: http://www.rand.org/pubs/working_papers/WR1034.html

Stecher, Brian M., Michael Garet, Laura S. Hamilton, Elizabeth D. Steiner, Abby Robyn, Jeffrey Poirier, Deborah Holtzman, Eleanor S. Fulbeck, Jay Chambers, and Iliana Brodziak de los Reyes, *Improving Teaching Effectiveness: The Intensive Partnerships for Effective Teaching After Five Years, Appendixes D and E*, Santa Monica, Calif.: RAND Corporation, RR-1295/2-BMGF, forthcoming.

Stecher, Brian, Mike Garet, Deborah Holtzman, and Laura Hamilton, "Implementing Measures of Teacher Effectiveness," *Phi Delta Kappan*, Vol. 94, No. 3, November 2012, pp. 39–43.

Steele, Jennifer L., Matthew D. Baird, John Engberg, and Gerald Hunter, *Trends in the Distribution of Teacher Effectiveness in the Intensive Partnerships for Effective Teaching*, Santa Monica, Calif.: RAND Corporation, WR-1036-BMGF, 2014. As of September 30, 2015: http://www.rand.org/pubs/working_papers/WR1036.html

Steinberg, Matthew P., and Lauren Sartain, "Does Better Observation Make Better Teachers?" *Education Next*, Vol. 15, No. 1, Winter 2015, pp. 71–76. As of September 30, 2015:
http://educationnext.org/better-observation-make-better-teachers/

Students First, *State of Education: State Policy Report Card, 2013*, Sacramento, Calif., 2013.

Taylor, Eric S., and John H. Tyler, "The Effect of Evaluation on Teacher Performance," *American Economic Review*, Vol. 102, No. 7, December 2012, pp. 3628–3651.

Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, and David Keeling, *The Widget Effect: Our National Failure to Acknowledge and Act on Difference in Teacher Effectiveness*, 2nd ed., Brooklyn, N.Y.: New Teacher Project, June 8, 2009. As of September 30, 2015:
http://tntp.org/publications/view/evaluation-and-development/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness

Whitney, Myra, Tishsha Hopson, Lytania Black, and Charles New, *Resource Book*, Memphis City Schools, September 27, 2011. As of October 19, 2015:
http://www.scribd.com/doc/66591232/TEM-Aligned-PD-Resource-Guide#

Xu, Zeyu, Umut Özek, and Matthew Corritore, *Portability of Teacher Effectiveness Across School Settings*, Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research, Working Paper 77, June 2012. As of September 30, 2015:
http://files.eric.ed.gov/fulltext/ED533217.pdf

Xu, Zeyu, Umut Özek, and Michael Hansen, *Teacher Performance Trajectories in High and Lower-Poverty Schools*, Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research, American Institutes for Research, Working Paper 101, updated March 2014.

Yuan, Kun, Vi-Nhuan Le, Daniel F. McCaffrey, Julie A. Marsh, Laura S. Hamilton, Brian M. Stecher, and Matthew G. Springer, "Incentive Pay Programs Do Not Affect Teacher Motivation or Reported Practices: Results from Three Randomized Studies," *Educational Evaluation and Policy Analysis*, Vol. 35, No. 1, March 2013, pp. 3–22.

Zamarro, Gema, John Engberg, Juan Esteban Saavedra, and Jennifer Steele, "Disentangling Disadvantage: Can We Distinguish Good Teaching from Classroom Composition?" *Journal of Research on Educational Effectiveness*, Vol. 8, No. 1, 2015, pp. 84–111.

To improve the U.S. education system through more-effective classroom teaching, in school year 2009–2010, the Bill and Melinda Gates Foundation announced four Intensive Partnership for Effective Teaching sites. The Intensive Partnerships Initiative is based on the premise that efforts to improve instruction can benefit from high-quality measures of teaching effectiveness. The initiative seeks to determine whether a school can implement a high-quality measure of teaching effectiveness and use it to support and manage teachers in ways that improve student outcomes. This approach is consistent with broader national trends in which performance-based teacher evaluation is increasingly being mandated at state and local levels.

To test the theory in practice, the foundation sought partnership sites. It selected three school districts—Hillsborough County Public Schools in Florida, Shelby County Schools in Tennessee, and Pittsburgh Public Schools in Pennsylvania. The foundation also selected four charter management organizations—Alliance College-Ready Public Schools, Aspire Public Schools, Green Dot Public Schools, and the Partnerships to Uplift Communities, all in California.

To evaluate Intensive Partnership implementation, researchers from the RAND Corporation and the American Institutes for Research interviewed annually central-office staff at each site and teachers and other staff in a sample of schools for each site. They also used data from annual teacher and school-leader surveys and documents that the sites and the foundation provided. This report summarizes the implementation status of key reform elements at each site when the Intensive Partnerships initiative launched and five years later in the spring of 2014.

EDUCATION

www.rand.org

$29.50

ISBN-10 0-8330-9221-9
ISBN-13 978-0-8330-9221-2

RR-1295-BMGF