

2015

# THE MIRAGE

Confronting the Hard Truth About Our Quest for Teacher Development

# THE MIRAGE

Confronting the Hard Truth About Our Quest for Teacher Development

*The Mirage describes the widely held perception among education leaders that we already know how to help teachers improve, and that we could achieve our goal of great teaching in far more classrooms if we just applied what we know more widely. Our research suggests that despite enormous and admirable investments of time and money, we are much further from that goal than has been acknowledged, and the evidence base for what actually helps teachers improve is very thin.*

**01 | EXECUTIVE SUMMARY**

**04 | SCOPE AND METHODOLOGY**

**06 | WHAT WE LEARNED**



**34 | THE WAY FORWARD**

**35 | RECOMMENDATIONS**

**40 | TECHNICAL APPENDIX**

**58 | ENDNOTES**



DO WE KNOW  
HOW TO HELP  
TEACHERS  
GET BETTER?

# EXECUTIVE SUMMARY

*Do we know how to help teachers get better?*

It's a critical question: By helping more teachers succeed in the classroom, we could put more students on the path to success.<sup>1</sup> For decades, conventional wisdom has been that if we could just get teachers the right type and right amount of support, educational excellence would be right around the corner. Just how to support teachers has been the preoccupation of school systems and organizations like ours, as well as the subject of countless research studies, op-eds and books.

Most discussions about teacher development presume that we already know the answer. *Of course* we know what good professional development looks like; we just haven't been able to do it at scale for all teachers, yet.<sup>2</sup>

We thought so, too. Two years ago, we embarked on an ambitious effort to identify what works in fostering widespread teacher improvement. Our research spanned three large public school districts and one midsize charter school network.

We surveyed more than 10,000 teachers and 500 school leaders and interviewed more than 100 staff members involved in teacher development.

Rather than test specific strategies to see if they produced results, we used multiple measures of performance to identify teachers who improved substantially, then looked for any experiences or attributes they had in common—from the kind and amount of development activities in which they participated to the qualities of their schools and their mindset about growth—that might distinguish them from teachers who did not improve. We used a broad definition of “professional development” to include efforts carried out by districts, schools and teachers themselves.

In the three districts we studied, which we believe are representative of large public school systems nationwide, we expected to find concentrations of schools where teachers were improving at every stage of their careers, or evidence that particular supports were especially helpful in boosting teachers’ growth.

After an exhaustive search, we were disappointed not to find what we hoped we would. Instead, what we found challenged our assumptions.



## FINDINGS

### Districts are making a massive investment in teacher improvement—far larger than most people realize.

We estimate that the districts we studied spend an average of nearly \$18,000 per teacher, per year on development efforts.<sup>3</sup> One district spends more on teacher development than on transportation, food and security combined.<sup>4</sup>

At this rate, the largest 50 school districts in the U.S. devote at least \$8 billion to teacher development annually.<sup>5</sup> Furthermore, the teachers we surveyed reported spending approximately 19 full school days a year—nearly 10 percent of a typical school year—participating in development activities. After a little more than a decade in the classroom, an average teacher will have spent the equivalent of more than a full school year on development.<sup>6</sup> This represents an extraordinary and generally unrecognized commitment to supporting teachers' professional growth as the primary strategy for accelerating student learning.

### Despite these efforts, most teachers do not appear to improve substantially from year to year—even though many have not yet mastered critical skills.

Across the districts we studied, the evaluation ratings of nearly seven out of 10 teachers remained constant or declined over the last two to three years.<sup>7</sup> Substantial improvement seems especially difficult to achieve after a teacher's first few years in the classroom; the difference in performance between an average first-year teacher and an average fifth-year teacher was more than *nine times* the difference between an average fifth-year teacher and an average twentieth-year teacher.<sup>8</sup> More importantly, many teachers' professional growth plateaus while they still have ample room to improve: As many as half of teachers in their tenth year or beyond were rated below "effective" in core instructional practices, such as developing students' critical thinking skills.<sup>9</sup>

**Even when teachers do improve, we were unable to link their growth to any particular development strategy.** We looked at dozens of variables spanning the development activities teachers experienced, how much time they spent on them, what mindsets they brought to them and even where they worked. Yet we found no

common threads that distinguished "improvers" from other teachers. No type, amount or combination of development activities appears more likely than any other to help teachers improve substantially, including the "job-embedded," "differentiated" variety that we and many others believed to be the most promising.<sup>10</sup>

**School systems are not helping teachers understand how to improve—or even that they have room to improve at all.** Teachers need clear information about their strengths and weaknesses to improve their instruction, but many don't seem to be getting that information. The vast majority of teachers in the districts we studied are rated Effective or Meeting Expectations or higher,<sup>11</sup> even as student outcomes in these districts fall far short of where they need to be. Perhaps it is no surprise, then, that less than half of teachers surveyed agreed they had weaknesses in their instruction.<sup>12</sup> Even the few teachers who did earn low ratings seemed to reject them; more than 60 percent of low-rated teachers still gave themselves high performance ratings.<sup>13</sup> Together, this suggests a pervasive culture of low expectations for teacher development and performance. These low expectations extended to teachers' satisfaction with the development they received. While two-thirds reported feeling relatively satisfied with their development experiences,<sup>14</sup> only about 40 percent reported that most of their professional development activities were a good use of their time.<sup>15</sup>

---

In short, we bombard teachers with help, but most of it is not helpful—to teachers as professionals or to schools seeking better instruction. We are not the first to say this: In the last decade, two federally funded experimental studies of sustained, content-focused and job-embedded professional development have found that these interventions did not result in long-lasting, significant changes in teacher practice or student outcomes.<sup>16</sup> And while countless other studies have been undertaken, researchers summarize the evidence base as weak and the results mixed at best.<sup>17</sup>

In spite of this, the notion persists that we know how to help teachers improve and could achieve our goal of great teaching in far more classrooms if we just applied that knowledge more widely. It's a hopeful and alluring vision, but our findings force us to conclude that it is a mirage. Like a mirage, it is not a hallucination but a refraction of reality: Growth is possible, but our goal of widespread teaching excellence is further out of reach than it seems.

Great teaching is very real, as are teachers who improve over time, sometimes dramatically so. Undoubtedly, there are development experiences that support that improvement. But we found no clear patterns in these success stories and no evidence that they were the result of deliberate, systemic efforts. Teacher development appears to be a highly individualized process, one that has been dramatically oversimplified. The absence of common threads challenges us to confront the true nature of the problem—that as much as we wish we knew how to help all teachers improve, we do not.

We say this with humility. In the course of our own work over the last two decades, we have made the same assumptions, missteps and miscalculations as the districts we studied. It is this experience that drives us to do better and urge others to do the same.

We believe it's time to take a step back in our pursuit of teacher improvement and acknowledge just how far we stand from the goal of great teaching in every classroom, even as we recommit ourselves to reaching it. We have no excuses—we cannot blame a lack of time, money or good intentions. Instead, we must acknowledge that getting there will take much more than tinkering with the types or amount of professional development teachers receive, or further scaling other aspects of our current approach. It will require a new conversation about teacher development—one that asks fundamentally different questions about what better teaching means and how to achieve it.

## RECOMMENDATIONS

Some may argue that we should drop our investment in teacher development in response to these findings. We disagree. Instead, we believe districts should take a radical step toward upending their approach to helping teachers improve—from redefining what “helping teachers” really means to taking stock of current development efforts to rethinking broader systems for ensuring great teaching for all students. While we found no set of specific development strategies that would result in widespread teacher improvement on its own, there are still clear next steps school systems can take to more effectively help their teachers. Much of this work involves creating the conditions that foster growth, not finding quick-fix professional development solutions. To do this, we recommend that school systems:

### **REDEFINE what it means to help teachers improve**

- Define “development” clearly, as observable, measurable progress toward an ambitious standard for teaching and student learning.
- Give teachers a clear, deep understanding of their own performance and progress.
- Encourage improvement with meaningful rewards and consequences.

### **REEVALUATE existing professional learning supports and programs**

- Inventory current development efforts.
- Start evaluating the effectiveness of all development activities against the new definition of “development.”
- Explore and test alternative approaches to development.
- Reallocate funding for particular activities based on their impact.

### **REINVENT how we support effective teaching at scale**

- Balance investments in development with investments in recruitment, compensation and smart retention.
- Reconstruct the teacher’s job.
- Redesign schools to extend the reach of great teachers.
- Reimagine how we train and certify teachers for the job.

# SCOPE AND METHODOLOGY

Our research included three large, geographically diverse school districts and one midsize charter network (*Figure 1*). The districts we studied collectively employ more than 20,000 teachers with annual operating budgets ranging from \$800 million to \$3 billion.<sup>18</sup> Between them, they serve almost 400,000 students, and on average, 69 percent of those students are low-income.<sup>19</sup>

We gathered information about teacher development in these districts by surveying 10,507 teachers and 566 school leaders, conducting interviews with 127 district staff members and school leaders and hosting smaller focus groups with teachers. We also analyzed professional development catalogs and budget data, as well as several other measures such as session attendance and district-provided coaching data. In this report, we focus primarily on what we learned from the three participating school districts, which we believe are representative of large public school districts across the country. We examine the experiences and growth of teachers in the charter school network in greater detail on page 30.

Unlike most research on professional development, our method was not to implement a particular development strategy and then track its results. Instead, we identified teachers whose performance appeared to improve substantially and worked backward to find any experiences, mindsets or environments they had in common, in contrast to those teachers whose performance did not improve substantially.

Each of the districts we studied implemented a multiple-measure evaluation system several years ago. We looked at two to four years of teacher performance data for each participating district, which allowed us to track the improvement of individual teachers from year to year and link them to our survey results about development experiences. Recognizing the inherent limitations of any single effectiveness measure, we chose to track growth over multiple measures: summative evaluation ratings, classroom observation scores (including component domain sub-scores) and value-added scores. By looking across several performance outcomes, we were able to test how consistently teachers' experiences, mindsets and environments were related to their performance, and compare how these factors were related to various measures. And though the near majority of teachers in each of the three districts received summative evaluation ratings in the top two categories, there was still variation in both raw evaluation scores and in observation component scores. This allowed us to differentiate between teachers based on performance level and growth over time.

We identified teachers who improved meaningfully using multiple definitions of growth across multiple measures of effectiveness. Beyond simply looking at changes in individual performance measures, we looked for teachers who grew more than their peers with similar experience and who started off at the same level of performance. We also grouped teachers into quartiles, assessing who was making the most and least growth over a two- to three-year period. Our goal was to find as many teachers as possible who seemed to have improved their instruction substantially so that we could assess differences between improvers and other teachers.

We tried to capture the full extent of how teachers spent time on their development over a two-year period. We asked them about a broad range of activities: traditional one-time professional development, extended development programs, independent teacher efforts, formal and informal peer collaboration, receiving direct coaching, completing university coursework, time with a formal evaluator, peer observations, administrator observations, feedback, technique practice, follow-up support and new teacher preparation and mentoring. We also collected feedback on these experiences from teachers and principals, allowing us to look at individual teacher mindsets and reactions, school leader reactions and the collective responses from teachers working in the same school.<sup>20</sup>

To calculate total spending on efforts to improve teacher practice, we chose not to focus only on straightforward "professional development" line items that surface in some district financial documents. Instead, we sought to understand the staff time and resources that are intended to improve instruction, either directly or indirectly. To do this, we looked at a range of resources, including line-item budgets and personnel data, financial and policy documents, teachers' contracts and interviews with district staff members and school leaders. We generated estimates on a sliding scale of three tiers, with the lowest tier representative of more traditional development activities and the higher tiers representative of more strategic investments, such as teacher evaluation and rewards for attaining higher levels of effectiveness. (*For more detailed information on our methodology, see the Technical Appendix, p. 40.*)

FIGURE 1 | OVERVIEW OF STUDY METHODOLOGY



We used multiple measures of performance to identify teachers who improved substantially.

We looked for any experiences or attributes in common, including professional development activities, mindset and school.

We studied three large districts and one charter management organization for a total of 20,000 teachers and 400,000 students.

We expected to find evidence that teachers who improve share experiences or mindsets that set them apart from teachers who don't improve. We found that it's just not that simple.



**WHAT  
WE  
LEARNED**

## THE INVESTMENT

What is the current investment in teacher development across the districts we studied?

## THE RESULTS

How much do teachers improve their performance over time and what distinguishes teachers who improve from those who don't?

## TEACHERS' PERSPECTIVES

What do teachers make of their own professional growth and their experiences with the system that's trying to support it?

# 1. THE INVESTMENT

## SCHOOL DISTRICTS ARE MAKING A MASSIVE INVESTMENT IN TEACHER DEVELOPMENT

Conventional wisdom suggests that school districts underinvest in supporting their teachers. But in the districts we studied, we found a consistently *huge* commitment to teacher improvement—much larger than most people probably realize and far exceeding what other industries spend on comparable efforts. When we look at the resources allocated to help teachers improve, including time and money toward training, mentoring, evaluating and providing ongoing job-embedded experiences, we calculate that the districts we studied spend an average of nearly \$18,000 per teacher, per year<sup>21</sup>—the equivalent of 6 to 9 percent of their annual operating budgets.<sup>22</sup> Based on those estimates, we project that the 50 largest school districts in the U.S. likely spend a combined \$8 billion every year on teacher development.<sup>23</sup> Teachers devote an enormous amount of time to their development, too: according to our survey results, approximately 150 hours a year, or nearly 10 percent of a typical school year.<sup>24</sup>

The districts we studied spend an average of nearly \$18,000 per teacher, per year on teacher development.

## Staff Support

Districts' commitment to helping teachers improve is perhaps most visible in the sheer number of staff spending significant amounts of their time supporting development. In *addition* to principals and assistant principals, for every 14 to 37 teachers across the districts we studied, there is one full-time equivalent staff member directly supporting teachers. These positions include coaches, instructional and curriculum specialists, professional learning community (PLC) leaders, teacher evaluation staff and more.

All told, in the districts we studied, we estimate that as many as 10 people or central departments can play a role in a single teacher's development. A teacher might be working with her school leadership, a curriculum specialist, an instructional coach and the district's professional development staff, just for starters.

## Teacher Time

Teachers are making a significant investment in their own development in the form of their time. In the districts studied, teachers reported spending an average of 17 hours per month on development activities run by their district or school, or those that are self-initiated. That comes to almost 150 hours per year—the equivalent of 19 school days, or nearly 10 percent of a typical school year.<sup>25</sup> If we consider *only* time mandated directly by district policy through development days and release time set aside for teacher improvement efforts, the time ranges from 39 to 74 hours per school year.

This investment of time seems to continue as long as teachers remain in the classroom. While new teachers we surveyed reported spending substantively more time on instructional coaching (13 hours per year) compared to their more experienced peers (5 hours), after a teacher's second year, that difference becomes much more negligible, with teachers at all levels of experience reporting about the same time spent on coaching.<sup>26</sup> Among other development activities, like extended professional development workshops, formal collaboration efforts and peer observations, the differences were equally minimal.<sup>27</sup>

The time adds up. In a little over a decade in the classroom, the average teacher in the districts we studied would have spent the equivalent of more than an entire school year (198 days) on their development, in some form or fashion.<sup>28</sup>

## Professional Learning Experiences

School districts have also built enormous catalogs of workshops and courses for their teachers in an effort to give them a wide variety of learning opportunities. The largest district we studied offered more than 1,000 professional learning courses during the 2013-14 school year.<sup>29</sup>

These offerings take place largely during the school year (although the districts we studied offer additional summer opportunities as well). Programming for new teachers and teachers new to the district (but with some prior teaching experience) is also offered at the start of the school year. Throughout the year, the schools all commit several days to district-wide professional development, in addition to time for school-specific professional development. And they devote time to various types of formal collaboration through venues like PLCs, with additional time earmarked for teachers to work as a whole team or in smaller groups.

**After a little over a decade  
in the classroom, the average  
teacher in the districts we  
studied would have spent the  
equivalent of more than an  
entire school year on  
professional development.**

## Estimating the Full Cost

We calculated the amount of money these districts invest in teacher improvement on a sliding scale. In the low range, we considered only the baseline costs associated with improving teacher practice, including the cost of time spent on direct support at the central office and school levels, materials and supplies for professional development, contracts with vendors, the cost of teacher time spent on professional development days, and formal collaboration and stipends for development activities.

In the middle range, we considered all of that plus other spending directly aligned to districts' strategies to help teachers improve their instruction, including evaluation systems, time and resources that indirectly support teachers at the central office and school levels, additional time teachers reported spending on coaching and peer observations, and investments in teachers' salaries for degree attainment.

Finally, in the high range, we accounted for all of the strategic investments one could argue should be considered teacher support spending, such as salary incentives for improved performance, the costs of instructional leadership development activities and select data strategy expenditures.

Using the mid-range estimates, the districts we studied spend between roughly \$73 million and \$181 million on teacher improvement annually (*Figure 2*).<sup>30</sup> That works out to between 6 and 9 percent of their annual budgets, or an average of \$18,000 per teacher, per year.<sup>31</sup> Even using only the low-range estimates, the districts we studied spend more on professional development than they do on other big-ticket items, like food services (an average of 4 percent across our districts, with a range of 3 to 5 percent) or transportation (an average of 1 percent, with a range of 0.04 to 2 percent).<sup>32</sup>

By a wide margin, the largest piece of that investment is in the salaries and other costs related to teachers and the hundreds of people who provide instructional support at all levels of each district. Across districts, between 77 and 87 percent of the estimated mid-range costs are related to teacher and staff time and salaries.<sup>33</sup>

Our analysis indicates that the investment these school districts are making in helping their teachers improve is massive. In fact, it far exceeds what other industries spend on support and development for their practitioners.<sup>34</sup>

For example, the average large government/military organization (defined as 10,000 employees or more) spent a little more than \$2 million on staff training in 2013.<sup>35</sup> By comparison, a school district we studied, with a similar number of teaching staff, spent more than \$90 million on teacher training and support in the same time period, excluding the costs of teachers' salaries for the time they spent in training, additional investments like salary bumps for improved performance and school leader time beyond meeting directly with teachers for support. Even using this more conservative estimate, on average, the districts we studied spent anywhere from nearly two to four times more of their budgets and four to nearly 15 times more per employee on support and development, compared to other industries.<sup>36</sup>

To be clear, an outsized investment in teacher support is not necessarily unwise or unmerited; after all, if teacher improvement were achieved at scale, it would have an enormous effect on students. The problem is our indifference to its impact—that all this help doesn't appear to be helping all that much.

**An outsized investment in teacher improvement is not necessarily unwise or unmerited. The problem is our indifference to its impact—that all this help doesn't appear to be helping all that much.**

FIGURE 2 | ESTIMATED TEACHER IMPROVEMENT SPENDING FOR DISTRICTS, FY 2014

	LOW	MEDIUM	HIGH
DISTRICT A	Total cost of teacher improvement <b>\$151 million</b>	<b>\$181 million</b>	<b>\$196 million</b>
	Percent of FY 2014 budget <b>5%</b>	<b>6%</b>	<b>6%</b>
	Cost per teacher <b>\$13,004</b>	<b>\$15,535</b>	<b>\$16,804</b>
DISTRICT B	Total cost of teacher improvement <b>\$50 million</b>	<b>\$73 million</b>	<b>\$91 million</b>
	Percent of FY 2014 budget <b>6%</b>	<b>9%</b>	<b>11%</b>
	Cost per teacher <b>\$14,232</b>	<b>\$20,886</b>	<b>\$25,914</b>
DISTRICT C	Total cost of teacher improvement <b>\$90 million</b>	<b>\$146 million</b>	<b>\$164 million</b>
	Percent of FY 2014 budget <b>6%</b>	<b>9%</b>	<b>10%</b>
	Cost per teacher <b>\$10,558</b>	<b>\$17,014</b>	<b>\$19,133</b>

Districts are making a massive investment in teacher improvement—far larger than most people realize.

## 2. THE RESULTS

### MOST TEACHERS WE STUDIED DO NOT APPEAR TO BE IMPROVING SUBSTANTIALLY FROM YEAR TO YEAR

The school districts we studied are dedicating extraordinary resources and time to help teachers get better, demonstrating a commitment to teacher support that is essential, laudable and generally unacknowledged. As a result, we would all hope to see evidence that most teachers are making substantial improvements over time and consistently reaching a level of mastery over core instructional techniques before their growth levels off. We would also hope to see relationships between districts' teacher development efforts and evidence of substantial teacher improvement.

By these standards, however, the teacher development efforts in the districts we studied are falling short. Most teachers' performance does not appear to improve substantially from year to year, especially after their first few years in the classroom. Too many peak before they master core instructional skills. And when teachers do improve by leaps and bounds, we could not trace that growth to systemic development strategies.

## Marching in Place

Most teachers in the districts we studied seem to be marching in place when it comes to their development. While they may be making small progress here or there, they ultimately end up in basically the same place, year after year. And while some do make meaningful improvement—the kind that results in observably better teaching or improved student learning—it is too rare. Consider this: Across the districts we studied, only three out of every 10 teachers tended to improve their performance substantially over the years studied, as measured by their overall evaluation scores (*Figure 3*).<sup>37</sup> Of the remaining teachers, five maintained relatively the same level of performance, while two actually saw their performance decline substantially, over a two- to three-year period.

In the districts we studied, average performance scores on evaluations and observations remained generally constant from year to year, with little—if any—meaningful movement forward.<sup>38</sup>

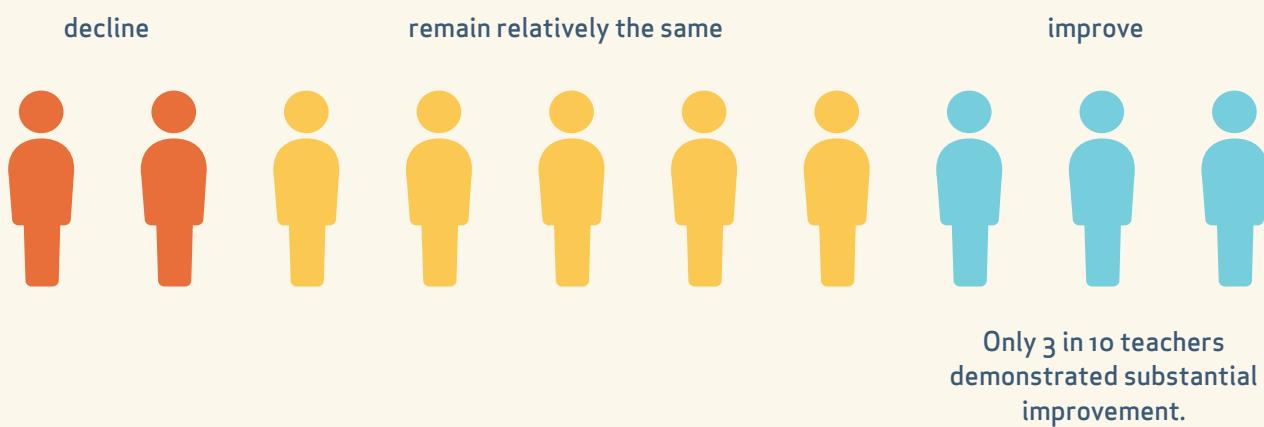
Similar patterns of limited progress hold when we look closely at individual teachers' performance over time on specific instructional skills rated in classroom observations. In the 2011-12 school year, for example, more than 1,200 teachers in one district earned a rating below “effective” on how well they develop students’ critical thinking skills. Two years later, nearly two-thirds of those teachers had still not earned a rating of “effective” on that skill strand.

In another district, of all the teachers who earned a low rating in 2011-12 for their ability to engage students, 28 percent of those who remained in the district two years later hadn’t improved in this area at all. Another 43 percent improved only enough to earn a “developing” rating instead of the lowest one. Only 26 percent had improved enough to become at least “effective” at this skill.<sup>39</sup>

All together, these patterns indicate just how difficult substantial improvement can be to attain, especially on the skills students need most for academic success.<sup>40</sup>

**FIGURE 3 | AVERAGE CHANGE IN PERFORMANCE ON EVALUATIONS**

Over several years, teachers saw their scores:



**Most teachers are marching in place—and some are even seeing their performance decline.**

## Rapid Growth... At First

Substantial improvement seems especially difficult to achieve after a teacher's first few years in the classroom. Most teachers in the districts we studied *did* improve substantially during these early years<sup>41</sup>—a well-established pattern that has been documented by many researchers and reflects a natural learning curve.<sup>42</sup>

But that's where the meaningful improvement for the average teacher seems to end. *Figure 4* illustrates the overall pattern of teacher growth we saw in these three districts. It tracks the average rates of change in teacher performance at different levels of experience over time. Teacher improvement here is measured using districts' evaluation tools, all of which rely on multiple measures, including the results of classroom observations, student assessment data and measures of professionalism.<sup>43</sup>

Teachers in their first five years grew at least two and a half to five times faster than all other teachers across the districts studied, over the last three years. After their fifth year of teaching, the average teacher grew even less, and the average teacher in their tenth year or beyond has a growth rate barely above zero.

This trend also held true even when we looked at the individual measures that feed into overall evaluation results. Looking only at the change in teachers' classroom observation scores across several years, for example, we found again that the highest rates of growth were consistently achieved by teachers in their first five years. The same is true of value-added scores.<sup>44</sup>

## A Low Plateau

Decreased growth over the course of the average teacher's career might not be a problem if it occurred after most had mastered core instructional techniques. Unfortunately, that does not appear to be the case in the districts we studied. The overall pattern of rapid growth that wanes after the early years results in a performance plateau that occurs where most teachers—and their students—still have room to improve.

Many studies, largely relying on value-added data, have shown that natural “returns to experience” slow down or even plateau after teachers’ first several years in the classroom.<sup>45</sup>

We found a similar plateau in average teacher effectiveness scores on various measures for teachers at different experience levels. Average teacher performance increases dramatically among teachers in their early years, but then tends to level off among teachers in the later experience

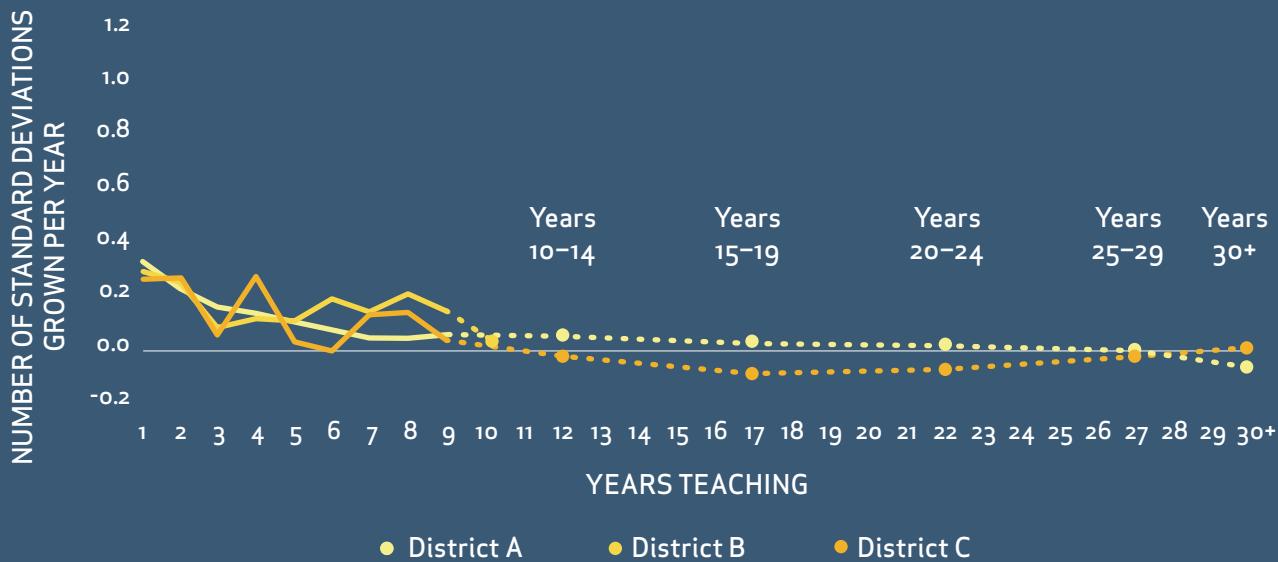
bands. For example, the difference in performance between first- and fifth-year teachers is around nine times as much as the difference in performance between fifth- and twentieth-year teachers (*Figure 5*).<sup>46</sup> Here, too, we saw this trend repeated across observations and value-added data.

But as growth wanes and performance plateaus, we found that many teachers are still struggling to become effective in key skills, even in their tenth year and beyond in the classroom.

For example, across the districts we studied, nearly half of all teachers in their tenth year or beyond earned less than an “effective” rating in developing students’ critical thinking skills—an essential instructional skill for successfully transitioning to the Common Core State Standards<sup>47</sup>—while between 29 and 46 percent of all those teachers struggled with engaging students in lessons (*Figure 6*).<sup>48</sup>

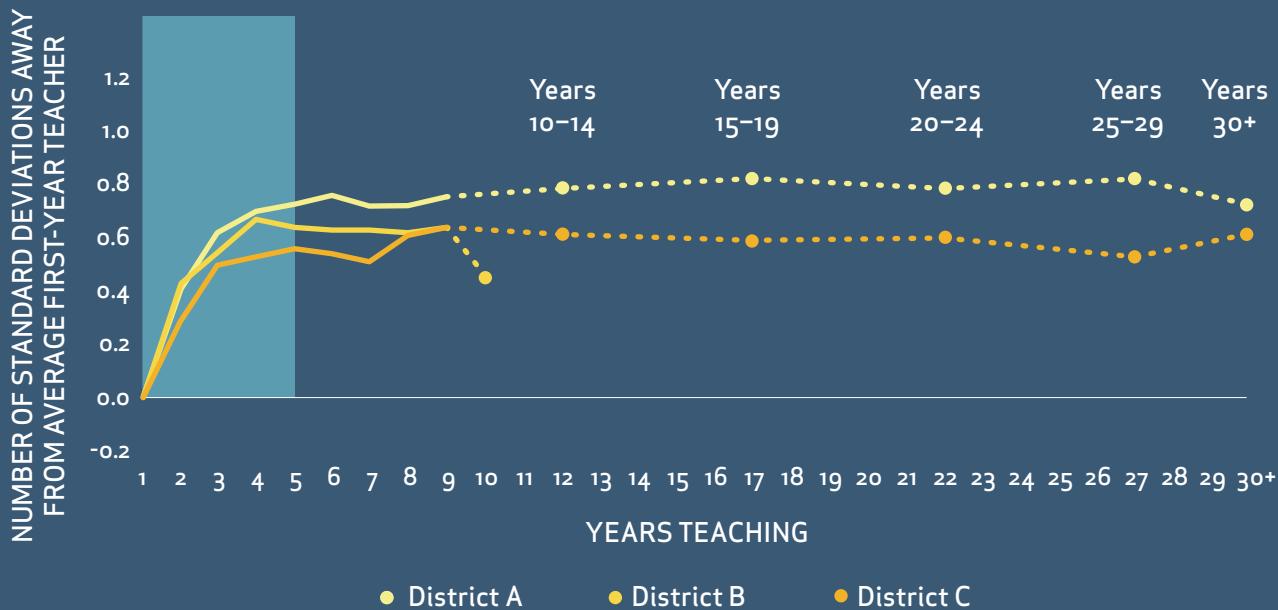
**Fig. 4 & 5 Note:** Because of sample size restrictions, we grouped teachers into 5-year experience bands starting with 10th-14th year teachers and ending at all teachers in their 30th year or beyond. In the figures on p. 15, the large dots represent the average growth rate (*Figure 4*) or performance (*Figure 5*) for all teachers in these experience bands. The lines connecting these points are dashed to signify that we did not look at averages in each of these experience years individually. For District B only, because of how we obtained experience information, we must group all teachers in their 10th year or more of teaching into a single group. This last point in District B represents the growth rate or performance of all teachers with at least 10 years of teaching experience.

FIGURE 4 | AVERAGE GROWTH RATE ON EVALUATION SCORES BY EXPERIENCE



Growth in teacher performance levels off after the early years.

FIGURE 5 | AVERAGE TEACHER PERFORMANCE BY EXPERIENCE



The average fifth-year teacher's performance looks very similar to the average teacher's performance after 10 or 15 years.

FIGURE 6 | EXPERIENCED TEACHERS RATED BELOW EFFECTIVE ON CORE INSTRUCTIONAL SKILLS

Teachers in their tenth year and beyond were rated **below effective** in:

**46-53%**

**29-46%**

**20-42%**

developing critical thinking skills

engaging students in lessons

checking for understanding

**Too often, teachers plateau before they master core instructional skills.**

And unfortunately, it is likely almost impossible for the average teacher to become “highly effective” in some key instructional skills, based on current growth rates. In one district, for example, it would take the average teacher 31 years—potentially an entire career—to become “highly effective” at developing students’ higher-level understanding; it would take 33 years for the average teacher in another district to do so. And for a teacher in another district in their sixth year of teaching or beyond, it would be nearly impossible to reach “highly effective” in skills like using questioning and discussion techniques and designing student assessments.<sup>49</sup>

This is a level of performance where most teachers still have plenty of room for improvement. It’s a level at which students are falling short of expectations, too. Across the districts we studied, around half of students (or less) were proficient against state standards in math and reading, with no district exceeding 56 percent proficiency in either subject.<sup>50</sup> Teachers are certainly not the only factor in these results, but improving the quality of instruction students receive is one of the most important things districts could do to change them. For example, in one district, where

some teachers in their sixth to ninth years of experience had substantially better evaluation outcomes than their peers, their students had better outcomes, too: In the classrooms of above-average teachers, 72 percent and 67 percent of students achieved proficiency in math and reading, respectively. Among average teachers, just 63 percent and 53 percent of students did so.<sup>51</sup>

**In one district, in the classrooms of above-average teachers, 72 percent and 67 percent of students achieved proficiency in math and reading, respectively. Among average teachers, just 63 percent and 53 percent of students did so.**

# A REASONABLE BAR FOR TEACHER PERFORMANCE?

We've established that many experienced teachers—in some cases, nearly half of all teachers in their tenth year and beyond across the districts we studied—receive ratings that indicate there is still room for improvement in core instructional skills. So it's reasonable to ask what "effective" and "highly effective" teaching practices look like in these skill areas. What must a teacher be able to do in order to earn these ratings? Are we holding teachers to an unrealistic bar?

Because each district's observation rubric is different, there's no universal agreement about exactly what "effective" and "highly effective" teaching look like for particular instructional skills. But the districts we studied have clear commonalities. In order to earn a rating of "effective" in competencies aligned with developing students' critical thinking skills, for example, teachers need to demonstrate to their observers that they are posing meaningful questions to students, which lead students to critically assess information and

rely on evidence to put forth a point of view. To earn a "highly effective" rating in this same category, teachers must masterfully do so in such a way that all students lead their own conversations and are posing questions to each other. To earn a rating of "effective" at engaging students in lessons, teachers must be able to acknowledge student abilities and create opportunities in response that result in most students being motivated by and equally engaged in appropriately challenging learning tasks. Those rated "highly effective" are able to do the same but for all students, leaving no one behind.

These are complex skills, to be sure. Achieving "effective" instructional practice isn't easy, and achieving "highly effective" practice is that much more challenging. But if we're going to get the results we need for students, teachers need to master these essential skills, and we must assess teacher development efforts by how well they help teachers get there.

## No Clear Pattern to Real Improvement

The plateau we observed in most teachers' performance made us even more interested in studying those teachers who did grow substantially over time, with the hope that they could point the way to a particular approach to professional development that works consistently. Yet, when we found them, we were unable to trace their growth to any particular kind or amount of support their districts were providing.<sup>52</sup> Meaningful improvement, it seems, defies routine; it is a highly individualized process that seems to vary from teacher to teacher. What works for one teacher may not work for another.

We searched for improvers in a variety of ways, ranging from a basic analysis of changes in evaluation scores from year to year to more sophisticated models that identified above-average growth among teachers with the same amount of experience and similar starting levels of performance. Ultimately, between 19 and 30 percent of teachers in the districts we studied met our most rigorous definition of improvers—and these teachers were present in 95 percent of the schools we studied. We also identified a comparison group of teachers who did not improve based on our methods for identifying growth.

What helped some teachers improve when so many of their peers did not? Did they have greater access to particular kinds of interventions? Did they spend more time on one activity or another? Perhaps they brought a different mindset to their professional development or their own growth? Did they work in a particular school or teach a particular subject?

We closely examined how the teachers we surveyed reported spending time on their development over the course of one to two years, assessing dozens of variables across multiple measures of growth and effectiveness. These variables spanned what they did, how much time they spent doing it, what they believe and even where they work. But we were disappointed not to find common threads that meaningfully distinguished improvers from other teachers. When we looked at activities in which improvers participated, as well as their attitudes and beliefs, they seemed more similar to non-improvers than different from them (*Figure 7*).<sup>53</sup>

FIGURE 7 | COMPARISON OF PROFESSIONAL DEVELOPMENT ACTIVITIES AND PERCEPTIONS  
BETWEEN IMPROVERS AND NON-IMPROVERS

## FREQUENCY OF DEVELOPMENT ACTIVITIES

	IMPROVERS	NON-IMPROVERS
Number of times observed over two years	8	7
Hours of coaching over two years	12	13
Hours of formal collaboration over two years	69	64
Hours spent per month in professional development	17	18

## SATISFACTION WITH PROFESSIONAL DEVELOPMENT

“drives lasting improvements to my instructional practice”	52%	48%
“is targeted to support my specific teaching context”	50%	48%
“is a good use of my time”	44%	40%
“is overall satisfactory”	67%	65%

## BELIEFS

Individual teacher is responsible for development	41%	40%
Feedback plays a crucial role in improving teacher practice	79%	74%

Improvers and non-improvers have more in common than not, and improvers are present in 95 percent of the schools we studied.

## HERE'S WHAT WE KNOW

### Improvers, on average, do not report spending more time on their development or on any particular activity.

Conventional wisdom suggests that more professional development is key to teacher improvement, but we found that improvers do not actually experience more of anything. Overall, improvers spend about 17 hours a month on their development, compared to 18 among teachers who did not show evidence of improvement.

Across the 11 kinds of professional development activities we asked about, we found few meaningful differences between the time improvers and other teachers spend on any of them.<sup>54</sup> We even looked at the extreme ends of the time equation—teachers who spent the very most and very least amount of time on particular activities—and found exactly the same trend. For example, 24 percent of all improvers reported the most time spent in extended professional development activities. Meanwhile, 26 percent of improvers also reported the *least* time spent on extended professional development activities.

Even when we looked for teachers who received what many would consider the most support districts can offer, all in conjunction, improvers were no more likely than other teachers to be part of the group. About 14 percent of improvers reported above-average exposure to all of the following: extended professional development activities, formal collaboration, coaching and receiving observations and feedback. But so did about 14 percent of non-improvers.<sup>55</sup>

### Improvers generally were no more satisfied with the development activities they experienced.

Much of the existing research on teacher improvement that informs policy relies on teachers' self-reports of how they changed their practice, or their satisfaction with particular development strategies, as proxies for those strategies' effectiveness.<sup>56</sup> Teacher satisfaction is certainly relevant to consider, but our data suggest that it is unrelated to actual teacher improvement.<sup>57</sup> Sixty-seven percent of improvers and 65 percent of non-improvers reported feeling satisfied with the professional development they received. When we asked teachers if most of the professional development they received was a good use of their time, 44 percent of those who improved said yes, compared to 40 percent of other teachers. And we found few differences between teachers who improved and those who did not when we asked which activity had helped them learn the most (*Figure 8*).<sup>58</sup>

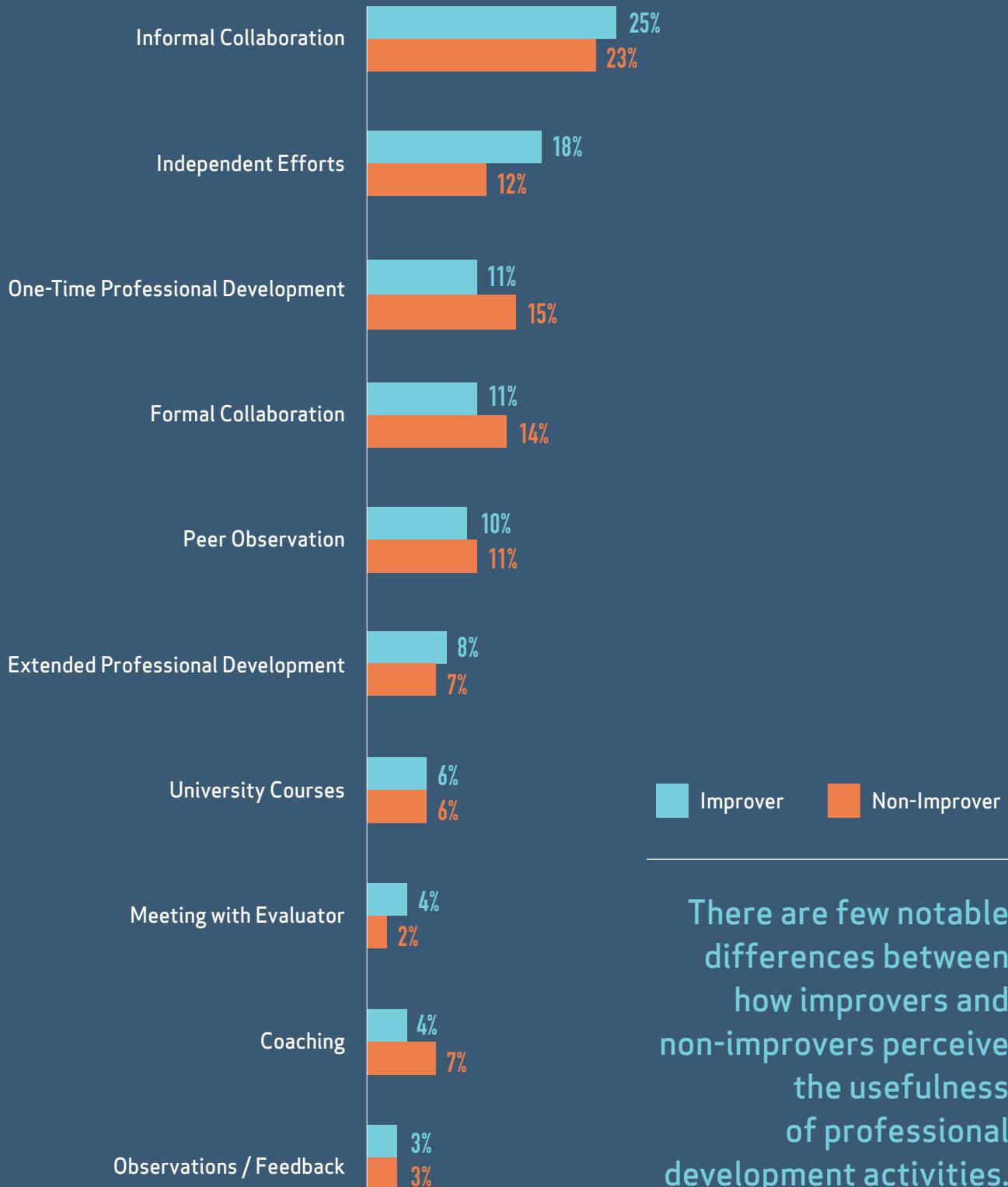
### Improvers do not seem to bring a different mindset to their development.

Improvers reported “reflecting daily on their practice” about as often as teachers who did not demonstrate evidence of improvement. They were as likely as other teachers to feel that they should bear the greatest responsibility for their own development, and they were no more likely to admit they had weaknesses in their instruction (40 percent agreed they had weaknesses, compared to 45 percent of other teachers).

### Improvers are not concentrated in any particular school, school level or subject.

No school in our sample seems to have solved the teacher improvement puzzle more than any other, since we found teachers who improved meaningfully in 95 percent of them. And even among teachers in their sixth year and beyond, we found improvers in 90 percent of schools.<sup>59</sup> These teachers were evenly distributed across subjects, too.<sup>60</sup> And while there is recent research indicating that several school factors can have a positive effect on teachers' improvement, our findings were unable to pinpoint specific drivers at the school level.<sup>61</sup>

FIGURE 8 | TEACHERS REPORTING ON “ACTIVITY THAT HAS HELPED ME LEARN HOW TO IMPROVE THE MOST”



These trends held true at every experience level. Even the rapid growth we saw during teachers' first few years on the job offered no clues for what might sustain that growth later in their careers. New teachers' growth looks consistent across the districts we studied, despite their different approaches to new teacher development.<sup>62</sup> For example, in one district, new teachers spend considerably more time on one-to-one mentoring than do teachers in the other two, but their growth is similar to new teachers' growth elsewhere.<sup>63</sup> And newer teachers who did break the typical growth trajectory for their experience level tended to participate in the same kind and amount of activities as those who did not, just like their more experienced peers.<sup>64</sup>

You can see the development mirage at work in these results. Some teachers really are improving substantially. But in reality, it's impossible to pinpoint a particular type, amount or combination of development activities that is currently helping the average teacher improve more than any other.

Every development strategy, no matter how intensive, seems to be the equivalent of a coin flip: Some teachers will get better and about the same number won't. What separates them may be a host of highly individualized variables or a combination of many we have not yet pinpointed. In practice, though, this means that districts don't have clear direction for how to help any given teacher improve—they are hoping for the best, rather than trying to demonstrate results first and build from that foundation.

**Every development strategy,  
no matter how intensive,  
seems to be the equivalent  
of a coin flip: Some teachers  
will get better and about the  
same number won't.**

# ARE THERE ANY HIDDEN INSIGHTS?

We did find a few consistent, small but statistically significant relationships associated with more teacher improvement on total observation and evaluation scores.<sup>65</sup> As teachers indicate that they are more open to feedback, their scores can be expected to increase modestly. As teachers report feeling more positively about their schools' efforts to help them improve, and as their perceptions of their evaluators improve, their scores can be expected to improve a bit, as well. And when we looked at the school level, we found a small relationship between the number of observations teachers reported receiving and their growth: As the average number of observations at the school increased, the concentration of improvers at that school increased by 2 percent.

The one factor that consistently showed a relationship to teacher growth, across measures and at both the individual teacher level and the school level, was alignment between teachers' perceptions of their instructional effectiveness and their formal evaluation ratings.<sup>66</sup> For example, improvers are almost twice as likely to rate their own performance as the same as their formal evaluation, while non-improvers are almost twice as likely to self-assess their own performance as stronger than their formal ratings.<sup>67</sup>

### 3. TEACHERS' PERSPECTIVES

#### SCHOOL SYSTEMS ARE NOT HELPING TEACHERS UNDERSTAND HOW TO IMPROVE—OR EVEN THAT THEY HAVE ROOM TO IMPROVE AT ALL

Finally, we surveyed teachers about how they experienced these development efforts—and how they view their own performance. It's reasonable to assume that if current teacher improvement efforts were functioning well, most teachers would have an accurate understanding of their instructional strengths and weaknesses, and would be receiving support focused on their particular development areas. Again, however, this does not appear to be the case in the districts we studied. Instead, half of these teachers don't think the help they are receiving is particularly useful for improving their practice, and many have been led to believe they have little room for improvement in the first place.

## Positive Self-Perceptions

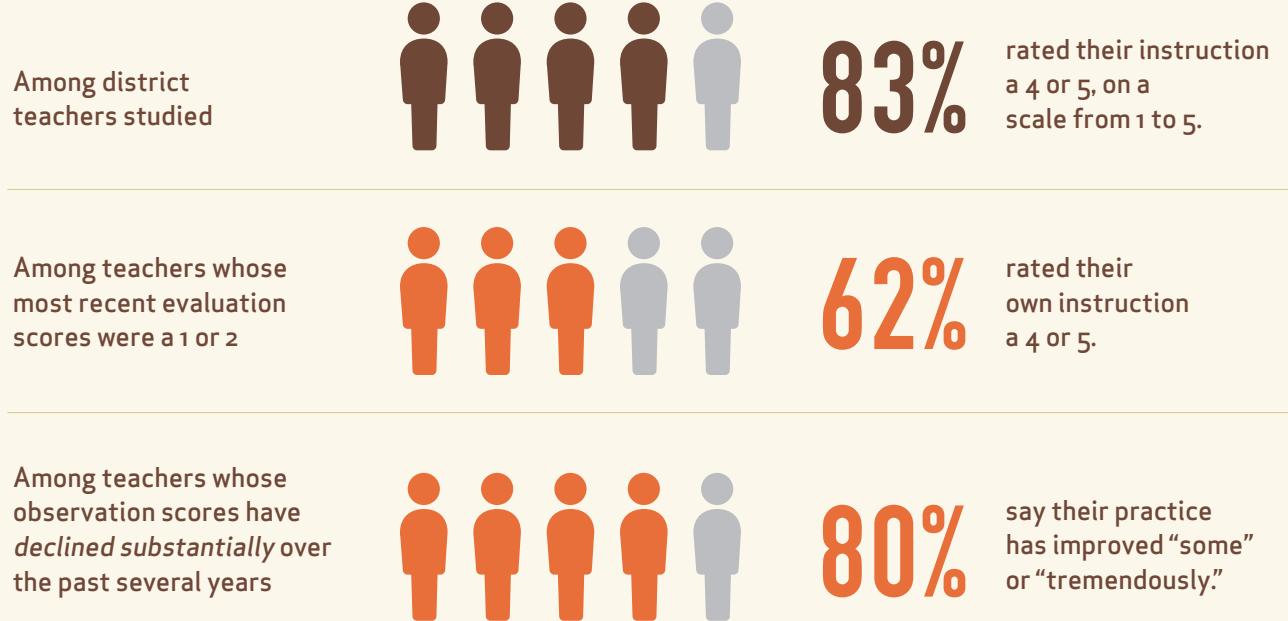
A striking trend that emerged in our survey responses was how differently teachers seem to perceive their performance and growth compared to third-party data. When we asked teachers to rate their own instruction on a five-point scale (with 5 being the highest), more than 80 percent gave themselves a 4 or a 5.<sup>68</sup> Only 47 percent “agreed” or “strongly agreed” that they have weaknesses in their instruction (*Figure 9*).<sup>69</sup> And asked how much their instruction had changed over the last several years, 87 percent of teachers said they had improved “some” or “tremendously.”<sup>70</sup>

Districts themselves are likely a leading cause of these self-ratings in tangible and intangible ways. The vast majority of teachers in these districts are routinely told that there isn’t any need for improvement, through ratings

of Effective or Meeting Expectations—or higher—on their official performance evaluations.<sup>71</sup> Among teachers in their fourth year and beyond, 77 percent to more than 95 percent of teachers in the districts we studied are rated Effective or Meeting Expectations (or better). And so are between 50 and 87 percent of all *brand new* teachers—in other words, they’re being told their instruction is already meeting their district’s expectations.<sup>72</sup>

But even the relatively few teachers who earn low evaluation ratings do not tend to accept them as accurate. Sixty-two percent of low-rated teachers still rated their own instructional practice as a 4 or 5.<sup>73</sup> Among teachers whose scores *declined* on classroom observations over the past several years, four out of five reported that their instruction had improved “some” or “tremendously.”<sup>74</sup>

**FIGURE 9 | TEACHER PERCEPTIONS OF PERFORMANCE AND IMPROVEMENT**



**Less than half of teachers surveyed agree:  
“I have weaknesses in my instruction.”**

## Little Faith in the System

We know that districts are investing in helping teachers improve and asking a lot of teachers in terms of improving their performance, too. But teachers seem skeptical about the usefulness of this support. Only about 40 percent of teachers told us that the majority of the professional development they received was a good use of their time.<sup>75</sup> And only about half felt that most of their development activities provided them with new skills and led to lasting improvements in their instruction.<sup>76</sup> Despite this, around two-thirds of teachers did report general satisfaction with the professional development they had received. This difference between satisfaction and perceived usefulness may be another indication that development efforts can offer teachers tangential benefits beyond actually helping them improve. It may also point to the low expectations for what kind of growth can and should be expected of teachers.<sup>77</sup>

Many teachers' complaints about their professional development appear to stem from a sense that it is not customized to fit their needs. For example, less than half of the teachers we surveyed told us they received professional development that was ongoing, tailored to their specific development needs or even targeted to the students or subject they teach.<sup>78</sup> Differentiation is a basic tenet of good teaching, and perhaps the same principle holds true for teacher improvement, too. It doesn't matter how many thousands of development activities a district offers if it fails to consistently connect teachers with the activities that are right for them at the right time.

As one teacher explained in a focus group, "If our students need choices, we need choices, too. We are differentiating for our kids, but no one is differentiating for me."<sup>79</sup> Likewise, teachers indicated that follow-through on the support they received was infrequent. Only one in five teachers said they "often" receive follow-up support or tailored coaching opportunities, and only one in 10 reported frequent opportunities for practicing new skills. Three-quarters told us they had been required to "sometimes" or "often" attend a professional development session on a topic or skill they already knew well.<sup>80</sup>

The districts we studied don't seem to be creating time for teachers to engage in the activities they say could be more effective. For example, even though nearly three-quarters of the teachers we surveyed said that observing other excellent teachers was a good use of their development time, they reported observing excellent peers less than twice a year.<sup>81</sup> By contrast, teachers spent an average of 24 hours per year participating in one-time professional development workshops, even though only 36 percent view them as a good use of time.<sup>82</sup> It seems, then, that beyond failing to help most teachers actually improve meaningfully, districts are not even meeting the arguably lower bar of giving teachers what they say they need.

**Only half of the teachers surveyed felt that most of their development activities led to lasting improvements in their instruction.**

Would teachers improve more if they participated in more activities they view as a good use of their time, or that actually focused on their individual development needs? Unfortunately, the answer is unclear. But it stands to reason that if current improvement efforts are getting such lackluster results, it would make sense for districts to help teachers first clearly understand what it is they need to improve upon, and then provide greater access to a variety of activities that, at a minimum, are perceived as more useful, and at best, may actually help them improve.

But the problem may not be as straightforward as teachers simply not receiving targeted professional development. We also saw evidence that many teachers may not trust the evaluation process and their formal evaluator's ability to help them improve.

In some cases, it may be that district and school leaders have failed to create enough trust in the development process by ensuring that teachers understand their strengths and weaknesses and how particular interventions are intended to help them meet those goals.

For example, just over a third of teachers “agreed” or “strongly agreed” that receiving performance evaluation ratings plays a crucial role in improving teacher practice.<sup>83</sup> And less than half of the teachers we surveyed agreed that their formal evaluator was able to direct them to development opportunities that were aligned to their needs.<sup>84</sup> When asked to identify an area for skill development, around two-thirds (64 percent) selected a development area that aligned with one their formal evaluator had also identified for them. But the remaining third either chose an area that did not align with their evaluator (28 percent), or did not report having been informed of any areas for improvement (8 percent).<sup>85</sup>

**“If our students need choices,  
we need choices, too.  
We are differentiating  
for our kids, but no one is  
differentiating for me.”**

-Teacher



## A Disjointed System

Through interviews and focus groups, we were able to gain greater insight into the maze of development activities teachers travel through and the various people with whom they engage along the way. Those conversations painted a picture of a well-intentioned system that, at least from a teacher's perspective, is as disjointed and impersonal as it is vast (*Figure 10*).<sup>86</sup> We heard that there are many central office employees focused on helping teachers, but that working consistently as a team is a challenge. Given that these development personnel often span different departments, report to different leadership and perform different functions, it's no wonder coordination can become difficult. We also heard from teachers that often, the people employed to support their development may not actually be on the same page about their development goals. They may not even coordinate with each other.

One district administrator we spoke to put it this way: “Truly, everybody is trying very hard to have a positive impact on the schools, but there is some redundancy and disconnect. The phrase ‘random act of school improvement’ is what pops into my head. We’re all out there trying to do our best but we’re not coordinating the efforts.”<sup>87</sup>

Teachers also seemed frustrated by the types of development they received and when they received it; it rarely met their expectations for what would be most helpful, even when it was “job-embedded” in spirit. Too often, teachers told us, their development experiences seemed repetitive or focused on information they could read and digest on their own.

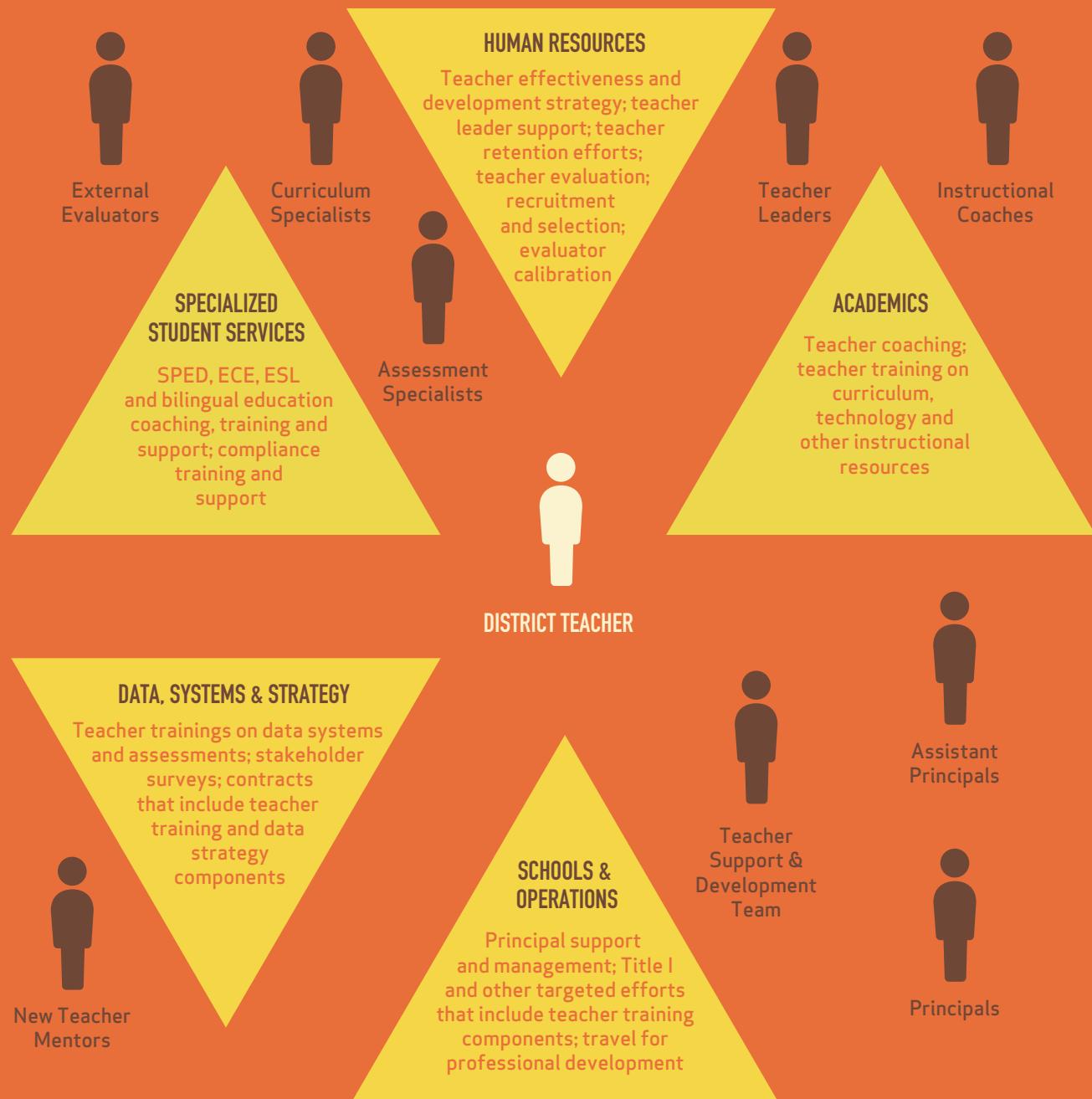
More broadly, teachers described a system that lacks any real vision or strategy—one that channels an enormous amount of time and resources to teacher development in the hope that they will turn into results. It’s a system in which few—from teachers to district leaders—seem to agree on what “teacher improvement” means or what “good teaching” looks like. In focus groups, teachers gave varied answers when asked how they measure improvement in their instruction, ranging from their own perceptions to others’ perceptions to student data. We heard similarly wide-ranging responses at the central office level. Not surprisingly, coordinated efforts to assess current development efforts were lacking as well. While central office staff were able to highlight some distinct support efforts that were being evaluated (or had been in the past), they could not point to systems currently in place to strategically assess all efforts across the board.

What is the vision of excellent instruction that every teacher should be striving to reach? Where do teachers stand right now compared to that standard of excellence? What, exactly, does every teacher need to do to start bridging the gap? How will teachers be able to tell whether they’re on the right track? Leaving these questions unanswered makes it impossible to help teachers set the right professional goals or identify the support they need to achieve them.

**“The phrase ‘random act of school improvement’ is what pops into my head. We’re all out there trying to do our best but we’re not coordinating the efforts.”**

—District Administrator

FIGURE 10 | THE TEACHER DEVELOPMENT MAZE



The current system for teacher improvement  
is huge but disjointed.

# AN EXCEPTION TO THE RULE?

The fourth school system we studied is a midsize charter management organization (CMO) operating across several cities. This CMO takes a markedly different approach to teacher improvement than the other districts we studied. While they have not solved the problem of teacher development entirely—and given the CMO's size, it is important to note our limited sample sizes here—their results seem promising, and point to several strategies other districts might consider as they reassess their efforts to help teachers improve.

---

## More Growth Over Time

Compared to the other districts we studied, the CMO seems to be supporting teachers to make greater improvements to their practice over time, based on both their observation scores and their overall evaluation ratings. Over three years, teachers in the CMO improved notably on their observations (a mean growth rate of .61 standard deviations per year, compared to .09, .11, and .02 respectively in Districts A, B and C).<sup>88</sup> The same is true for growth on overall evaluation ratings, where the CMO has a mean growth rate that is more than four times higher than that of the district with the next highest growth rate.

This is particularly noteworthy because teachers at all experience levels show more substantial growth than teachers with comparable experience in the other districts we studied (*Figure 11*).<sup>89</sup> In other words, teachers in the CMO are growing more rapidly in their early years, but so too are teachers with many years of classroom experience. In fact, about seven out of 10 teachers in the CMO showed substantial growth in their practice, as opposed to about three out of 10 in the districts we studied.

Students attending the CMO are getting consistently better results, too. When we look at teachers' value-added scores, we see that CMO teachers are making a greater impact on their students' learning, year to year, than teachers in surrounding schools. And overall test scores in math and reading are higher across the charter network than in surrounding schools as well.

.

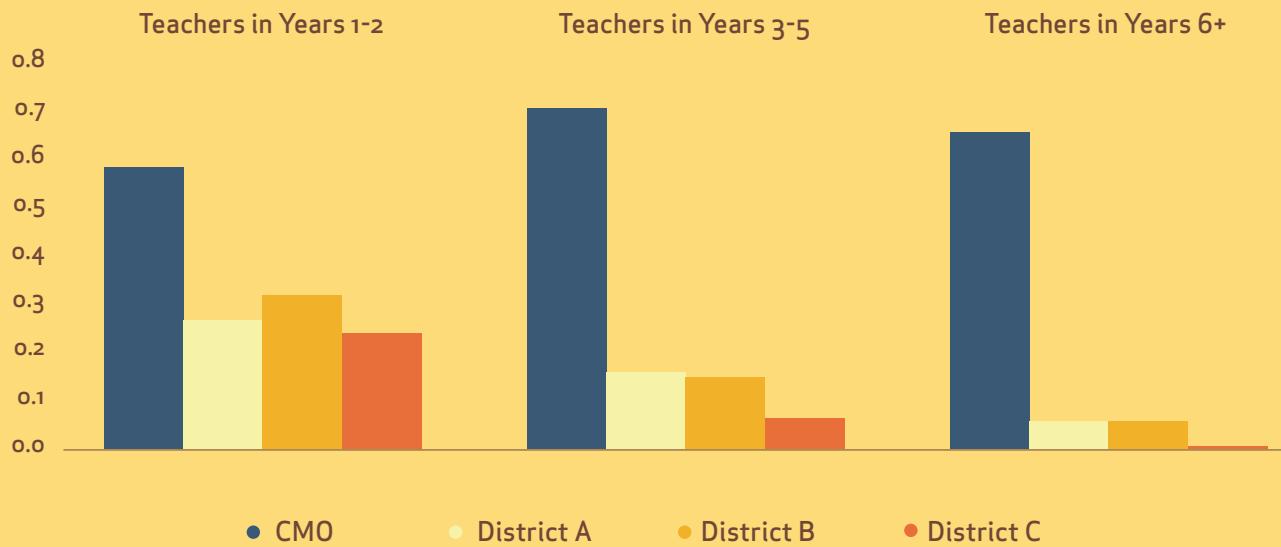
## What Are They Doing Differently?

The question is, what is the CMO doing differently from the districts we studied that might be garnering these different outcomes for teachers (and better results for students)? We wondered if there would be dramatic differences between improvers and non-improvers within the CMO that would point to particular strategies that seem to be having a marked effect on CMO teachers who make greater strides than their peers.

But when we compared improvers and non-improvers within the CMO, we found very few distinguishing features. In other words, even here, where we see higher rates of growth overall, there doesn't seem to be a magic formula of teacher supports that we can link to that growth. In terms of their development experiences and their mindsets, CMO teachers who grow look a lot like CMO teachers who don't grow. In some respects, meaningful improvement in the CMO—while more frequent than in the other districts we studied—is just as much of an individualized process, lacking in any particular pattern.

Nonetheless, we did find some differences on an institutional level in comparison to the districts we studied; specifically, a more disciplined and coherent system for organizing themselves around teacher development, and a network-wide culture of high expectations and continuous growth.<sup>90</sup>

FIGURE 11 | STANDARDIZED GROWTH RATES ON OBSERVATIONS BY TEACHING EXPERIENCE



**Teachers in the CMO grew more on observations compared to district teachers with similar years' experience.**

### Clear Roles and Responsibilities

This starts with how staff roles and responsibilities are organized. The CMO is very clear about who does what (and why) when it comes to teacher development. While a small number of central office staff do support teachers through observations and feedback, most central office staff are not dropping in and out of teachers' classrooms. Instead, the central office focuses primarily on setting instructional expectations, overseeing and coaching school leaders on progress toward those expectations, generating data to support teachers and school leaders and organizing CMO-wide professional learning experiences.

That's where the majority of central office teacher support stops. The rest of the CMO's teacher support efforts occur at the school site, through rethinking the traditional job functions of principals and assistant principals. Principals view themselves primarily as managers of their assistant principals, whose primary responsibility is coaching teachers and ensuring that high-quality instruction is occurring in

classrooms every day. While everyone is working toward the same goal—teacher improvement in order to see improved student learning—there is real discipline in what function everyone plays, and a specific strategy for how more teacher growth can and should occur.

### A Culture of High Expectations and Continuous Learning

That strategy is rooted in a robust and deliberate culture of high expectations and continuous learning. In focus groups, CMO teachers reflected on the sense that everyone in their school community is constantly working toward better instruction, and pushing each other to do their best work. One experienced teacher explained it this way: “Because I have been teaching for as long as I have, I have a lot of friends with similar years of experience who are doing the same thing from day to day and not necessarily growing. What’s unique about being at [my school] is that there is always going to be someone to push you. I don’t think I’ll ever be able to stagnate here.”<sup>91</sup>

We also found evidence of these high expectations in teachers' perceptions of their own performance. CMO teachers tend to more readily acknowledge that they still have room to improve. Eighty-one percent of teachers in the CMO agreed that they have weaknesses in their instruction, compared to 41 to 60 percent among teachers in the other three districts we studied.<sup>92</sup> Asked to rate their own teaching on a scale from 1 to 5, just 4 percent of teachers in the CMO gave themselves a top rating, compared to 24 percent or more of teachers elsewhere (*Figure 12*).<sup>93</sup> School leaders are more critical of their own abilities in comparison to district school leaders, as well.<sup>94</sup>

### Regular Feedback and Practice

This culture seems, at least in part, to be a product of deliberate actions that prioritize regular feedback. Each CMO teacher receives weekly observations from his or her coach, followed by a 30-45 minute debrief. And compared to teachers in our other three districts, CMO teachers are far more likely to report opportunities to practice teaching outside the classroom (82 percent reporting "sometimes" or "often" practicing, compared to 17 to 38 percent elsewhere).<sup>95</sup> All of that may help explain why CMO teachers are more likely to believe that observations and feedback are "effective for their improvement" (65 percent agreeing, compared to 36 to 50 percent in the other districts we studied).<sup>96</sup>

CMO teachers also spend two to three hours every week with other teachers, reflecting on instructional practices and outcomes from the past week, practicing new skills or reflecting on changes to be made next, and preparing for their upcoming units. Alongside these ongoing feedback and reflection cycles, there are several structured CMO-wide learning days throughout the year, as well as deep dives into student data outcomes.

**CMO teachers are far more likely to report opportunities to practice teaching outside the classroom.**

### A More Strategic Investment in Growth

Overall, teachers in the CMO report spending slightly more time on development activities than teachers in the other districts we studied (22 hours per month on average compared to 16 to 19 hours elsewhere).<sup>97</sup> However, those hours are spent on activities that appear to provide substantively greater opportunities for individualized support that focuses on specific development goals—and they occur within a culture that expects continual improvement.

This level of individualized support for teachers is expensive; in fact, the CMO spends significantly more per teacher and more of its total operating budget on teacher improvement efforts compared to the other districts we studied (on average, \$33,000 per teacher and 15 percent of its annual operating budget, compared to \$18,000 and 6 to 9 percent of annual budgets elsewhere).<sup>98</sup> Most of this difference comes from different allocations of time for school-level staff, as well as more teacher time spent on development—rather than additional support personnel, for example.<sup>99</sup> Critically, CMO leaders are constantly assessing the effectiveness of their efforts through data review and reflection.

### Understanding the Implications

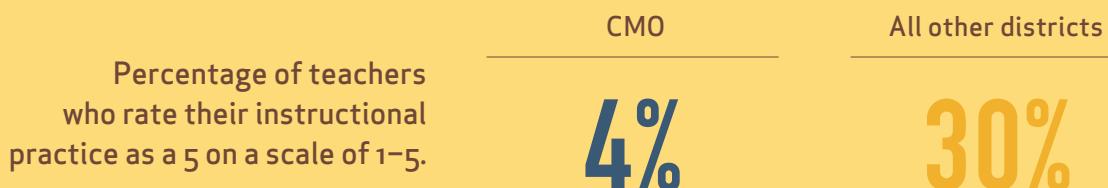
Does this mean that districts actually need to spend even *more* to get better results? Without question, the CMO is further confirmation that reducing investments in teacher support is not the solution. But the evidence also reveals the broader nature of the problem: Having a meaningful impact on teacher performance over time depends as much on the conditions in which development takes place as on the nature of the development itself.

Is it possible that this CMO attracts a certain type of teacher, one who holds especially high standards for his or her own performance? Certainly. But by establishing a clear vision and high expectations for excellence and giving teachers specific, actionable feedback on their areas for improvement, the system seems to be doing its part. Their culture of high expectations is met with an equal sense of commitment to helping teachers succeed.

FIGURE 12 | SELF-REPORTED PERCEPTIONS OF TEACHER PERFORMANCE



**Teachers in the CMO are more likely than district teachers to identify weaknesses in their instruction...**



**...and are less likely to give themselves a top rating.**

It is important to note other caveats. The CMO is considerably smaller than the other districts we studied. Can this intensive development model work at scale? We can't say for sure. With its relatively small number of teachers in total, and higher teacher turnover rates than other districts, it's hard to say conclusively that this approach would garner the same results at a much larger scale, or that growth for individual teachers would be sustained over many more years.<sup>100</sup> The CMO also recognizes that they need to increase the impact of their teachers in order to get better outcomes for students moving forward; while teachers' value-added scores indicate they have produced better than statistically expected results for their students over the past several years, they haven't seen a dramatic rise in student outcomes in all subjects and in all locations. They too have to find new ways to get all of their teachers to the next level of effectiveness.

Nonetheless, the evidence suggests that there is promise in the CMO's strategy of creating a culture and an organizational structure centered on teacher development and its impact on student learning, being deliberate about central office and school-level roles and responsibilities, and providing teachers with targeted, regular feedback from trusted leaders. Other school districts should consider how they could apply similar strategies in their own teacher development efforts.

**“What’s unique about being at my school is that there is always going to be someone to push you. I don’t think I’ll ever be able to stagnate here.”**

—CMO Teacher

# THE WAY FORWARD

It's clear that the school districts we studied are deeply invested—philosophically but also quite literally—in unlocking the untapped potential of their teachers. There is no corruption, venality or cynicism in the millions of dollars they devote to this effort, only a genuine, admirable desire to help teachers succeed at one of the toughest jobs in the world. If good intentions alone were enough to help teachers improve, every teacher would already be great.

Unfortunately, our research shows that our decades-old approach to teacher development, built mostly on good intentions and false assumptions, isn't helping nearly enough teachers reach their full potential—and probably never will. The incredible talent and creativity among our teachers lies untapped because we aren't creating the right combination of urgency and support in which teachers will take up the challenge—and be supported in the right ways—to continue growing. The pervasive beliefs that "we know what works," that more support for teachers is inherently good regardless of the results, and that development is the key to instructional excellence have all contributed to a vision of widespread teaching excellence just over the horizon that is mostly a mirage.

That doesn't mean we should give up. Those who would take our findings as evidence of "wasteful" spending or as an argument for drastically cutting support for teachers miss the point. Improving teacher effectiveness at scale—so that the vast majority of teachers master core instructional skills and students learn in rich, engaging and rigorous classroom environments—is critical to the long-term success of our education system and worthy of a substantial investment of time, attention and dollars. In fact, the CMO we studied spends substantially more than the districts on teacher development, but they also see many more teachers improving their practice substantially. Summarily cutting supports to teachers

would be a disaster; it would result in massive disruption, low morale and high attrition of top-performing teachers.<sup>101</sup> But the evidence shows that the challenge of helping teachers achieve real, meaningful improvement has been massively underestimated and oversimplified. It also offers a compelling argument about the limits of traditional notions of "professional development" in helping teachers improve.

Our research suggests that getting better at teaching is a lot like getting into better physical shape: a task that is difficult, highly individualized and resistant to shortcuts. Just as there is no single diet and exercise plan that will work for everyone, it's all but certain that there is no single development experience or activity that will get results for every teacher. We cannot try to force one solution on some 3.5 million individualized challenges. Yet we continue to search for the elusive treatment that will boost teachers' success overnight, in the same way we search for easy workout routines and lose-weight-fast strategies for improving our health.

While we found no set of specific development strategies that would result in widespread teacher improvement on its own, there are still clear next steps school systems can take to help their teachers more effectively. Much of this work is about creating the conditions for successful teacher development—conditions that do not currently exist.

# RECOMMENDATIONS

## 1 REDEFINE

### what it means to help teachers improve

School districts genuinely want to help their teachers, but what exactly does that mean? Currently, “helping teachers” generally means providing them with *more*—more workshops, more coaches, more seminars, more time for reflection or collaboration. Leading academics on behavior change emphasize that adapting to new and different expectations is a deeply complex process involving many factors, including what motivates us to break old habits and build new ones.<sup>102</sup> In other words, becoming more skilled at any job—especially one as complex as teaching—involves many other variables. For example, teachers can’t make the most of development opportunities if they don’t understand the end goal of those opportunities or don’t feel a sense of urgency to make improvement happen in the first place.

Our research suggests that, while understandable and well-intentioned, layering on *more* support is not the solution. Instead, we believe school systems need to make a more fundamental shift in mindset and define “helping teachers improve” not just in terms of providing them with a package of discrete experiences and treatments, but with information, conditions and a culture that facilitate growth and normalize continuous improvement.

This requires districts to clarify the goal of teacher development and approach it with some broader questions in mind: Are teachers getting accurate information about their performance? Do they have a clear vision of success to aim for and clear metrics to track their progress? Are school leaders equipped to guide teachers through the process? Does everyone involved view improvement as a top priority, or is it just something on the back burner? Specifically, we recommend that school systems:

#### Define “development” clearly, as observable, measurable progress toward an ambitious standard for teaching and student learning.

This is the first, most important step for any school district setting out to change their approach to teacher development. Districts need to develop a vividly clear vision of instructional excellence that can be observed and measured (through classroom observations and student assessment results, for example), and make advancing teachers toward this vision the primary goal of every development activity. This means setting clear goals for improvement in teacher practice and student achievement and reducing the emphasis on unreliable proxies for effectiveness, such as satisfaction, attendance or self-perceived improvement.

We believe the basic act of setting a clear and ambitious vision for excellent teaching and ensuring that principals and teachers understand that vision will have a galvanizing effect, as it seems to have had in the charter school network we studied. This vision is instilled in school cultures in part through formal evaluation systems, like observation rubrics and practice guides, but just as importantly in informal ways, like the conversations that veteran teachers have with new hires about expectations and administrative decisions about which teachers get promoted. The message should be clear: *In this school, we all strive to reach this level of performance, every day. Acknowledging our failures is just as important as celebrating our successes.*

Like all organizational change, a clear vision for excellence takes time to be understood and internalized by staff, especially in schools that have consistently set lower expectations. School leaders must lead this effort and ensure that the vision is internalized and teachers feel supported to take risks as they try to achieve it. Patience will pay off because no efforts at teacher improvement will succeed if there isn’t a shared understanding of the end goal.

**Give teachers a clear, deep understanding of their own performance and progress.**

Helping teachers starts by giving them a clear vision of success and honest feedback about their strengths and weaknesses. Currently, most teachers are told in innumerable ways that their level of performance is good enough. The resulting culture is an enormous drag on growth. Districts need to make sure that teachers have accurate information about how their performance compares to the vision of instructional excellence—which skills they've already mastered, and which they need to improve.

This isn't simply about evaluation ratings and the amount of feedback teachers receive, both of which are important, but also ensuring that such feedback is rigorous, tied to a clear vision for instruction and viewed by teachers as credible. Many of the teachers we surveyed seem to have little faith in their district's professional development efforts. Districts might focus on finding and training observers that teachers are likely to trust, and ensuring that school leaders are better equipped to provide teachers with trustworthy feedback. They might also consider supplementing observations with other resources, such as a video library of exemplar teaching or opportunities to observe highly effective teachers in their grade or subject.

**Encourage improvement with meaningful rewards and consequences.**

Changing one's professional practice can be difficult and uncomfortable. It often requires teachers to confront weaknesses, disrupt old routines and learn new skills. Even the most intrinsically motivated educator may need additional incentives to start and persist through the improvement process.

A thoughtful accountability system can help address the lack of urgency around teacher improvement we observed in the districts we studied and positively reinforce growth.<sup>103</sup> Creating meaningful rewards and consequences can send a clear message that improvement should be a top priority, and energize teachers about opportunities to innovate and grow. For example, districts can modify their observation rubrics and evaluation systems to focus on teachers' progress toward the vision of great teaching. This accountability for teacher improvement should extend to school leaders, too, and—critically—to central office staff in charge of teacher development.

**This vision is instilled in school cultures in part through formal evaluation systems, like observation rubrics and practice guides, but just as importantly, in informal ways, like the conversations that veteran teachers have with new hires about expectations and administrative decisions about which teachers get promoted.**

## REEVALUATE

### existing professional learning supports and programs

#### Inventory current development efforts.

School districts cannot accurately evaluate the impact of current or future development efforts without baseline information about their current approach. Before making any changes, they should create a comprehensive inventory of all the teacher development activities and initiatives they currently offer and calculate the costs associated with those supports. It's also likely that this process will uncover duplicative or misdirected efforts that can be eliminated quickly.

#### Start evaluating the effectiveness of all development activities.

Districts should stop making assumptions about which approaches to development work best and actually evaluate their impact instead, based on the standards they have set for measurable teacher improvement.<sup>104</sup> This means structuring development initiatives so that their impact can be measured—for example, by ensuring that there is a comparison group of teachers not receiving the same support—to assess the extent to which their results differ. And if a district's teacher evaluation system does not differentiate teacher performance well enough, the district may need to invest in stronger management, independent observers or a redesign of the evaluation system to ensure that it can capture real improvements in teachers' instruction.

This common-sense step of measuring the efficacy of particular activities should have a significant and positive impact. Imagine, for example, how your mindset would change if you were a literacy coach who is now being assessed not based just on your principal's subjective judgment or teacher satisfaction, but on whether the teachers you work with actually improve. You would likely find ways to monitor teachers' progress in a more systematic way and focus on teachers with the highest potential for improvement. And if you are a teacher, knowing that the literacy coaching is being evaluated based on whether it actually helps you do your job better will likely make you more invested in your district's efforts to help.

#### Explore and test alternative approaches to development.

Since current development efforts are not coming even close to working at scale, districts should make it a priority to try new approaches that push the limits of how much teachers can really improve. This could mean providing more time for teachers to practice instructional techniques with school leaders or expert peers in lieu of collaboration time for the sake of collaboration; programming opportunities for teachers to view colleagues at their own or nearby schools during time otherwise spent on administrative duties; identifying a single person to act as coordinator for all development opportunities so individual teachers aren't receiving disconnected and potentially contradictory guidance; or rooting development efforts in the particular needs of the individual teacher's students at a given time.

Districts might try focusing development efforts on teachers who seem to have higher potential to improve, such as early-career teachers and teachers on the cusp of being highly effective, as opposed to teachers who persistently struggle and should receive shorter-term interventions; or devolving some of the investment in teacher improvement directly to teachers, for example as a lump sum each year to spend on their development as they choose.

#### Reallocate funding for particular activities based on their impact.

Districts should redirect funding away from development activities that show little or no evidence of helping teachers improve and toward other activities that show greater potential (or toward pilots of brand-new approaches). For example, if literacy coaching is helping middle school teachers but not elementary school teachers, a district should expand the coaching initiative to more middle schools and try a different approach in elementary schools. Or perhaps a particular mentoring program doesn't seem to be helping teachers at all, year after year. In that case, the district should consider eliminating the program entirely in favor of activities showing more success, or new approaches that it hasn't yet tried.

### 3 REINVENT

#### how we support effective teaching at scale

As important as it is to clarify what development efforts should accomplish, it's just as crucial to be honest about what they can't. Our research suggests that even in the best-case scenario, focusing only on the kind and amount of development opportunities teachers receive will not result in improvement for most teachers, and that success will continue to be difficult to predict or replicate. Even an infinite amount of the best possible development is unlikely to make the vision of great teaching in every classroom a reality.

Given the apparent limits of professional development even in the best circumstances, we recommend that school systems embrace a new paradigm in which development is just one strategy among many for improving instructional quality. Districts need to combine the changes above with efforts to promote great teaching in other ways, some of which are proven and others of which are untested. We should be prepared to shift resources to these other levers if innovation in teacher improvement does not help substantially more teachers succeed in the classroom. We offer the following suggestions as a starting point:

#### **Balance investments in development with investments in recruitment, compensation and smart retention.**

Even as districts continue trying to help more teachers improve on the job, they should also prioritize recruiting teachers who already have a track record of success and retaining teachers after they actually become highly effective. In these areas, there are proven strategies, such as hiring teachers earlier and by mutual consent;<sup>105</sup> targeting effective teachers for retention through measures like simply asking them to stay<sup>106</sup> and added compensation for strong performance and additional responsibilities;<sup>107</sup> and exiting chronically low performers who have been given support and a fair chance to improve.

In most cases, the impact of keeping a high-performing teacher in the classroom even one or two more years will exceed that of helping a developing teacher reach a minimal standard of effectiveness. Where initiatives designed to spark teacher improvement don't prove successful, systems should repurpose funding to levers like these that districts can be confident will have a positive impact.

#### **Reconstruct the teacher's job.**

Currently, we expect teachers to be responsible for almost every single aspect of their classroom. Mastering the job requires mastering a daunting list of individual skills, from analyzing student data to designing assessments to using smart Internet searches to find the best content for students. That could be why there's no clear path to helping most teachers become truly great: Maybe it's simply unrealistic to expect millions of people to be great at everything that goes into such a complex job.

What if districts tried changing the job itself, for example by dividing it into many different roles, allowing for more specialization that plays to individual teachers' strengths? Entry-level positions would come with a smaller workload and a smaller scope of responsibilities—perhaps just focusing on small group instruction, or grading, or engaging families. As teachers build a track record of success, they could move up to other roles that gradually expand their responsibilities, to the point of becoming a lead teacher or managing larger instructional teams. This approach would help schools deliver higher-quality instruction to more students without requiring every teacher to master all the toughest instructional skills from day one—all while creating a natural career ladder for teachers that doesn't currently exist in most school systems and adding new, potentially more diverse, pipelines of talent into the profession.

## **Redesign schools to extend the reach of great teachers.**

In our current factory-era model of one teacher in a classroom of 25 students, it is difficult to scale the reach of top-performing teachers. Ultimately, the answer to ensuring excellent instruction for all students may not be to try to get all 3.5 million public school teachers in this country to a consistent level of excellence. Rather, it's worth exploring ways to combine the disaggregation of the teacher's role, as described above, with alternative models for school design that allow higher-performing teachers to reach more students.<sup>108</sup> For example, this might mean introducing blended learning technologies, even in small doses, to free up time each day for top teachers to reach more students.

## **Reimagine how we train and certify teachers for the job.**

In the short term, state regulators and school systems should hold higher standards for preparation programs so that more teachers enter the profession having mastered foundational instructional skills and are able to become effective within a reasonable time period.

Over the long term, however, we believe we must more radically reconsider how we help teachers learn the knowledge and skills necessary to thrive in the classroom. An extensive body of research has demonstrated that the type and amount of preparation teachers receive is poorly correlated to their actual performance.<sup>109</sup> Expecting teachers to master a wide range of instructional practices before setting foot in a real classroom may simply be unrealistic and inefficient. We believe we should shift our teacher training and licensure approach to focus on mastery of a clearly delineated progression of skills.

In this new paradigm, training would largely take place on the job, through practice—similar to an apprenticeship system, but more cost-effective, as these roles would fill an operational need and perform regular job duties, like tutoring, running small group instruction or supervising students during lunch and recess. We would not expect new teachers to have mastered all aspects of the role on day one, but rather to demonstrate mastery of a core set of gateway skills in a gateway role. For example, a new

teacher might start by being responsible for tasks such as grading student work, engaging parents, checking homework or running extracurricular activities. After demonstrating mastery in those skills, he or she would take on more advanced responsibilities, such as creating assessments, lesson plans or unit courses.

This would require state regulators and school systems to develop a new system of progressive licensure that is aligned with this sequence of roles, through which new teachers would progress based on demonstrated skills in the classroom and impact on student learning. Ultimately, only teachers who master all aspects of teaching or are able to manage a team that can deliver all aspects of teaching (classroom management, content, instructional delivery, student cognitive development) would gain full certification and become eligible for privileges such as tenure. These highly skilled professionals would also be compensated accordingly.

This approach to licensure would reinforce a culture of continuous learning and recognize that the greatest predictor of future success in the classroom has always been past performance. We believe it will not only improve instructional quality for more students, but accelerate skill mastery and improvement early in a teacher's career (when we know growth is most likely to occur); dramatically expand career path options for teachers; and open the profession to a wider and more diverse range of prospective educators.<sup>110</sup>

---

Our suggestions on how to redefine, reevaluate and reinvent efforts to help teachers improve reflect the lessons of this research, as well as our direct experience training and developing thousands of teachers over the last two decades (during which we have fallen victim to many of the same pitfalls we found in the districts we studied). Our hope is that these ideas will spark a candid new dialogue about teacher improvement and inspire school districts and training providers to try new approaches, measure their impact, find out what really works and share what they learn. It will take a collective, long-term effort to break through the mirage and finally unleash the full talent and creativity of our nation's teachers.

# THE MIRAGE: TECHNICAL APPENDIX

## 1. DATA

### DISTRICT DESCRIPTIVES

This report relies on data from three large, diverse districts and one charter school network.

Student racial compositions range from:

21-72% African American

1-37% Caucasian

9-34% Hispanic

2-8% Other races

### DATA SOURCES

*District Budget Data.* To investigate teacher improvement spending, each site provided budget data from fiscal year 2014 along with access to staff from relevant departments for interviews around personnel and non-personnel expenditures related to efforts intended to help improve teacher practice.

*Teacher Performance Data.* In each district, we used two to four years of teacher performance data (between 2010-11 and 2013-14), which each district collected as part of their formal evaluation system. While each district has a unique evaluation model, all have multiple measures that are factored into final scores. In our analysis, we consider performance derived from several measures:

- Final indicator-level observation scores (using the district's final ratings on each rubric indicator).
- Average overall observation scores (using the district's final overall observation score, which is typically created by averaging scores from multiple points in the year). In some districts, scores are available from multiple raters, but in other districts, scores are only provided by school leaders.
- Value-added scores (using the value-added score created by the state or district). Each state uses a different methodology for calculating value-added scores.
- Summative evaluation scores (using the district's final annual evaluation score, calculated using the district's official methodology).

*Student Performance Data.* In each district, we obtained three years of student performance data (between 2011-12 and 2013-14). We also collected publicly available data from state and district websites. We received information about student proficiency on state assessments, as well as student scale scores on state assessments. We were able to link these student data to teachers and to schools to create aggregate measures of student proficiency rates and average student performance for both teachers and schools.

*Other District Administrative Data.* In addition to information regarding performance, each school district provided teacher and administrator roster/demographic information, as well as school-level demographic information. These sources were used to calculate annual retention rates, as well as control for these factors in regression models.

Across the three districts and the CMO in our study, 61-84% of students qualify for free or reduced price lunch (FRPL). The total number of students and the percentage of students who qualified for free or reduced price lunch in 2013-14 are based on data from District A's state department of education's online database, District B's website, District C's state department of education's online database and data provided directly from the CMO.

*Surveys.* In all districts, the population of teachers was sent an online survey between January 27, 2014 and October 6, 2014. Survey respondents were demographically similar to the distribution of teachers in each district as a whole. Response rates were as follows: District A: 35%; District B: 26%; District C: 63%; CMO: 53%.

All school leaders in each district received a similar version of the survey. Response rates were as follows: District A: 34%; District B: 30%; District C: 46%; CMO: 50%.

These surveys were designed to address a variety of topics, ranging from teachers' reports of their participation in development activities to their mindsets around growth and development to their perceptions of their school environments. The school leader survey covered many of the same topics, asking the leaders to reflect on the development experiences of their teachers, assess their confidence in supporting teacher development and get their perspective on district support for development. In Appendix B, we provide more detail regarding the creation of measures from the individual survey items and subsequently used in our analysis of the link between performance and teacher self-reports.

*Teacher Focus Groups.* Between September 8, 2014 and March 9, 2015, we held 25 teacher focus groups across the three districts and the CMO. We created a purposive sample for focus groups, inviting teachers based on their classification as "improvers" vs. "non-improvers," definitions set via our analysis of teacher performance data. Of the invited teachers, 15 improvers and 27 non-improvers participated in District A; 20 improvers and 5 non-improvers participated in District B; 32 improvers and 28 non-improvers participated in District C; and 2 improvers and 5 non-improvers participated in the CMO.

## 2. ANALYSIS

In this report, we address the following research questions:

1. What is the financial investment being made in teacher development efforts across our partner districts?
2. To what extent do teachers improve their performance over time in each district, and does that improvement vary for teachers of different experience levels?
3. To what extent do teachers who improve their performance report taking part in similar development activities, sharing similar beliefs or mindsets, or working in similar school environments, compared to teachers who did not improve?

### RESEARCH QUESTION 1:

#### What is the financial investment being made in teacher improvement efforts across our partner districts?

We collected data through intensive document review and interviews with district staff at the central office and school level. Data were collected from a variety of sources, including but not limited to: district-wide budget reports, departmental line item budgets, personnel data, organizational charts, collective bargaining agreements, district policy documents like teacher evaluation handbooks and instructional calendars. In addition, we formally interviewed 127 central office and school-based staff members, including six principals across school levels—two elementary, two middle and two high school—in each district, and had follow-up and validation conversations with staff across the three districts and the CMO in order to understand staff roles and responsibilities, gather estimates for what percentage of time each staff role spent on direct teacher improvement and indirect teacher improvement efforts and understand all non-personnel spending on teacher improvement efforts in fiscal year 2014.

Using all of these data, we built personnel (PS) and non-personnel (NPS) teacher improvement budgets for each central office department and school-level support, estimated the cost of teacher time on improvement efforts and estimated the cost of investments in teachers' salaries for improvement efforts and excluding principals and assistant principals. See Appendix A for detailed explanations of each component of the teacher improvement cost calculation.

### TEACHER IMPROVEMENT

$$\begin{aligned} & (\text{Central Costs (PS and NPS)} + \text{School Costs (PS and NPS)} + \text{Teacher Time on Development} + \text{Teacher Salary Investments}) \\ & = \text{Total Cost} \end{aligned}$$

#### Tiers of teacher improvement spending

We also generated estimates on a sliding scale, tiering them into three groups, ranging from the most conservative definition of teacher improvement spending to a broader approach that considered anything that could be interpreted as teacher improvement efforts in each district. To do so, we determined the tier for each individual personnel and non-personnel line item within each of the components of the teacher improvement equation. The table on the following page summarizes the definitions of the three spending tiers. More detailed information can be found in Appendix A.

#### Full-time equivalents ratios

The staff counts and the percentage of time spent on direct teacher improvement and indirect teacher improvement efforts at the central office and school level were used to calculate the number of full-time equivalents (FTEs) each district dedicated to teacher improvement work in 2013-14.

$$\begin{aligned} & ((\text{N Role} * \% \text{ Direct Improvement}) + (\text{N Role} * \% \text{ Indirect Improvement})) \\ & = \text{FTEs} \end{aligned}$$

This data was used to calculate the ratio of teachers to central office and school-level personnel, using only staff who dedicate at least 50 percent of their time to direct teacher improvement efforts.

$$\begin{aligned} & (\text{N Teachers} / (\text{N Role} * \% \text{ Direct Improvement})) \\ & = \text{Span of Control} \end{aligned}$$

*(limited only to central office and school-based staff other than principals and assistant principals whose  
% Direct Improvement ≥ 50%)*

<b>LOW</b>	<i>The baseline costs districts are incurring to improve teacher practice.</i>	Central Costs	Personnel: Select direct and indirect teacher improvement staff time identified as "traditional" support costs (excluding teacher evaluation, principal managers and leadership development staff, and select data strategy staff) Non-personnel: Training and support resources, materials and contracts for teacher support
		School Costs	Personnel: School leader time for meetings with teachers for improvement (not evaluation-related); other school-based support staff time on direct teacher improvement efforts; and teacher development-related substitute coverage Non-personnel: All school-based non-personnel expenditures on teacher improvement
		Teacher Time on Development	Contracted time, survey time estimates for formal collaboration and payments made to teachers to attend professional development sessions
		Teacher Salary Investments	Stipends for teachers for teacher-leader roles, participating in selective leadership development programs and earning education credits
<b>MEDIUM</b>	<i>The baseline costs plus other spending that is grounded in work directly aligned to districts' strategies to improve teacher practice.</i>	Central Costs	Personnel: Additional direct and indirect teacher improvement staff time, including all direct and indirect time related to teacher evaluation Non-personnel: Training and support resources for improvement staff (coaches, etc.) and teacher evaluation non-personnel expenditures
		School Costs	Personnel: School leader time for teacher evaluation (minimum district requirements), evaluator calibration, and strategy for teacher development; and other school-based support staff time on indirect teacher improvement efforts
		Teacher Time on Development	Survey time estimates for coaching and peer observations; teacher time meeting with their formal evaluator (minimum district requirements)
		Teacher Salary Investments	Lanes spending
<b>HIGH</b>	<i>All costs that one could argue should, but may not always, be considered teacher improvement spending.</i>	Central Costs	Personnel: All direct and indirect staff time including all direct and indirect time for principal managers working with principals to support teacher development and data strategy staff Non-personnel: Expenditures for data strategy and leadership development
		School Costs	Personnel: School leader time for teacher evaluation (maximum estimate) and other school leader district-required activities related to teacher improvement
		Teacher Time on Development	Teacher time meeting with their formal evaluator (survey estimate)
		Teacher Salary Investments	Performance bonuses

### RESEARCH QUESTION 2:

To what extent do teachers improve their performance over time in each district, and does that improvement vary for teachers of different experience levels?

In an effort to identify improvement trends across these districts, we used several strategies to identify whether or not individual teachers improved over time. Given that we were looking at changes between school years, by definition, the teachers included in this portion of the analysis had to remain present in the data for the years studied. Descriptions of the various approaches to calculate our growth flags are included in more detail below.

#### Tracking "meaningful" change

In order to identify teachers whose performance changed meaningfully over the last two to three years, we first subtracted a teacher's 2011-12 (2012-13 in District C) overall evaluation score from their 2013-14 score. We then compared this difference to the distribution of all 2013-14 evaluation scores in the same district. Teachers whose scores increased by at least a half a standard deviation (based on the 2013-14 site-specific distribution of evaluation scores among all teachers) were considered to have "improved meaningfully"; we considered teachers whose scores decreased by at least a half a standard deviation to have "declined"; all other teachers were not considered to have changed their score meaningfully. Teachers whose initial score was too low or too high to be eligible to improve or decline were not included in the analysis.

We chose half a standard deviation as our threshold for meaningful change because it aligned well to the typical differences seen among early career teachers. Across all three of our districts, a half a standard deviation was larger than the average difference in 2013-14 performance between first- and second-year teachers, but smaller than the difference between first- and third-year teachers.

#### [Tracking growth rates over time](#)

We constructed simple annual growth rates in Districts A and B and the CMO by subtracting each teacher's 2011-12 performance score from their 2013-14 score and dividing this number by two to represent the average growth made per year between these two years. Only teachers who had a performance measure in all three years spanned were included. For District C, we simply subtracted each teacher's 2012-13 performance score from their 2013-14 score.

Because each district has its own performance scales, we standardized each teacher's growth rate by dividing each rate by the standard deviation of the performance score among all teachers in the district in the 2013-14 school year. Thus, standardized growth rates represent the number of standard deviations a teacher tended to change each year.

*Figure 4* represents the average standardized growth rate for all teachers based on their years of teaching experience in 2011-12 in Districts A and B and 2012-13 in District C.

#### [Tracking change over time on specific rubric indicators](#)

In addition to changes in teachers' final observation scores between years, we were also interested in whether or not teachers' scores on specific rubric indicators changed over time. None of the districts included a single final rating at the indicator level. Instead, each time a teacher received a formal observation, every indicator received a categorical rating. There were four category choices in Districts A and B and five categories in District C. In order to construct an overall annual rating on specific instructional indicators, we first converted each categorical rating to an integer, with the lowest possible ratings converted to a 1, the second lowest converted to a 2, and so on. We then averaged each teacher's ratings from the school year in that indicator to obtain a value between 1 and 4 in Districts A and B, and 1 and 5 in District C. Based on that final average, we assigned the following labels:

- "Low": Averages less than or equal to a 2 in Districts A and B, and less than or equal to a 2.33 in District C.
- "Developing": Averages greater than a 2 but less than a 3 in Districts A and B, and greater than a 2.33 but less than a 3.67 in District C.
- "Effective": Averages equal to or greater than a 3 but less than a 3.5 in Districts A and B, and equal to or greater than 3.67 but less than 4.33 in District C.
- "Highly Effective": Averages equal to or greater than a 3.5 in Districts A and B, and equal to or greater than 4.33 in District C.

Because the districts had a different number of rating categories, these thresholds were set to represent equivalent distances on each district's scale. Put another way, a score of a 3 on a scale ranging from 1 to 4 represents a point two-thirds up the scale; two-thirds up a scale ranging from 1 to 5 is approximately 3.67, so we used these two points to set the "effective" bar.

To track indicator ratings over time we repeated the above process in each year of data and assessed how teachers at each performance designation in one year performed in subsequent years.

To project the number of years until the average teacher was "highly effective" in a given indicator, we created a line of best fit representing the annual trend in overall indicator scores among teachers we could track each year and identified when that line, extended into the future, would surpass our bar for "highly effective." Specifically, we ran a simple linear regression using the year (centered on 2013) to predict the annual indicator score. We then identified how many years past 2013 would be required until the regression line was estimated to surpass "highly effective."

#### [Pseudo returns to experience](#)

To explore how teaching experience affects performance, we created Figure 5 to display what we call "pseudo returns to experience." While we were informed by the returns to experience literature, given our short panel of data, we did not have the opportunity to follow a more traditional returns to experience model, so we created this less sophisticated alternative method.

First, we had to determine the best way to define years of teaching experience for individual teachers. Only District A specifically tracked years of teaching experience in each year under study. In District B, we were able to identify years of teaching experience each year from the district's payroll data, which connected to teaching experience via its step system. We were unable to obtain teaching experience information or the necessary payroll data from District C. Instead, we used teachers' self-reported years of teaching experience. For teachers who did not respond to our survey, we substituted the number of years since the teacher's hire date. Because we were only able to obtain teacher survey and district hire dates for teachers working in the 2013-14 school year, we were unable to identify teacher experience information for teachers who left the district prior to 2013-14.

Because of the different ways we identified years of teaching experience across our districts, we tested the robustness of our findings by also using years in the district, which we defined consistently across sites. Our results were qualitatively similar.

With years of teaching experience assigned, we analyzed data from each site separately. In each site, we first standardized each teacher's overall evaluation score against the average overall evaluation score in the same school year and same evaluation group. For example, in some districts, the weights used in a teacher's overall evaluation score depend on the evaluation measures available in their setting. A teacher in a specific scoring group would be standardized against all other similar teachers in the same school year.

We then pooled the standardized performance results across the last several years (three in District A, four years in District B and two years in District C). Next, we took these standardized evaluation scores and centered them on the average score of all first year teachers in the pooled data set by subtracting the average standardized score among first year teachers from all teachers' scores. Last, we calculated the average of this "centered, standardized score" among teachers who were in the given experience level. This means some teachers have the potential to be represented in these results multiple times. For example a 4th year teacher in 2011-12 would have his 4th year results contribute to the 4th year average; his 5th year results contribute to the 5th year average, and his 6th year results contribute to the 6th year average. Similarly, a teacher who was in her 20th year in 2011-12 could have three years of results all contribute to the 20-24 experience band. This approach does not try to make any correction for differential attrition.

### **RESEARCH QUESTION 3:**

**To what extent do teachers who improve their performance report taking part in similar development activities, sharing similar beliefs or mindsets, or working in similar school environments, compared to teachers who did not improve?**

While Research Question 2 explored aggregate district trends in teacher performance, with Research Question 3, we also sought to identify individual teachers as "improvers" or "non-improvers" and focused on factors related to improvement in individual teacher's performance.

#### **Improvers vs. non-improvers**

We identified teachers who improved significantly using multiple definitions of growth.

Beyond simply looking at changes in individual performance measures, we looked for teachers who grew more than their peers with similar experience and who started off at the same level of performance. We also grouped teachers into quartiles, assessing who was making the most and least growth over a two- to three-year period. We tracked this type of movement across four different measures of growth: change in total observation scores, change in value-added scores, change in total evaluation scores and change in standardized overall evaluation scores.

Individual teachers are flagged as "improvers" or "non-improvers" based on the following definitions:

**1. District Rating Change:** This definition identifies teachers by calculating change in district evaluation ratings over time in two ways:

- a. *Simple Change:* Teachers who went up, down, or stayed the same were identified by subtracting their overall evaluation rating between 2011-12 and 2013-14 in Districts A and B, and between 2012-13 and 2013-14 in District C. Additionally, teachers who had the highest rating in both years were categorized as "Always Effective."
- b. *Detailed Change:* Using all three years of data (2011-12, 2012-13, and 2013-14) in Districts A and B, and two years in District C, teachers who had the type of movement outlined below were identified.
  - i. Transformative Growth or Decline – Movement up or down 2 rating levels, and never dropped (improved) a rating over the time period
  - ii. Consistent Growth or Decline – Movement up or down 1 rating level, and never dropped (improved) a rating over the time period
  - iii. Remained the Same – Remained the same rating in all years
  - iv. Always Effective – Earned the highest possible rating in each year  
*(Note: These were not used for the CMO as they do not provide their teachers with final categorical evaluation ratings at the end of the year.)*

**2. Beat the Average Growth:** This definition identifies teachers by calculating whether or not they beat the average growth for their experience level using 2011-12 to predict 2013-14 performance. (In District C, we used 2012-13 performance to predict 2013-14 performance.) To do this, we regressed the 2013-14 outcome on a cubic polynomial of the 2011-12 or 2012-13 outcome on the same measure and experience (entered as separate dummy variables from first year to 10+ years). All teachers with positive residuals were considered to have beaten their average growth.

**3. Fixed Amount Growth:** This definition identifies teachers by using the same regression model as the one specified in "Beat the Average Growth," but only teachers whose actual 2013-14 score surpassed their estimated 2013-14 outcome by at least 0.5 standard deviations (based on that outcome's distribution among all teachers in the most recent year of data) were considered improvers; all others were non-improvers. In other words, teachers who had residuals that were equal to or surpassed a half a standard deviation were identified as improvers; all others were non-improvers.

**4. Fixed Amount Growth-Split:** This definition uses the same approach as "Fixed Amount Growth". However, to be considered a non-improver in this definition, a teacher must have a 2013-14 score that was at least 0.5 standard deviations below expectation, i.e. residuals less than or equal to negative half a standard deviation. Teachers whose performance was within a half a standard deviation of expectation were excluded from this growth definition.

**5. Quartiles of Growth:** This definition uses the same regression model outlined in the three previous definitions. For all teachers, we calculated the difference between actual 2013-14 performance and estimated 2013-14 performance, i.e., we calculated a residual, and split these results into four quartiles, with the top quartile representing the 25% of teachers who most exceeded their expected performance.

### Teacher-Level Analysis

To investigate potential differences between teachers who did and did not improve over time, performance data were linked to survey data. First, we performed simple descriptive analyses and t-tests to determine whether or not teachers flagged as “improvers” or “non-improvers” differed significantly in terms of the following:

- The type and dosage of teacher professional learning experiences,
- the presence of certain mindsets and
- the characteristics of their environments.

We completed this analysis for various levels of teaching experience separately and together to look for potential differences between improvers and non-improvers at different stages of their career. We also created quartiles of the teacher time reports for each professional learning experience investigated in the survey to investigate potential differences in the distribution of improvers and non-improvers at the highest and lowest ends of the spectrum.

Additionally, we performed a series of linear regression analyses to investigate potential relationships between teacher performance and increased teacher support efforts, increasingly positive mindsets and teacher perceptions of their environment on performance. We first looked at all items in separate models, controlling for years of teaching experience and prior performance. In an additional series of linear regressions, we sought to determine whether teachers who had more “optimal” development experiences could be expected to have higher performance by regressing the various survey constructs in combination with each other.

We also performed a series of logistic regressions, using the same set of survey constructs, to test whether or not certain development experiences, mindsets, or environments increased the likelihood of being identified as an improver.

### School-Level Analysis

We followed a very similar approach to analysis of school-level trends. Because school-level survey response rates were uneven, we were concerned about attributing responses from just a small fraction of teachers in the building to “school characteristics” that would be used as predictors of performance or likelihood of growth of teachers in the school. While we investigated a variety of decision rules related to response rates and teachers whose multi-year growth rate could be tracked, we settled on the following requirements to both maximize the number of schools included in the analysis as well as plausibly make the case that teacher perceptions could stand in as school-level measures:

- *Survey Requirements:* At least five survey responses and at least 25% of the teaching population at the school
- *Growth Measure Requirements:* At least five teachers with performance data available from the specified time frame and at least 25% of the teaching population at the school with available growth data

For schools who met these criteria, we conducted four separate analyses to explore relationships between concentrations of teachers who improve and professional development experiences, mindsets and characteristics and perceptions of school environments.

First, we ran correlations between the percent of teachers identified as improvers in each school to the average school-level response to the various survey constructs used in the teacher-level analysis along with additional items related to school leader perceptions and teacher and leader survey response alignment. As in the other analyses using teacher growth as the outcome, we tested this relationship with multiple definitions of growth, based on teachers who improved their overall rating category; teachers who had evaluation scores that exceeded those of other similar teachers; teachers who were in the top quartile of overall evaluation scores; and teachers who were in the top two quartiles of overall evaluation scores.

Next, we simply looked at a dichotomous school-level outcome: schools were categorized as having “high growth” and “low growth” based on the percentage of teachers who met our growth definitions. Schools were considered “low growth” if fewer than 10% of their teachers were flagged as improvers, and as high growth if more than 50% of teachers were identified as improvers. Alternative cut-offs were required for District B due to sample sizes, with “high growth” defined as 33% and “low growth” as 10%. Using t-tests, we determined if teacher responses regarding professional development activities, mindsets or school culture differed in high growth and low growth schools.

Additionally, linear regression analysis was used to regress teacher participation in professional development, mindsets and school culture on the percentage of teachers identified as “improvers” at the school. These models also controlled for: FRPL from 2013-14, the percent of minority students in 2013-14, enrollment in 2013-14, attrition from 2012-13 to 2013-14, the percent of teachers with one to two years of experience in the school in 2012-13, whether or not teachers were in the same school in 2012-13 and 2013-14, and whether or not the school had the same principal in 2012-13 and 2013-14. These school-level regression analyses produced results qualitatively similar to teacher-level regressions.

Finally, we used the same predictors in models to determine whether student proficiency in reading and math, respectively, were related to aggregate teacher experiences, mindsets or perceptions of school environment.

# APPENDIX A

## Detailed Summary Method for Estimating Teacher Improvement Spending

TNTP collected and analyzed budget information from fiscal year 2014, or the 2013-14 school year, to capture all expenditures related to improving teacher instructional practice. To calculate the total cost incurred to improve teacher practice, all direct and indirect teacher improvement efforts related to **Personnel Spending (PS)** and **Non-Personnel Spending (NPS)** at the central office and school level, the cost of teacher time dedicated to these efforts, and the salary investments districts make in teacher improvement were included.

### TEACHER IMPROVEMENT

$$\begin{aligned} & (\text{Central Costs (PS and NPS)} + \text{School Costs (PS and NPS)} + \text{Teacher Time on Development} + \text{Teacher Salary Investments}) \\ & = \text{Total Cost} \end{aligned}$$

**Direct Teacher Improvement:** Personnel and non-personnel expenditures associated with direct teacher contact (e.g., teacher evaluation, new teacher support, professional development for teachers, teacher coaching, etc.). More specifically:

1. *Direct Personnel Spending* represents staff who work directly with teachers on improving their practice, such as principals, coaches, etc.
2. *Direct Non-Personnel Spending* represents any expenditure associated with teacher training, new teacher support, teacher evaluation, career pathways spending, and contract expenses with a teacher training component.

**Indirect Teacher Improvement:** Personnel and non-personnel expenditures intended in part or in total to improve teacher practice but not targeted directly to the teacher, including:

1. *Indirect Personnel Spending* represents staff that manage direct teacher improvement efforts or spend time providing strategic or operational support to teacher improvement efforts.
  - a. Managerial support are costs associated with managing direct support to teachers.
  - b. Strategic support are costs associated with planning or approving policies and programs geared towards improving teacher practice.
  - c. Operational support are costs to provide logistical support and execution of teacher improvement efforts such as trainings.
2. *Indirect Non-Personnel Spending* are any expenditures associated with direct training for school or central office staff who are “one person away” from the teacher on topics geared towards improving instructional practice (e.g., Principal trainings or time they spend focusing on improving their ability to improve practice but not trainings for principal managers who ultimately train principals).

### A1. CENTRAL COSTS

#### Central Personnel Spending (PS):

The average compensation (salary and benefits) for a given role and estimates from central office staff interviews about the percent of time spent on direct and indirect teacher improvement efforts are used to calculate this cost. Coding was applied to staff titles to assign them to spending tiers.

$$\begin{aligned} & ((\text{Avg. Role Compensation} * \% \text{ Direct Improvement}) + (\text{Avg. Role Compensation} * \% \text{ Indirect Improvement})) * N \text{ Role} \\ & = \text{Central PS} \end{aligned}$$

Tiers include the following:

**Low:** Direct Time for All Staff (excluding staff who work on teacher evaluation, principal managers and leadership development staff, and some data strategy staff based on job description)  
 + Indirect Time for Staff in Professional Development Departments or with Roles Designed to Directly Support Teacher Improvement

**Medium:** + Direct and Indirect Time for Teacher Evaluation Staff  
 + Indirect Time for All Staff (excluding principal managers and leadership development staff and some data strategy staff based on job description)

**High:** + All Direct and Indirect Time for Principal Managers, Leadership Development, and Data Strategy Staff

**Central Non-Personnel Spending (NPS):** Depending on the site, line item level budgets or overall non-personnel teacher support spending data were provided. Coding was applied to expenditures to assign them to spending tiers.

$$\begin{aligned} & ((\text{Item Spend} * \% \text{ Direct Improvement}) + (\text{Item Spend} * \% \text{ Indirect Improvement})) \\ & = \text{Central NPS} \end{aligned}$$

Tiers include the following:

**Low:** Costs related to traditional teacher professional development and contracts with teacher training components

**Medium:** + All costs related to teacher evaluation and professional development for coaches and content managers

**High:** + Professional development for other teacher support staff and school leaders and contracts for data and strategy

## A2. SCHOOL COSTS

### School Personnel Spending (PS):

Personnel spending at the school level includes three separate components:

$$\begin{aligned} & (\text{Support Personnel Cost} + \text{School Leader Time Cost} + \text{Teacher Development-Related Substitute Coverage}) \\ & = \text{School PS} \end{aligned}$$

**1. Support Personnel Cost:** The average compensation for a given role and estimates from staff interviews about the percent of time spent on direct and indirect teacher improvement efforts are used to calculate this cost. Coding was applied to staff titles to assign them to spending tiers.

$$\begin{aligned} & ((\text{Avg. Role Compensation} * \% \text{ Direct Improvement}) + (\text{Avg. Role Compensation} * \% \text{ Indirect Improvement})) * \text{N Role} \\ & = \text{Support Personnel Cost} \end{aligned}$$

Tiers include the following:

**Low:** All Direct Time

**Medium:** + Indirect Time

**High:** Same as Medium Tier

**2. School Leader Time Cost:** A sample of principals were interviewed in each site across school levels to gain additional insights into school embedded support efforts and school leader time. Calculations for this component use average hourly rates for school leaders and the number of hours school leaders spend on teacher improvement activities as sourced from interviews, other central information gathered about school leader time requirements, and teachers' contracts. A description for each portion of the equation follows.

$$\begin{aligned} & (\text{Teacher Evaluation Time Cost} + \text{Other School-Level Meetings Time Cost} + \text{District Requirements Time Cost}) \\ & = \text{School Leader Time Cost} \end{aligned}$$

Tiers include the following:

**Low:** Meetings with Teachers for Improvement (Not Evaluation Related)– e.g., Faculty Meetings with PD Components or Student Data Meetings (Interview Data)

**Medium:** + Minimum District Requirements for Evaluation Activities and Time Requirements

+ Strategy Meetings for Teacher Development (Interview Data)

+ Time Requirements related to Evaluator Calibration and Training

**High:**

+ District Requirements for Evaluation Activities but Time Estimates from Principal Interviews and Additional Walkthroughs

+ All Instructional Leadership Activities

- a. **Teacher Evaluation Time Cost:** During interviews, principals were asked to estimate how much time they spend on the various evaluation activities per teacher. District minimum evaluation requirements were obtained from 2013-14 Evaluation Handbooks. Data was captured on: Initial Beginning-Of-Year (BOY) Meetings: Prep, Meeting; Formal Observations: Pre-Conference, Observation, Writing Feedback, Post-Conference; Informal Observations: Observation, Writing Feedback, Post-Conference; Walkthroughs: Walkthrough, Feedback; and Summative End-Of-Year (EOY) Meetings: Prep, Meeting.

$$\begin{aligned} & (\text{Total Hours for All Teachers} * \text{Average Leader Hourly Rate}) \\ & = \text{Teacher Evaluation Time Cost} \end{aligned}$$

- b. **Other School-Level Meetings:** During interviews, principals were asked to list the teacher support meetings at their school in which they are involved along with the frequency and duration. Teacher Collective Bargaining agreements were also used to gather information about school-level meeting requirements and their frequency and duration. Where contracts were not specific about "Direct" or "Indirect" school leader time with teachers, interview information was used or estimates were derived based on the described content and purpose of the meeting. These meetings fall into two categories: 1) Meetings with Teachers for Improvement (Not Evaluation Related), and 2) Strategy Meetings for Teacher Development.

$$\begin{aligned} & (\text{Total Annual Meeting Hours} * \% \text{ Teacher Improvement} * N \text{ Principals} * \text{Avg. Principal Hourly Rate}) + (\text{Total Annual Meeting Hours} * \\ & \% \text{ Teacher Improvement} * N \text{ AP} * \text{Avg. AP Hourly Rate}) \\ & = \text{Other School-Level Meetings Time Cost} \end{aligned}$$

CMO teachers receive extensive coaching from school leaders, so an additional component was created:

$$\begin{aligned} & (\text{Hours per Teacher} * N \text{ Teachers} * \text{Avg. Leader Hourly Rate}) \\ & = \text{CMO Coaching Support Cost} \end{aligned}$$

- c. **District Required Time Cost:** Information obtained from central office interviews, school leader interviews, and district websites was used to generate a list of district requirements for school leaders related to teacher support and to determine: 1) the count of leaders in attendance, 2) the duration of the activity, and 3) the frequency. District staff or school leaders were also asked to estimate what percentage of each type of requirement was related to teacher improvement. Examples of activities included in this cost are: leadership development series, evaluator calibration training and school leader coaching.

$$\begin{aligned} & (\text{Annual Hours of Activity} * \% \text{ Teacher Improvement} * N \text{ Principal} * \text{Avg. Principal Hourly Rate}) + (\text{Annual Hours of Activity} * \\ & \% \text{ Teacher Improvement} * N \text{ AP} * \text{Avg. AP Hourly Rate}) \\ & = \text{District Required Time Cost} \end{aligned}$$

3. **Teacher Development-Related Substitute Coverage:** The cost for teacher development-related substitute coverage is included in the **Low** tier.

#### School Non-Personnel Spending (NPS):

All school-level NPS spending is coded as direct teacher improvement efforts. These costs are in the **Low** tier.

### A3. TEACHER TIME ON DEVELOPMENT

Spending in this component accounts for any time teachers are being paid to partake in development activities at the district or school level. It does not include time they spend independently on improving their instruction. The cost of teacher time spent in efforts to improve their instruction is based on average hourly wage and includes costs related to:

$$\begin{aligned} & (\text{PD Attendance Payments} + \text{Contracted Time} + \text{In-School Embedded Support} + \text{Meeting with Evaluator}) \\ & = \text{Teacher Time on Development} \end{aligned}$$

Tiers include the following:

**Low:** PD Attendance Payments

+ Contracted Time

+ In-School Embedded Support (Teacher Survey Data for Formal Collaboration only)

**Medium:** + In-School Embedded Support (Teacher Survey Data for Coaching and Peer Observations)

+ Minimum District Requirements for Meeting with Evaluators

**High:**

+ Teacher Survey Data for Meeting with Evaluator (instead of Minimum District Requirements)

**PD Attendance Payments:** *Payments made to teachers for attending professional development as sourced from district budgets*

**Contracted Time:** *The hours of formal, district-mandated professional learning as sourced from Collective Bargaining Agreements (CBAs) or work requirements*

We used the Education Resource Strategies (ERS) Professional Growth & Spending Calculator<sup>1</sup> – Teacher Time Worksheet and information from each district's CBA or work requirements to calculate the cost of teachers' contracted time in professional development.

$$\begin{aligned} & (\text{Annual Non-Instructional PD Hours in Contract} * \text{Cost of Teacher Hour} * N \text{ Teachers}) \\ & = \text{Contracted Time} \end{aligned}$$

Annual Non-Instructional PD Hours in Contract = Annual Hours in Contracted Non-Student PD Days + Annual Hours of Release for PD

Cost of Teacher Hour = (Average Teacher Compensation – Cost of Lanes Spending) / Annual Contracted Work Hours per Teacher

**In-School Embedded Support:** *The hours of formal collaboration, coaching, and peer observations as sourced from the teacher survey*  
 This cost leverages ERS Professional Growth & Support Spending Calculator's estimate for "Regular and Frequent PG & Collaboration Time During Instructional Day," yet instead of summing the weekly time like ERS, annual time was used from teacher survey reports. We have summed the annual hours of formal collaboration, coaching, and peer observations, which most closely matches ERS's examples of "required collaborative planning time, weekly coaching, etc." We assumed 30 minutes for each peer observation instance. When appropriate, we adjusted the annual hours of formal collaboration from teacher survey reports to prevent counting the annual hours of release for professional development twice, given potential overlap between the two based upon policy.

$$\begin{aligned} & (\text{Average Annual # of Hours Spent on Formal Collaboration, Coaching, and Peer Observation} * \text{Cost of Teacher Hour} * N \text{ Teachers}) \\ & = \text{In-School Embedded Support} \end{aligned}$$

**Meeting with Evaluator:** *The hours of evaluator meetings was calculated in two ways: 1) Using minimum district evaluation requirements for Initial BOY Meetings, Pre-Observation Conferences, Post-Observation Conferences, and Summative EOY Meetings gathered from teacher evaluation handbooks and principal interviews (see School Leader Time Cost – Teacher Evaluation Time Cost above), and 2) Using data from the teacher survey.*

$$\begin{aligned} & (\text{Average Annual # of Hours Spent Meeting with Formal Evaluator} * \text{Cost of Teacher Hour} * N \text{ Teachers}) \\ & = \text{Teacher Evaluation Time} \end{aligned}$$

#### A4. TEACHER SALARY INVESTMENTS

The cost of teacher salary investments includes the following:

$$\begin{aligned} & (\text{Stipends} + \text{Lanes Spending} + \text{Performance Bonuses}) \\ & = \text{Teacher Salary Investments} \end{aligned}$$

Tiers include the following:

**Low:** Stipends (e.g., for taking on leadership roles, earning education credits, and participating in development programs)

**Medium:** + Lanes Spending

**High:** + Performance Bonuses

**Stipends:** *Monetary supplements for teachers in leadership roles, who participate in selective programs designed to improve their leadership skills, or for earning education credits*

**Lanes Spending:** *The portion of a teacher's salary due to degree attainment.* District salary schedules for 2013-14 and teacher level information were used to determine this cost. The cost is calculated by taking what each district actually spends on teachers' salaries and subtracting what they would spend if they did not pay teachers more for advanced degrees. Increases due to years of experience are not included.

**Performance Bonuses:** *Monetary rewards for teacher performance*

<sup>1</sup>Education Resource Strategies. (2013). *Professional Growth & Support Spending Calculator*. Retrieved from <http://www.erstrategies.org/cms/files/1782-gates-pgs-calculator-doc.pdf>

# APPENDIX B

## Overview of the Development Profile Analysis

The Development Profile Analysis linked performance data to survey data and other available teacher- and school-level information to compare teachers who improved to those who did not improve over time. This analysis was conducted at the teacher and school level and investigated potential differences around teacher experiences, mindsets and environments. See Technical Appendix: Appendices B1 to B4 for findings from the Development Profile Analysis and additional details on the variables and constructs investigated.

**Experiences:** Teacher self-reports from the survey regarding the frequency with which they engaged in various professional development activities during the 2012-13 and 2013-14 school year were used to investigate relationships to teacher improvement. Additionally, to investigate potential differences that might emerge from early career support, teachers in their first two years of experience when the survey was administered were asked about their experiences with teacher preparation and mentoring.

**Mindsets:** Teacher self-perceptions of their practice, growth and self-efforts they engage in for their development were used to investigate potential differences in mindsets between improvers and non-improvers.

**Environments:** Teacher environments and their perceptions of their environments were investigated using a combination of teacher survey data, leader survey data and extant data to look for potential differences between teachers identified as improvers and non-improvers.

## B1. DEVELOPMENT PROFILE SIMILARITIES FOR IMPROVERS AND NON-IMPROVERS

Few differences emerged between improvers and non-improvers in the Development Profile Analysis. The table below contains additional details on the survey questions, percentages and Ns from this analysis as reported in the paper. The Fixed-Split Standardized Evaluation definition of growth is used to present results.

FREQUENCY OF DEVELOPMENT ACTIVITIES		Improvers	Non-Improvers
Hours of Development Activities		Hours	
Receiving direct coaching from an assigned district or school-level staff member (e.g., individualized support in my classroom with feedback and/or modeling of techniques, etc.) (two years)		12.43 Hours (n=1,250)	12.66 Hours (n=1,067)
Formally meeting with small teacher teams in my school for support (e.g., PLCs or other formally organized small groups) (two years)		69.41 Hours (n=1,259)	64.02 Hours (n=1,072)
About how many hours in a given month, on average, do you spend engaged in some sort of professional development activity: a. Organized/run by your district; b. Organized/run by your school; c. You pursued independently. (2013-14)		16.86 Hours a Month (n=1,467)	18.01 Hours a Month (n=1,212)
Participating in extended professional development programs (e.g., a focused series including multiple sessions and ongoing support throughout the year). (*Responses were quartiled to investigate percentages of improvers and non-improvers at the extreme ends.) (two years)		24.17% Top/ 25.52% Bottom (n=1,258)	23.06% Top/ 26.89% Bottom (n=1,071)
Number of Observations		Observations	
Please indicate how many classroom observations you received from a formal evaluator (e.g., a person who has an impact on your final evaluation rating) in each of the years listed below. Please include observations of any length. (two years)		7.58 Observations (n=1,267)	7.36 Observations (n=1,057)
Combination of Experiences		Percent ≥ Median Hours	
A group of teachers was identified who reported receiving the median or above of support relative to other teachers across multiple activities in 2013-14 including: Extended Professional Development, Formal Collaboration, Coaching, Observations and Feedback. When looking at the percentage of improvers and non-improvers who fell into this group, results were even. (296 teachers were captured in this group across all three districts).		13.60% (n=1,154)	14.18% (n=980)

SATISFACTION WITH DEVELOPMENT EXPERIENCES		Improvers	Non-Improvers
Overall Satisfaction		% Yes	
Are you satisfied, overall, with the professional development you receive from your school and district?		67.19% (n=1,524)	65.36% (n=1,273)
Detailed Satisfaction		% Strongly Agree or Agree	
The majority of the professional development I receive from my school and district drives lasting improvements to my instructional practice. *Note: This result is statistically significant at p<.05 on this growth flag, but results are not consistent across sites or definitions of growth.		51.96% (n=1,559)	47.91% (n=1,315)
The majority of the professional development I receive from my school and district is targeted to support my specific teaching context.		50.16% (n=1,569)	47.84% (n=1,319)
The majority of the professional development I receive from my school and district is a good use of my time. *Note: This result is statistically significant at p<.05 on this growth flag, but results are not consistent across sites or definitions of growth.		43.95% (n=1,561)	39.64% (n=1,317)

MINDSETS		Improvers	Non-Improvers
Role of Feedback and Weaknesses in Instruction		% Strongly Agree or Agree	
Receiving feedback on instructional practice plays a crucial role in improving teacher practice. *Note: This result is statistically significant at p<.01 on this growth flag, but results are not consistent across sites or definitions of growth.		78.55% (n=1,450)	73.94% (n=1,205)
I have weaknesses in my instruction. *Note: This result is statistically significant at p<.05 on this growth flag, but results are not consistent across sites or definitions of growth.		39.89% (n=1,439)	44.65% (n=1,205)
Teacher Responsibility		% Individual Teacher	
In your opinion, who should bear the greatest responsibility for improving teacher instructional practice? (Teacher preparation programs (undergraduate or graduate), Central district staff (coaches, mentors and professional development facilitators, etc.), School leaders, In-school teacher-leaders (coaches, mentors, content specialists, etc.), Individual teachers)		40.58% (n=1,417)	39.73% (n=1,168)
Reflection on Instructional Practice		% Daily Reflection	
How frequently do you reflect on your instructional practice? (Never, Once a year, Once a semester, Monthly, Weekly, Daily) *Note: This result is statistically significant at p<.05 on this growth flag, but results are not consistent across sites or definitions of growth.		75.53% (n=1,459)	70.81% (n=1,206)

## B2. DEVELOPMENT PROFILE DIFFERENCES BETWEEN IMPROVERS AND NON-IMPROVERS

**Teacher-Level Analysis Findings:** Teacher-level regression models, run separately for each district, indicated that increasingly positive teacher responses on four variables—Openness to Feedback, Evaluator Quality, School Support Structure and Rating Alignment Scale\*—were associated with small increases in observation scores, standardized evaluation scores and/or value-added scores. Each model controlled for prior performance and years of teaching experience. See Appendix Table B3 for additional details on the construction of these variables.

**Observation Scores.** Across a series of linear regression models, four predictors were significantly related to increases in teacher observation scores: Openness to Feedback, Evaluator Quality, School Support Structure and Rating Alignment Scale. The number of teachers contributing data varied across models. In Districts A and C, between 1,500 and 2,700 teachers were included. In District B, between 200 and 400 teachers were included across models.

For every one-point increase on our Openness to Feedback measure, observation scores could be expected to increase by 0.72 points in District A ( $p<.001$ ), 0.04 points in District B ( $p<.05$ ) and 0.04 points in District C ( $p<.001$ ). The more positively teachers rated the quality of their evaluators, the more their observation scores increased. A one-unit increase in the evaluator quality construct was associated with observation score increases of 0.74 points in District A ( $p<.001$ ), 0.08 points in District B ( $p<.001$ ) and 0.10 points in District C ( $p<.001$ ). As teachers provide more positive responses on the school support structure index, observation scores could be expected to increase by 0.33 points in District A ( $p<.001$ ), 0.04 points in District B ( $p<.01$ ) and 0.03 points in District C ( $p<.001$ ). Finally, as teachers reported ratings which were more aligned to the formal assessment of their practice in 2013-14, observation scores were expected to increase by 2.49 points in District A ( $p<.001$ ), 0.17 points in District B ( $p<.001$ ) and 0.05 points in District C ( $p<.001$ ).

**Standardized Evaluation Scores.** Approximately the same number of teachers were included in these models as were included in models predicting observation scores. In these models, two variables were significantly related to evaluation scores: Evaluator Quality and Rating Alignment Scale.

A one-unit increase in teacher perceptions of evaluator quality was associated with an increase in standardized evaluation ratings of 0.09 standard deviations in District A ( $p<.001$ ), 0.17 standard deviations in District B ( $p<.001$ ) and 0.07 standard deviations in District C ( $p<.001$ ). Rating alignment was also significantly related to increases in standardized evaluation scores; as teachers reported ratings more aligned to the formal assessment of their practice in 2013-14, standardized evaluation scores could be expected to increase by 0.33 standard deviations in District A ( $p<.001$ ), 0.47 standard deviations in District B ( $p<.001$ ), and 0.53 standard deviations in District C ( $p<.001$ ).

**Value-added Scores.** Notably, only two districts had enough teachers with value-added scores and survey data to conduct these regressions (in District A, roughly 2,200 teachers contributed data, and in District C, roughly 450 teachers are included). In these models, rating alignment was the only significant predictor. As teachers reported ratings more aligned to the formal assessment of their practice in 2013-14, value-added scores were expected to increase by 0.54 points in District A ( $p<.001$ ) and 0.99 points in District C ( $p<.001$ ).

*\*Note: All teachers who received the highest rating in 2013-14 in each site were removed from the analysis to look more specifically at teachers not already identified as the highest performers.*

**School-Level Analysis Findings:** School-level regression models, run with all districts pooled, indicated that increasingly positive teacher responses (aggregated to the school level) on two variables—Average Number of Observations and Rating Alignment\*—were associated with a small increase in the percent of improvers at a school. Each model included a thematically related subset of variables constructed by aggregating individual teacher survey responses to the school level, as well as controls related to school demographics and aggregate teacher demographics. See Appendix Tables B3 and B4 for additional details on the construction of these variables.

**Percent of teachers improving on observation scores.** There were approximately 370 schools included in regression models predicting the percent of teachers in a school improving on observation scores (using the “quartiles of growth” definition). For every increase in the average number of observations reported by teachers in a school, the percent of teachers identified as improvers at the school was expected to go up by 3% ( $p<.05$ ). When considering teachers’ self-reported evaluation scores as compared to the formal assessments of their practice in 2013-14, for every one-unit increase in school alignment scores, the percent of teachers identified as improvers at a school was expected to increase by 10% ( $p<.01$ ).

**Percent of teachers improving on standardized evaluation scores.** There were approximately 370 schools included in regression models predicting the percent of teachers in a school improving on standardized evaluation scores (using the “quartiles of growth” or “fixed-split growth” definition). For every addition to the average number of observations reported by teachers in a school, the percent of teachers identified as improvers at the school was expected to go up by 3% ( $p<.05$ ) or 2% ( $p<.05$ ) using “quartiles of growth” and “fixed-split growth,” respectively. When considering teachers’ self-reported evaluation scores as compared to the formal assessments of their practice in 2013-14, for every one-unit increase in school alignment scores, the percent of teachers identified as improvers at a school was expected to increase by 28% ( $p<.01$ ) or 25% ( $p<.01$ ) using “quartiles of growth” and “fixed-split growth,” respectively.

**Percent of teachers improving on value-added scores.** There were approximately 200 schools included in regression models predicting the percent of teachers in a school improving on value-added scores (using the “quartiles of growth” definition or the “fixed-split growth” definition). Only District A and C were included in the VAM analysis due to sample size limitations at the school level in District B. For every additional observation reported by teachers in a school on average, the percent of teachers identified as improvers at the school was expected to go up by 3% ( $p<.05$ ), using “quartiles of growth.” As teachers at a school, on average, self-report ratings more aligned to or deflated in relation to the formal assessments of their practice in 2013-14, the percent of teachers identified as improvers at a school was expected to increase by 10% ( $p<.05$ ), using “fixed-split growth.”

*\*Note: All teachers who received the highest rating in 2013-14 in each site were removed from the analysis to look more specifically at teachers not already identified as the highest performers.*

### B3. SURVEY ITEMS USED TO COMPARE IMPROVERS TO NON-IMPROVERS AT THE TEACHER AND SCHOOL LEVEL

These items were tested at the individual teacher level and in the school-level analysis.

EXPERIENCES	SURVEY QUESTIONS AND CONSTRUCT DETAILS	VARIABLE CALCULATION
One-time PD	Attending one-time professional development sessions or meetings (e.g., in-person or online run by your district, school, or a vendor)	0-200 hours (continuous scale)
Extended PD	Participating in extended professional development programs (e.g., a focused series including multiple sessions and ongoing support)	
Independent Efforts	Engaging in independent efforts to improve my instruction (e.g., researching strategies or content, testing strategies, studying student data, watching my practice via video, etc.)	
Formal Collaboration	Formally meeting with small teacher teams in my school for support (e.g., PLCs or other formally organized small groups)	
Informal Collaboration	Spending time with colleagues (e.g., informal time you set aside to discuss content, data, instruction, etc., but not a formal coaching or small group relationship)	
Time with Evaluator	Spending time with my formal evaluator (e.g., discussing my instructional practice, reviewing student data, etc.)	
Direct Coaching	Receiving direct coaching from an assigned district or school-level staff member (e.g., individualized support in my classroom with feedback and/or modeling of techniques, etc.)	
University Courses or Certifications	Completing university level coursework (e.g., to earn additional salary credits, degrees, or certifications, etc.)	
Peer Observations	Number of instances in 2012-13 + 2013-14	0-20 instances (continuous scale)
Observations	Number of instances in 2012-13 + 2013-14	
Feedback	Number of instances in 2012-13 + 2013-14	
Receiving Follow-up	Receive follow-up support to ensure I am implementing new instructional practices effectively. <i>Scale: Often, Sometimes, Rarely, Never</i>	Categorical Frequency
Outside Practice Time	Have the opportunity to practice teaching techniques in a setting outside my classroom before using them with my students. <i>Scale: Often, Sometimes, Rarely, or Never</i>	
Job-Embedded PD	Direct Coaching + Time with Formal Evaluator	Sum of all activity hours from 2012-13 and 2013-14
Combined PD	Extended PD Programs + Independent Efforts + Informal Collaboration	
Peer Time	Formal Collaboration + Informal Collaboration + Peer Observations	
Practice Opportunities	Receive follow-up support to ensure I am implementing new instructional practices effectively; Have the opportunity to practice teaching techniques in a setting outside my classroom before using them with my students. <i>Scale: Often, Sometimes, Rarely, or Never</i>	Mean of Two Questions
Total Hours	Total Hours of Individual Activities from 2012-13 + 2013-14	Sum of Hours
Total Hours a Month	How many hours of district, school and independent PD are you engaged in during one month?	Numeric Responses

EARLY CAREER SUPPORT	SURVEY QUESTIONS AND CONSTRUCT DETAILS	VARIABLE CALCULATION
Certification	Please select the kind of program through which you were certified? Traditional / Alternative certification program	Binary Variable
Classroom Practice	Approximately how much time did you spend practicing teaching in a classroom throughout your teacher preparation program prior to starting your first year of teaching? <i>Scale: My preparation program did not include classroom practice, 4 weeks, 5-8 weeks, 9-12 weeks, 1 semester, More than 1 semester, A full year, More than a full year</i>	Categorical Frequency
Outside Practice	Approximately how often were you able to practice teaching outside of the live classroom environment throughout your teacher preparation program (e.g., presenting a lesson or practicing a certain skill with a mentor or professor)? <i>Scale: My preparation program did not include this kind of practice opportunity, Once a year, Once every few months, Once a month, Once a week or more</i>	Categorical Frequency
Preparation Practice Total	Combination of Classroom Practice and Outside Practice	Mean of Two Questions
Teacher Readiness	From the list below, please place a check beside all the areas where you feel you were NOT prepared to perform well in your first year of teaching. <i>List of classroom practice competencies provided to check all that apply.</i>	Count of all areas listed
Preparation Quality	My teacher preparation program included sufficient classroom practice opportunities for me to master the basic skills I needed to be a teacher. / My teacher preparation program prepared me to be effective in the classroom in my first year of teaching. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i>	Mean of Two Questions
Mentor Provided	In your FIRST year of teaching, did you work with a mentor teacher (i.e., person assigned to provide you support during your first year of teaching) who was assigned by your school or district? If you are in your first year of teaching, please answer for this school year.	Binary Variable
Mentor Frequency	How frequently did you work with your mentor teacher during your first year of teaching? <i>Scale: Never, A few times a year, Once or twice a month, At least once a week</i>	Categorical Frequency
Mentor Impact	Overall, to what extent did your mentor teacher improve your teaching in your first year of teaching? <i>Scale: Not at all, To a small extent, To a moderate extent, To a great extent</i>	Likert Scale

MINDSETS	SURVEY QUESTIONS AND CONSTRUCT DETAILS	VARIABLE CALCULATION
Teacher Responsibility for Development	In your opinion, who <b>should</b> bear the greatest responsibility for improving teacher instructional practice? <i>Teacher preparation programs (undergraduate or graduate), Central district staff (coaches, mentors and professional development facilitators, etc.), School leaders, In-school teacher-leaders (coaches, mentors, content specialists, etc.), Individual teachers</i>	Binary Variable: <i>Teacher is Responsible vs Other</i>
Admits to Having Weaknesses	I have weaknesses in my instruction. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i>	Likert Scale
Learning/Growth Mindset	I believe I have more to learn as a teacher. / I have weaknesses in my instruction. / I have a clear understanding of my instructional practice strengths and weaknesses. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i>	Construct created with exploratory factor analysis (range of scores: -5.41 to 1.04)
Self-Effort	How frequently do you: Reflect on your instructional practice / Try new teaching strategies in your classroom / Seek out resources to help you grow / Meet with teachers throughout your school or district who teach in your same grade or subject to plan and share resources. <i>Scale: Never, Once a year, Once a semester, Monthly, Weekly, Daily</i>	Mean Across Variables
Open to Feedback	Receiving feedback on instructional practice plays a crucial role in improving teacher practice. / Receiving performance evaluation ratings plays a crucial role in improving teacher practice. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i> How effective do you believe receiving frequent and honest feedback against clear performance standards is for improving your instructional practice? <i>Scale: Very effective, Effective, Somewhat effective, Somewhat ineffective, Ineffective, Very ineffective</i>	Construct created with exploratory factor analysis (range of scores: -2.92 to 1.51)
Driver of Own Development	Strongly Agree or Agree: I have a clear understanding of my instructional practice strengths and weaknesses. At Least Weekly: Seek out resources to help you grow Individual Teacher: In your opinion, who should bear the greatest responsibility for improving teacher instructional practice? Myself: If you had to pick the person/group of people who have been most instrumental in improving your instructional practice over the course of your career, who would it be?	Additive combination of responses; range of scores 0 to 4
Change in Status Quo	Strongly Agree or Agree: I have weaknesses in my instruction. Strongly Agree or Agree: The Common Core Standards are an important and positive change for teachers and students. Self-Improvement: Indicate they have "Improved Some", "Stayed the Same" or "Declined" Self-Rating: Rates self as a 4 or less on the 5 point scale. Somewhat Agree or Less: There are teachers at my school who set an example for highly effective teaching Somewhat Agree or Less: The majority of the professional development I receive from my school and district: 1) Drives lasting improvements to my instructional practice 2) Drives lasting changes in my student learning outcomes	Additive combination of responses; range of scores 0 to 7
External Assessments	Teachers indicate that: Anyone can assess me as long as they have knowledge. Strongly Agree or Agree: Receiving feedback on instructional practice plays a crucial role in improving teacher practice. Strongly Agree or Agree: Receiving performance evaluation ratings plays a crucial role in improving teacher practice. How do you know you have improved: The feedback I get through my performance evaluation has improved or Others have told me that I am improving (e.g., formal evaluators, peer teachers, students, etc.).	Additive combination of responses; range of scores 0 to 4
Rating Alignment Scale	Using teacher ratings from 2013-14, a teacher is given a score of 1 to 5, based on how aligned they are to this rating in their self-assessment. A teacher is given a 5 if they are aligned, a 4 if they are off by 1, a 3 if they are off by 2, a 2 if they are off by 3, and a 1 if they are off by 4.	Categorical Variable Created (1 to 5)
Rating Inflation, Alignment and Deflation	Using teacher ratings from 2013-14, a teacher is given a score of 1 if they inflate their self-assessment of practice, a 2 if they are aligned exactly, and a 3 if they deflate their assessment of their own practice relative to their actual performance rating in 2013-14.	Categorical Variable Created (1,2,3)

ENVIRONMENTS	SURVEY QUESTIONS AND CONSTRUCT DETAILS	VARIABLE CALCULATION
Perceptions of Evaluator Quality	My formal evaluator has an accurate understanding of my instructional strengths and development areas. / My formal evaluator is able to direct me to development opportunities aligned with my needs. / My formal evaluator has communicated my instructional practice strengths and weaknesses to me. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i>	Construct created with exploratory factor analysis (range of scores: -2.91 to 1.59)
Data Culture	My school uses the results of student assessments to make decisions about how to provide targeted support to teachers. / My school uses the results from teacher evaluations to make decisions about how to provide targeted support to teachers. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i>	Mean of Two Questions
School Support Structure (Construct)	<p>The expectations for effective teaching are clearly defined at my school. / My school uses the results of student assessments to make decisions about how to provide targeted support to teachers. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i></p> <p>Spending time with my formal evaluator (e.g., getting feedback on my performance, reviewing student data, etc.) <i>Scale: Very effective, Effective, Somewhat effective, Somewhat ineffective, Ineffective, Very ineffective</i></p>	Construct created with exploratory factor analysis (range of scores: -3.40 to 1.40)
School Support Structure (Index)	<p>Strongly Agree or Agree: The expectations for effective teaching are clearly defined at my school. / My school uses the results of student assessments to make decisions about how to provide targeted support to teachers. / Teachers in my school have time to visit each other's classrooms (e.g., to observe highly effective practice or provide feedback and support). / My school has the resources it needs to allow teachers additional flexibility during the day to focus on their development.</p> <p>Very Effective or Effective: Spending time with my formal evaluator (e.g., getting feedback on my performance, reviewing student data, etc.)</p>	Additive combination of responses; range of scores 0 to 5
Performance/Strong Leadership Culture	<p>Strongly Agree or Agree: There is a low tolerance for ineffective teaching at my school.</p> <p>Leader Responsibility: In your opinion, who should bear the greatest responsibility for improving teacher instructional practice?</p> <p>Very Effective or Effective: Spending time with my formal evaluator (e.g., getting feedback on my performance, reviewing student data, etc.)</p> <p>Teacher "Yes": The area of development they identified is aligned to what they have heard from their evaluator this year.</p>	Additive combination of responses; range of scores 0 to 4

#### B4. ADDITIONAL INVESTIGATIONS IN THE SCHOOL-LEVEL ANALYSIS

The below table contains the additional variables investigated in the school-level analysis beyond the items in Appendix B3. All survey items were averaged at the school level.

SCHOOL ITEMS	SURVEY QUESTIONS AND CONSTRUCT DETAILS	VARIABLE CALCULATION
Instructional Culture Index (ICI)	Teachers at my school share a common vision of what effective teaching looks like./ The expectations for effective teaching are clearly defined at my school./ My school is committed to improving my instructional practice. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i>	Additive combination of responses; range of scores 1 to 10
School Characteristics	Items include: Teacher attrition from 2012-13 to 2013-14, percent of minority students in 2013-14, total enrollment in 2013-14, percent of teachers with 1 to 2 years of experience in 2012-13, a teacher being in the same school in both 2012-13 and 2013-14, having the same principal in a school in both 2012-13 and 2013-14, and school-level student proficiency rates from 2011-12 to 2013-14	School Level Percentages
School Leader Confidence	Please indicate your level of confidence in your ability to effectively implement the following. (For the purposes of this question, please do not consider time as a factor but rather your confidence level in carrying out these responsibilities.) Assigning accurate observation ratings to teachers based on evidence from classroom observations/ Delivering feedback that helps teachers improve instructional practice/ Identifying meaningful professional development opportunities for teachers based on their specific needs or content area/ Developing and facilitating meaningful professional development opportunities for teachers based on their specific needs or content area/ Discussing student data with teachers and helping them plan accordingly/ Following up with teachers after professional development has been conducted to assess if they are using new strategies. <i>Scale: Very confident, Confident, Somewhat confident, Not very confident, Not confident, Not at all confident</i>	Mean of Six Questions
School Leader District Support Perceptions	I feel supported by my district to prioritize teacher development as one of my main areas of focus as a school leader./ My district provides me with the skills and knowledge I need to help my teachers improve their instructional practice. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i>	Mean of Two Questions
School Leader PD Spending Control	My school currently spends money on the kinds of professional development activities that make lasting improvements to teacher instructional practice. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree, N/A- My school does not have control over our professional development budget.</i>	Six-point Likert Agreement Scale with an N/A option
Teacher and Leader Survey Congruence	<p>The average responses to the following survey questions were compared between the teacher and school leader surveys at the school level:</p> <ol style="list-style-type: none"> <li>1) Are you satisfied, overall, with the professional development you receive from your school and district? Yes/No</li> <li>2) The majority of the professional development I receive from my school and district: <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i> <ol style="list-style-type: none"> <li>a. Drives lasting improvements to my instructional practice.</li> <li>b. Drives lasting changes in my student learning outcomes.</li> <li>c. Is targeted to support my specific teaching context.</li> </ol> </li> <li>3) How tailored is the professional development you receive from your school to the specific areas of development in your instructional practice? <i>Scale: Very tailored, Tailored, Somewhat tailored, Not very tailored, Not tailored, Not at all tailored</i></li> <li>4) Receiving feedback on instructional practice plays a crucial role in improving teacher practice. <i>Scale: Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree</i></li> <li>5) Please indicate how effective you believe the following activities are for making lasting improvements to your instructional practice. <i>Scale: Very effective, Effective, Somewhat effective, Somewhat ineffective, Ineffective, Very ineffective</i> <ol style="list-style-type: none"> <li>a. Formally meeting with small teacher teams in my school for support (e.g., PLCs or other formally organized small groups)</li> <li>b. Spending time with colleagues (e.g., informal time you set aside to discuss content, data, instruction, etc., but not a formal coaching or small group relationship)</li> </ol> </li> <li>6) In thinking about your professional development, how often do you: Have a requirement to attend a session on a topic or skill in which you are already competent or aware of? <i>Scale: Often, Sometimes, Rarely, Never</i></li> </ol>	Variable Created That Is the Difference Between Mean Teacher and Leader Responses to Each Question at the School Level

# ENDNOTES

<sup>1</sup>See for example: Chetty, R., Friedman, J., & Rockoff, J. (2011). *The Long Term Impacts of Teachers: Teacher Value-added and Student Outcomes in Adulthood*. (NBER Working Paper No. 17699). Cambridge, MA: National Bureau of Economic Research; Aaronson, D., Barrow, L., & Sanders, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, Volume 25(1)*, 95-135; Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, Volume 73(2)*, 417-458; Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, Volume 94*, 247-252.

<sup>2</sup>There are many reports, papers and op-eds that could be cited. The following is just a sampling, meant not to call attention to one organization or person over any others: Archibald, S., Coggshall, J., Croft, A., & Goe, L. (2011). *High Quality Professional Development for All Teachers: Effectively Allocating Resources*. Washington, DC, National Comprehensive Center for Teacher Quality; Berry, B. (2014, November 19) Deja Vu in American education: The woeful state of professional development. Retrieved from: <http://www.teachingquality.org/content/blogs/barnett-berry/d%C3%A9j%C3%A0-vu-american-education-woeful-state-professional-development>; Gulamhussein, A. (2013). *Teaching the Teachers: Effective Professional Development in an Era of High Stakes Accountability*. Alexandria, VA: Center for Public Education. Learning Forward. (2015, March 17); PD Brain Trust Wants your Input on Professional Learning Redesign. *Education Week*. Retrieved from: [http://blogs.edweek.org/edweek/learning\\_forwards\\_pd\\_watch/2015/03/pd\\_brain\\_trust\\_wants\\_your\\_input\\_on\\_professional\\_learning\\_redesign.html](http://blogs.edweek.org/edweek/learning_forwards_pd_watch/2015/03/pd_brain_trust_wants_your_input_on_professional_learning_redesign.html); Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional*

*development in the United States: Trends and Challenges*. Dallas, TX: National Staff Development Council.

<sup>3</sup>The average cost per teacher across Districts A, B and C using the Medium tier estimate is \$17,811.83.

<sup>4</sup>The sum of the total cost of transportation, food services and security from the fiscal year 2014 budget in District B was compared to the Low tier teacher improvement cost.

<sup>5</sup>This analysis is based on the 2011-12 ranking of the 50 largest school districts in the nation by student enrollment (most recent year available). National Center for Education Statistics. (2012). *Table 215.10: Selected statistics on enrollment, teachers, dropouts, and graduates in public school districts enrolling more than 15,000 students: Selected years, 1990 through 2011*. Retrieved from [http://nces.ed.gov/programs/digest/d13/tables/dt13\\_215.10.asp](http://nces.ed.gov/programs/digest/d13/tables/dt13_215.10.asp); United States Census Bureau. (2012). *Public Elementary-Secondary Education Finance Data*. Retrieved from <http://www.census.gov/govs/school/>

<sup>6</sup>These calculations use average "hours a month" of support from the Teacher Survey: About how many hours in a given month, on average, do you spend engaged in some sort of professional development activity: a. Organized/run by your district; b. Organized/run by your school; c. You pursued independently. Total Average Hours a Month=16.60 (n=9,075). Assuming nine months in a school year, an eight-hour teacher workday and 198 days in a school year, this results in 9.43% of the year and 149.39 hours. These numbers represent District A, B and C combined.

<sup>7</sup>74.14% of teachers in District A (n=8,724) and 56.95% of teachers in District B (n=1,812) did not improve their evaluation rating from 2011 to 2013; 63.06% of teachers in District C (n=4,044) did not improve their evaluation rating from 2012 to 2013. These percentages are based only on teachers with evaluation ratings in all indicated years but exclude teachers who

earned the highest possible evaluation rating in both years.

<sup>8</sup>Because we cannot identify years of teaching experience past year 10 in District B, this district is excluded from the analysis. However, results held when we used years of district experience instead. Sample sizes varied by experience and district but were always above 250.

<sup>9</sup>These percentages are 51.52% in District A (n=5,765), 53.11% in District B (n=1,654) and 45.99% in District C (n=3,540). See Technical Appendix: Analysis for definition of "effective."

<sup>10</sup>See Technical Appendix: Analysis for definitions of growth and analysis approach. See Technical Appendix: Appendix B for detailed outcomes and variable definitions.

<sup>11</sup>All districts use a 5-point final evaluation rating scale. For Districts A and C, the bar for Effective or Meeting Expectations includes teachers in the top three rating categories. For District B, this includes the top two categories.

<sup>12</sup>Teacher Survey: I have weaknesses in my instruction. (Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree). 46.82% Strongly agree or Agree (n=9,003)

<sup>13</sup>Teacher Survey: How would you rate the current quality of your instructional practice with 1 being Ineffective and 5 being Highly Effective? (Please note that these categories do not need to directly align with the rating scale in your district.) (1 (Ineffective), 2, 3, 4, 5 (Highly Effective)). All districts use a 5-point final evaluation rating scale. For Districts A and C, "low rated" teachers include the bottom two rating categories. For District B, this includes the bottom three rating categories. 62.14% of "low rated teachers" selected 4 or 5 (n=8,798)

<sup>14</sup>Teacher Survey: Are you satisfied, overall, with the professional development you receive from your school and district? (Yes/No). 67.47% Yes (n=9,567)

<sup>15</sup>Teacher Survey: The majority of the professional development I receive from my school and

district is a good use of my time. (Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree). 41.45% Strongly agree or Agree (n=9,799)

<sup>16</sup>Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement* (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2010). *Middle School Mathematics Professional Development Impact Study: Findings After the First Year of Implementation* (NCEE 2010-4009). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

See also: Arens, S. A., Stoker, G., Barker, J., Shebby, S., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2012). *Effects of curriculum and teacher professional development on the language proficiency of elementary English language learner students in the Central Region*. (NCEE 2012-4013). Denver, CO: Mid-continent Research for Education Learning; Bos, J., Sanchez, R., Tseng, F., Rayyes, N., Ortiz, L., & Sinicropi, C. (2012). *Evaluation of Quality Teaching for English Learners (QTEL) Professional Development*. (NCEE 2012-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

<sup>17</sup>See for example: Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches*. (REL 2014-010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance,

Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>; Hill, H. C., Beisiegel, M., & Jacob, R. (2013). *Professional Development Research: Consensus, Crossroads, and Challenges*. Educational Researcher; Suk Yoon, K., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007-No. 033)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

<sup>18</sup>The annual operating budgets for fiscal year 2014 were provided by each district.

<sup>19</sup>Demographic information represents data available on district or state websites from 2013-14. See Technical Appendix: Data for additional details.

<sup>20</sup>See Technical Appendix: Appendix B3 and B4 for a full description of the experiences, mindsets and environment variables investigated.

<sup>21</sup>Ibid Endnote 3

<sup>22</sup>Based on the Medium tier teacher improvement cost and total fiscal year 2014 budget, District A spent 5.91%, District B spent 8.94% and District C spent 8.88% of its budget on teacher improvement.

<sup>23</sup>Ibid Endnote 5

<sup>24</sup>Ibid Endnote 6

<sup>25</sup>Ibid Endnote 6

<sup>26</sup>Teachers with one to two years of experience reported 13.17 hours of instructional coaching in 2013-14 while teachers with 10 or more years reported 5.09 hours a year ( $p<.001$ ). Teachers with three to five years of experience reported statistically significantly more hours than teachers with 10 or more years of experience ( $p<.01$ ), but the difference is greatly diminished: 6.99 hours versus 5.09 hours

( $n=7,511$ ). See Technical Appendix: Appendix B3 for details on survey items used in this analysis.

<sup>27</sup>No statistically significant differences emerged in average hours of extended professional development workshops between teacher experience groups in 2013-14. For formal collaboration, teachers with one to two years of experience reported slightly fewer hours relative to other experience groups (27.98 hours compared to between 32.69 to 35.72 hours), ( $n=7,560$ ,  $p<.001$ ). For peer observations, teachers with one to two years of experience reported 2.19 instances a year, compared to between 1.32 to 1.39 for other experience groups ( $n=7,532$ ,  $p<.001$ ). This is less than a one-observation difference between groups. See Technical Appendix: Appendix B3 for details on survey items used in this analysis.

<sup>28</sup>Ibid Endnote 6

<sup>29</sup>This calculation is based on 2013-14 teacher professional development course attendance data (through April 1, 2014), using only instruction-related courses from one of the districts studied.

<sup>30</sup>The Medium tier teacher improvement cost in District A is \$180,957,227.72, in District B is \$73,143,171.06 and in District C is \$145,775,188.41.

<sup>31</sup>Ibid Endnotes 3 and 22

<sup>32</sup>This finding uses the Low tier estimate from each site as a comparison to fiscal year 2014 expenditures on transportation and food services.

<sup>33</sup>These figures are the sum of the Medium tier central personnel, school personnel, teacher time on development and teacher salary investments as a percentage of the total Medium tier teacher improvement cost. These costs represent 77.30% of District A's Medium tier cost, 87.33% of District B's Medium tier cost and 79.62% of District C's Medium tier cost.

<sup>34</sup>Association for Talent Development. (2014). *2014 State of the Industry*; Training

Magazine. (2013). *2013 Training Industry Report*.

<sup>35</sup>Training Magazine defines "Total training spending" as "All training-related expenditures for the year, including training budgets, technology spending, and staff salaries." Training Magazine. (2013). *2013 Training Industry Report*, 22-23.

<sup>36</sup>To compare district teacher improvement costs to other industry reported training costs, a restricted district cost was calculated below our Low tier estimates. This cost only includes the Low tier central office personnel and non-personnel costs, school-level direct support personnel costs, the cost of school leader meetings with teachers for improvement (not evaluation related), and school-level non-personnel costs in each district. See also: Association for Talent Development. (2014). *2014 State of the Industry*; Training Magazine. (2013). *2013 Training Industry Report*.

<sup>37</sup>Sample sizes in Districts A, B and C are 9,789, 2,148 and 4,140, respectively. The percent of teachers who improved in Districts A, B and C are 29.56%, 37.48% and 32.63%, respectively; the percent who declined are 14.33%, 16.29% and 22.05%, respectively. Overall evaluation scores represent the final composite score calculated by each district. In all three districts, these composites represent weighted averages of classroom observations and (potentially) value-added data, student surveys, student achievement, professionalism, and other measures depending on the district and teacher. See Technical Appendix: Analysis for a description of how we classified annual changes in overall evaluation scores as improving or declining.

<sup>38</sup>We calculated the average evaluation and observation scores among all teachers who had evaluation results the past three years (two in District C). In Districts A and B, the average 2013-14 evaluation and observation scores were about 0.17 to 0.23 standard deviation units (based on the 2013-14 site-specific distribution of

evaluation scores among all teachers) higher than in 2011-12, for average growth rates between approximately 0.09 to 0.11 standard deviations per year. Some of the score improvement in District A was driven by changes to the weights assigned to classroom observations. In District C, 2013-14 evaluation and observation scores were less than 0.03 standard deviations higher. Sample sizes for evaluation score comparisons in Districts A, B and C were 9,403, 2,245 and 5,548, respectively.

<sup>39</sup>The sample size in District B is 1,248 and in District A is 1,094. "Not improving at all" represents the percent of teachers who had 2013-14 indicator scores that were equal to or lower than their 2011-12 score on the same indicator. See Technical Appendix: Analysis for description of "effective," "low" and "developing" ratings for instructional skills.

<sup>40</sup>See for example: Common Core State Standards: National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Washington, DC. Retrieved from [http://www.corestandards.org/wp-content/uploads/ELA\\_Standards.pdf](http://www.corestandards.org/wp-content/uploads/ELA_Standards.pdf); National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC. Retrieved from [http://www.corestandards.org/wp-content/uploads/Math\\_Standards.pdf](http://www.corestandards.org/wp-content/uploads/Math_Standards.pdf)

<sup>41</sup>In all three districts, first and second year teachers in 2011-12 (2012-13 in District C) had significantly higher ( $p<0.001$ ) overall evaluation scores in 2013-14 than in 2011-12 (2012-13 in District C). Only teachers who had evaluation results in both years were included.

<sup>42</sup>See for example: Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). *The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty*

schools. NBER Working Paper 14021; Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247-252; Ladd, H. F. & Sorensen L. C. (2014). Returns to teacher experience: Student achievement and motivation in middle school. CALDER Working Paper No. 112; Papay, J. P. & Kraft, M. A. (Forthcoming). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*.

<sup>43</sup>Sample sizes varied by district and experience level but never dropped below 60 for any point represented in Figure 4. See Technical Appendix: Analysis for description of how growth rates were calculated including a description of Figure 4, specifically.

<sup>44</sup>Given sample size restrictions, we could only compare VAM-based growth rates at different experience levels in Districts A and C.

<sup>45</sup>Ibid Endnote 42

<sup>46</sup>Sample sizes varied by district and experience band but never dropped below 385 for any point represented in Figure 5. See Technical Appendix: Analysis for the description of "pseudo returns to experience" and additional details on how Figure 5 was constructed.

<sup>47</sup>Ibid Endnote 40

<sup>48</sup>Sample sizes in Districts A, B and C are 5,765, 1,655 and 3,540, respectively. See Technical Appendix: Analysis for description of how "effective" was defined for specific instructional skills.

<sup>49</sup>See Technical Appendix: Analysis for description of how we projected the number of years it would take the average teacher to be "highly effective" in a core instructional skill if current trends continue. Sample sizes for these specific projections are 2,231 in District B, 5,124 in District C and 6,635 in District A.

<sup>50</sup>Proficiency rates are based on math and reading performance in grades 3 to 10, though some

districts and subjects only had test results through grade 8.

<sup>51</sup>For all teachers in District B linked to at least five student test scores, we calculated a proficiency rate in math and reading across all years. We then identified teachers in their sixth to ninth year of teaching in each year of data whose standardized evaluation score was a half a standard deviation or more better than the average standardized evaluation score among all teachers in this experience range in the same academic year. These teachers were labeled "Above Average." Teachers with scores within a half a standard deviation were labeled "Average." For math results there were 39 Above Average teachers and 46 Average teachers; for reading there were 53 Above Average teachers and 73 Average teachers. We then pooled across all years of results and calculated average teacher-level proficiency rates for these two groups of teachers. When comparing these two groups of teachers' average proficiency rates, we made no attempt to account for student background characteristics or other factors that are associated with student test performance and could vary by teacher.

<sup>52</sup>See Technical Appendix: Analysis for definitions of growth and Appendix B1 to B2 for a summary of the similarities and differences between improvers and non-improvers.

<sup>53</sup>The Fixed Split – Standardized Evaluation definition of growth was used to display results across Districts A, B and C combined. Improvers were in 488 out of 513 schools.

<sup>54</sup>See Technical Appendix: Appendix B3 for a full list of professional development activities investigated and B1 for full results on the activity similarities between improvers and non-improvers.

<sup>55</sup>In addition to these similarities, Districts B and C provided centrally available data on teacher coaching data in 2013-14. In District C, non-improvers were actually more likely to have received coaching than improvers, and in District B, improvers and non-improvers

were equally as likely to have received coaching and had a similar number of coaching sessions on average. In District B, 16.44% of improvers (n=590) and 19.66% of non-improvers (n=468) received coaching support, and in District C, 9.73% of improvers (n=1,388) and 22.41% of non-improvers (n=1,071) received coaching support ( $p<.001$ ). Additionally, in District C, where records also indicated the specific instructional skills in which coaching occurred, no more than 38.24% of teachers who received coaching support on a specific instructional skill in 2013-14 saw an improvement in their evaluation score on that instructional skill from 2012-13 to 2013-14. A larger percentage of teachers who did not receive coaching support on the same skill saw improvement in their evaluation score from year to year. Only teachers who had final evaluation scores on an instructional skill in both school years were included in the analysis by instructional skill (n=4,409).

<sup>56</sup>See for example Cohen, D. K. & Hill, H. C. (2001). *Learning Policy: When State Education Reform Works*. New Haven, CT: Yale University Press; Desimone, L. M., Porter, A. C., Garet M. S., Suk Yoon, K., & Birman, B. F. (2002). Effects of Professional Development on Teachers' Instruction: Results from a Three-Year Longitudinal Study. *Educational Evaluation and Policy Analysis*, Vol. 24, 81-112; Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Suk Yoon, K. (2001). What Makes Professional Development Effective? Results from a National Sample of Teachers.

*American Educational Research Journal*, Vol. 38, No. 4, 915-945; Supevitz, J., Mayer, D., and Kahle, J. (2000). Promoting Inquiry-Based Instructional Practice: The Longitudinal Impact of Professional Development in the Context of Systemic Reform. *Educational Policy*, Vol. 14 (3), 331-356; Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, Vol. 44, 921-958; Bill & Melinda Gates Foundation. (2014). *Teachers Know Best: Teachers' Views on Professional Development*; National Center for Literacy Education. (2014). *Remodeling Literacy Learning Together: Path to Standards Implementation*. National Council of Teachers of English.

<sup>57</sup>In addition to the data collected through our teacher survey, District A provided centrally available teacher survey data from 2013-14 that is collected following attendance in professional development sessions. When looking at the results between improvers and non-improvers, there were no statistically significant differences in the percent who strongly agree or agree with a question regarding the extent to which the content was appropriate to them. Improvers: 99.46% (n=1,493) and Non-improvers: 99.52% (n=1,467).

<sup>58</sup>Teacher Survey: Which of the activities helped you learn the most about how to improve your instructional practice during your teaching career? (n=1,831). Results are statistically significant at  $p<.001$ , but this trend does not hold across all three districts.

<sup>59</sup>Schools with at least five teachers in each district were used as the denominator.

<sup>60</sup>Teacher self-reported subject areas from the survey and school levels from district provided rosters were used to investigate proportional distribution alignment between the full population of teachers and the percent of improvers in each category.

<sup>61</sup>Kraft, M. A. & Papay, J. P. (2014). Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience. *Educational Evaluation and Policy Analysis*, Vol. 36, No. 4, 476-500. For additional research on culture and its impact see: Bryk, A. S. & Schneider, B. (2004). *Trust in Schools: A Core Resource for Improvement*. New York, NY.: Russell Sage Foundation.

<sup>62</sup>Across all three districts, teachers in their first two years grew 0.26 to 0.27 standard deviations per year on their overall evaluation score over the next two to three years.

<sup>63</sup>Teacher Survey: How frequently did you work with your mentor teacher during your first year of teaching? (*Never, A few times a year, Once or twice a month, At least once a week*). % At least once a week: District A: 76.21%, District B: 28.57% and District C: 41.25% (n=774).

<sup>64</sup>The development profile analysis was conducted separately for teachers with one to two years, three to five years, six to nine years and 10 or more years of experience. The trends remained consistent with the overall analysis findings. See Technical Appendix: Analysis for details on this analysis.

<sup>65</sup>See Technical Appendix: Analysis for additional details on the regression models for the development profile analysis.

<sup>66</sup>See Technical Appendix: B2 for detailed regressions findings at the teacher and school level.

<sup>67</sup>Teacher Survey: How would you rate the current quality of your instructional practice with 1 being Ineffective and 5 being Highly Effective? (Please note that these categories do not need to directly align with the rating scale in your district.) (1 (*Ineffective*), 2, 3, 4, 5 (*Highly Effective*)). District A: Improvers: 37.64% Inflated and 55.75% Aligned (n=348) / Non-improvers: 77.64% Inflated and 21.95% Aligned (n=483). District B: Improvers: 35.71% Inflated and 60.71% Aligned (n=28) / Non-improvers: 60.61% Inflated and 31.82% Aligned (n=66). District C: Improvers: 22.32% Inflated and 56.25% Aligned (n=112) / Non-improvers: 81.94% Inflated and 14.97% Aligned (n=648). This analysis excluded teachers who received the highest rating at the end of the 2013-14 school year.

<sup>68</sup>Teacher Survey: How would you rate the current quality of your instructional practice with 1 being Ineffective and 5 being Highly Effective? (Please note that these categories do not need to directly align with the

rating scale in your district.) (1 (*Ineffective*), 2, 3, 4, 5 (*Highly Effective*)). 83.17% selected 4 or 5 (n=9,015).

<sup>69</sup>Teacher Survey: I have weaknesses in my instruction. (*Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree*). 46.82% Strongly agree or Agree (n=9,003).

<sup>70</sup>Teacher Survey: Please select the statement that best describes the kind of change you have seen in your instructional practice since 2010-11. (If you have not been teaching since 2010-11, please just consider the current duration of your teaching career.) (*Declined, Remained relatively the same, Improved some, Improved tremendously*). 87.27% selected “Improved Some” or “Improved Tremendously” (n=9,034).

<sup>71</sup>Final evaluation rating files provided by each district for the 2013-14 school year were used.

<sup>72</sup>This data represents teachers who received 2013-14 evaluation ratings and for whom we had years of teaching experience. Where experience data was not available, years of experience as reported in the teacher survey was used. For Districts A and C, this includes the top three rating categories. For District B, this includes the top two categories.

<sup>73</sup>All districts use a 5-point final evaluation rating scale. This includes only the bottom two rating categories in each district.

<sup>74</sup>Teacher Survey: How would you rate the current quality of your instructional practice with 1 being Ineffective and 5 being Highly Effective? (Please note that these categories do not need to directly align with the rating scale in your district.) (1 (*Ineffective*), 2, 3, 4, 5 (*Highly Effective*)). 80.33% percent of teachers who had observation scores decline between the first and last years of our datasets report that their instructional practice has “Improved Some” or “Improved Tremendously” (n=5,893).

<sup>75</sup>Teacher Survey: The majority of the professional development I receive from my school and district is a good use of my

time. (*Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree*). 41.45% Strongly agree or Agree (n=9,799).

<sup>76</sup>Teacher Survey: The majority of the professional development I receive from my school and district drives lasting improvements to my instructional practice. (*Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree*). 50.58% Strongly agree or Agree (n=9,760).

<sup>77</sup>Teacher Survey: Are you satisfied, overall, with the professional development you receive from your school and district? (*Yes/No*). 67.47% selected Yes (n=9,567).

<sup>78</sup>Teacher Survey: The majority of the professional development I receive from my school and district: a. Is ongoing, with follow-up opportunities to review how effectively I am growing and receive additional support: 42.74% Strongly agree or Agree (n=9,801); b. Is tailored to my specific needs or development areas: 48.37% Strongly agree or Agree (n=9,843); c. Is targeted to support my specific teaching context (e.g., content area, the needs of the students in my classroom, etc.): 47.33% Strongly agree or Agree (n=9,811). (*Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree*). 36.19% Strongly agree or Agree (n=9,028).

<sup>81</sup>The exact number of peer observations in 2013-14 was 1.47, on average, across districts (n=7,705). 71.83% of teachers report that “Observing the classroom practice of teachers known for excellent instruction” is Very Effective or Effective for making lasting improvements to their instructional practice (n=5,225).

<sup>82</sup>The exact number of hours of one-time PD in 2013-14 was 23.55, on average, across districts (n=8,056). 36.47% of teachers report that “Attending one-time professional development sessions or meetings (e.g., in-person or online sessions run by your district, school, or a vendor)” is Very Effective or Effective for making lasting improvements to their instructional practice (n=7,554).

<sup>83</sup>Teacher Survey: Receiving performance evaluation ratings plays a crucial role in improving teacher practice. (*Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree*). 36.19% Strongly agree or Agree (n=9,028).

<sup>84</sup>Teacher Survey: My formal evaluator is able to direct me to development opportunities aligned with my needs. (*Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree*). 46.69% Strongly agree or Agree (n=7,441).

<sup>85</sup>Teachers were asked to select the skill in which they feel the least confident in their instructional practice and then were asked: “Does the development area you selected align with information you have received from your formal evaluator (e.g., person who has an impact on your final evaluation rating) in the past year (2012-13 to now)?” (Yes, No, or N/A - *My formal evaluator has not communicated any areas of development to me during this time*). 64.31% Yes, 27.71% No and 7.98% N/A (n=7,431).

<sup>86</sup>This information is based on interviews with district staff and principals and focus groups with teachers in Districts A, B and C.

<sup>87</sup>This is a quote from a district administrator interview.

<sup>88</sup>Sample sizes are 144 for the CMO and 9,420, 2,243 and 5,548 in Districts A, B and C, respectively.

<sup>89</sup>See Technical Appendix: Analysis for a description of how we standardized growth rates in order to compare rates across districts. Sample sizes varied by experience and district but ranged from 24 (for CMO teachers in their sixth year of teaching or beyond) to 6,677 (for District A teachers in their sixth year of teaching or beyond).

<sup>90</sup>See DeArmand, M., Gross, B., Bowen, M., Demeritt, A., & Lake, R. (2012). *Managing Talent for School Coherence: Learning from Charter Management Organizations*. Seattle, WA: Center on Reinventing Public Education for discussion on the role of coherence in CMO talent management more broadly.

<sup>91</sup>This quotation is from a focus group with CMO teachers.

<sup>92</sup>Teacher Survey: I have weaknesses in my instruction. (*Strongly agree, Agree, Somewhat agree, Somewhat disagree, Disagree, Strongly disagree*). Strongly agree or Agree: CMO: 81.22% (n=229); District A: 51.01% (n=3,729); District B: 59.97% (n=707); District C: 41.36% (n=4,567).

<sup>93</sup>Teacher Survey: How would you rate the current quality of your instructional practice with 1 being Ineffective and 5 being Highly Effective? (Please note that these categories do not need to directly align with the rating scale in your district.) (1 (*Ineffective*), 2, 3, 4, 5 (*Highly Effective*)). Provided a Rating of 5: CMO: 4.46% (n=224); District A: 23.60% (n=3,738); District B: 26.59% (n=707); District C: 35.84% (n=4,570).

<sup>94</sup>CMO school leaders reported statistically significantly lower levels of confidence in their abilities on all the following questions compared to district school leaders. School Leader Survey: Please indicate your level of confidence in your ability to effectively implement the following (for the purposes of this question, please do not consider time as a factor but rather your

confidence level in carrying out these responsibilities.): (*Very confident, Confident, Somewhat confident, Not very confident, Not confident, Not at all confident*). 1) Assigning accurate observation ratings to teachers based on evidence from classroom observations; 2) Delivering feedback that helps teachers improve instructional practice; 3) Identifying meaningful professional development opportunities for teachers based on their specific needs or content area; 4) Developing and facilitating meaningful professional development opportunities for teachers based on their specific needs or content area; 5) Discussing student data with teachers and helping them plan accordingly; 6) Following up with teachers after professional development has been conducted to assess if they are using new strategies.

<sup>95</sup>Teacher Survey: In thinking about your professional development, how often do you: Have the opportunity to practice teaching techniques in a setting outside my classroom before using them with my students? (*Often, Sometimes, Rarely or Never*). Often or Sometimes: CMO: 82.01% (n=239); District A: 27.40% (n=3,784); District B: 17.45% (n=762); District C: 37.64% (n=4,758).

<sup>96</sup>Teacher Survey: Please indicate how effective you believe the following activities are for making lasting improvements to your instructional practice: Receiving classroom observations with verbal and/or written feedback. (*Very effective, Effective, Somewhat effective, Somewhat ineffective, Ineffective, Very ineffective*). Very Effective or Effective: CMO: 65.22% (n=161); District A: 36.49% (n=3,217); District B: 45.78% (n=509); District C: 50.12% (n=3,755).

<sup>97</sup>Teacher Survey: About how many hours in a given month, on average, do you spend engaged in some sort of professional development activity: a. Organized/run by your district; b. Organized/run by your school; c. You pursued independently. Average Hours a Month: CMO:

22.39 (n=244); District A: 16.01 (n=3,702); District B: 18.74 (n=743); District C: 16.72 (n=4,630).

<sup>98</sup>The average Medium tier cost per teacher in the CMO is \$33,044.89. The average Medium tier cost per teacher across Districts A, B and C is \$17,811.83. Based on Medium tier teacher improvement costs and fiscal year 2014 budgets, the CMO spent 15.15% of its total operating budget on teacher improvement compared to 5.91% in District A, 8.94% in District B and 8.88% in District C.

<sup>99</sup>School leader time costs, including meetings with teachers for improvement (not evaluation related), strategy meetings for teacher improvement, teacher evaluation time costs and district-required time costs, represent 22.58% of the CMO's total Medium tier teacher

improvement cost compared to 2.43% in District A, 4.59% in District B and 5.36% in District C. School-level support personnel and teacher development-related substitute coverage costs represent 4.61% of the CMO's total Medium tier cost compared to 17.82% in District A, 18.17% in District B and 5.78% in District C. Teacher time on development costs represent 35.79% of the CMO's total Medium tier cost compared to 30.35% in District A, 25.65% in District B and 27.11% in District C. CMO teachers also spend anywhere from 3.49 to 6.68 times the number of hours in contracted time (professional development days and release time for professional development) than teachers in Districts A, B and C. See Technical Appendix: Appendix A for additional details on the approach to calculating costs associated with teacher support spending.

<sup>100</sup>Turnover is estimated by calculating the percent of teachers with an evaluation result in one year but not the next. Thus it does not capture teachers who stay with the district or CMO but move to non-teaching positions.

<sup>101</sup>TNTP. (2012). *The Irreplaceables: Understanding the Real Retention Crisis in America's Urban Schools*. Brooklyn, NY: TNTP.

<sup>102</sup>See for example: Lewin, K. (June 1947). "Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; Social Equilibria and Social Change" (PDF). *Human Relations*. Vol. 1, No. 1, 5-41.; Schein, E. (2010). *Organizational Culture and Leadership* (4th Edition). San Francisco, CA: Jossey-Bass.

<sup>103</sup>See for example: Dee, T. & Wyckoff, J. (2013). *Incentives, Selection, and Teacher Performance Evidence from IMPACT*. (NBER Working Paper No. 19529). Cambridge, MA: National Bureau of Economic Research.

<sup>104</sup>For an example on assessing impact see Guskey, T. R. (2002). *Does It Make a Difference? Evaluating Professional Development*. *Educational Leadership*. Vol. 59, No. 6, 45-51.

<sup>105</sup>Daly, T., Keeling, D., Grainger, R., Grundies, A. (2008). *Mutual Benefits: New York City's Shift to Mutual Consent in Teacher Hiring*. Brooklyn, NY: The New Teacher Project.

<sup>106</sup>Ibid Endnote 101

<sup>107</sup>TNTP. (2014). *Shortchanged: The Hidden Costs of Lockstep Teacher Pay*. Brooklyn, NY: TNTP.

<sup>108</sup>Hassel, E. A. & Hassel, B. C. (2013). *An Opportunity Culture for all: Making teaching a highly paid, high-impact profession*. Chapel Hill, NC: Public Impact.

<sup>109</sup>See for example: Koedel, C., Ehlert, M., Podgursky, M., Parsons, E. (2012). Teacher preparation programs and teacher quality: Are there real differences across programs? *University of Missouri Department of Economics Working Paper Series*; Osborne, C., von Hippel, P., Lincoff, J., Mills, N., Bellows, L. (2013, March); The small and unreliable effects of teacher preparation programs on student test scores in Texas. Presented at the spring Association of Education Finance and Policy conference, New Orleans, LA.

<sup>110</sup>TNTP. (2013). *Leap Year: Assessing and Supporting Effective First-Year Teachers*. Brooklyn, NY: TNTP.

## ABOUT TNP

TNP believes our nation's public schools can offer all children an excellent education. A national nonprofit founded by teachers, we help school systems end educational inequality and achieve their goals for students. We work at every level of the public education system to attract and train talented teachers and school leaders, ensure rigorous and engaging classrooms, and create environments that prioritize great teaching and accelerate student learning. Since 1997, we've partnered with more than 200 public school districts, charter school networks and state departments of education. We have recruited or trained more than 50,000 teachers, inspired policy change through acclaimed studies such as *The Widget Effect* (2009) and *The Irreplaceables* (2012), and launched one of the nation's premiere awards for excellent teaching, the Fishman Prize for Superlative Classroom Practice. Today, TNP is active in more than 40 cities.

## ACKNOWLEDGMENTS

Many individuals across TNP were instrumental in creating this report. Dina Hasiotis led our two-year effort and our talented research team: Erin Grogan, Karen Lawrence, Adam Maier and Alex Wilpon, with additional support from Claire Allen-Platt, Heather Barondess, Trevor Bynoe, Megan Goodrich, Kevin Haynes and Danielle Proulx.

Andy Jacob and Kate McGovern led writing and editing. Elizabeth Vidyarthi, Jacqui Seidel and Keith Miller led design.

Four members of TNP's leadership team—Timothy Daly, Daniel Weisberg, David Keeling and Jennifer Mulhern—were deeply involved in every stage of the project. Ariela Rozman and Karolyn Belcher also contributed valuable feedback throughout the process.

We are grateful for the contributions of our Technical Advisory Panel: Eric Hanushek, Thomas Kane, Marguerite Roza, Douglas Staiger and James Wyckoff. Their candid feedback on our methodology and findings helped push our thinking and shape the final report. We also wish to thank Ashley Woo, along with other members of the Education Resource Strategies team, for sharing knowledge and providing feedback on calculating district investments in teacher improvement.

Finally, we are deeply indebted to the staff of the school districts that took part in our study, and to the thousands of teachers, principals and district staff who answered our questions and helped us understand their experiences.

The report, graphics and figures were designed by Kristin Girvin Redman and Bethany Friedericks at Cricket Design Works in Madison, Wisconsin.

## DISCLOSURE

The districts studied for this report are among the more than 60 school systems with which TNP is currently engaged as a consultant and/or service provider. None of these districts held editorial control over this report, and the report was independently funded.

