



2009

THE WIDGET EFFECT

Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness



The New Teacher Project

Daniel Weisberg • Susan Sexton • Jennifer Mulhern • David Keeling



THE WIDGET EFFECT

Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness



02 | FOREWORD

03 | EXECUTIVE SUMMARY

09 | THE PROBLEM:
Teachers as
Interchangeable Parts

10 | CHARACTERISTICS:
The Widget Effect In
Teacher Evaluation

19 | FLAWS
In Evaluation Practice

24 | POLICY IMPLICATIONS
Of the Widget Effect

26 | RECOMMENDATIONS:
Reversing the Widget Effect

31 | END NOTES

33 | METHODOLOGY

AUTHORS | Daniel Weisberg • Susan Sexton • Jennifer Mulhern • David Keeling

CONTRIBUTING AUTHORS | Joan Schunck • Ann Palcisco • Kelli Morgan

SECOND EDITION

Acknowledgments | This report is the result of invaluable contributions from many individuals across The New Teacher Project. The authors would especially like to thank Dahlia Constantine, Timothy Daly, Vinh Doquang, Adele Grundies, Crystal Harmon, Dina Hasiotis, Ellen Hur, Gabrielle Misfeldt, David Osta, Ariela Rozman and Jeffrey Wilson for their efforts and insights. Additionally, we would like to thank Rachel Grainger, Judith Schiamberg and Andrew Sokatch for their work on the initial design of the project, and Caryn Fliegler and Elizabeth Vidyarthi for their help in the report's publication.

We would also like to recognize the advisory panels in each of the four study states for helping us shape the study design, understand state policy and refine the report's recommendations.

We are indebted to each of the districts represented in our study and their staff members who provided invaluable assistance to us with data collection and interpretation. We are grateful for the commitment from district leadership and central office staff as well as leadership and staff at local teachers unions, all of whom invested many hours of their valuable time to provide us with data, information and local context.

Finally, we thank each of the approximately 15,000 teachers and 1,300 administrators who dedicated time to completing our surveys. Your opinions and thoughts continue to inspire us to work to ensure that each and every student has access to outstanding teachers.

Funding Support | Primary funding for this report was provided by the Robertson Foundation, the Bill & Melinda Gates Foundation and the Joyce Foundation. Additional funding was provided by the Carnegie Corporation of New York, the Laura and John Arnold Foundation, the Charles and Helen Schwab Foundation and the Walton Family Foundation. We thank all of our funders for their generous support; however, we acknowledge that the findings and recommendations presented in this report are those of the The New Teacher Project alone and do not necessarily reflect our funders' opinions or positions.

“There are at least ‘several hundred’ incompetents now in the school system [says the superintendent]. Other observers think there are several thousands, while still others insist that ‘several’ would be nearer the mark. Whether these incompetents were unfit to teach at any time, or have been rendered unfit by the passing years, is a matter of opinion. The question is, why are they allowed to remain?”¹

So wrote The New York Times—in 1936.

In the 73 years since, we have made little progress toward answering the question of why poor instruction in our schools goes unaddressed. The question has been the subject of vigorous discussion, but most commentary has attempted to answer it by debating the failure of school districts to dismiss teachers who perform poorly.

The contours of this debate are well-known. One side claims that teacher tenure and due process protections render dismissal a practical impossibility, shielding ineffective teachers from removal in all but the most egregious instances. The other argues that the process provides only minimal protection against arbitrary or discriminatory dismissal, but that administrators fail to document poor performance adequately and refuse to provide struggling teachers with sufficient support. For decades these positions have remained largely unchanged.

The established arguments, however, fail to recognize that the challenge of addressing performance in the teaching profession goes far beyond the issue of dismissal. In fact, as this report illustrates, school districts fail to acknowledge or act on differences in teacher performance almost entirely. When it comes to officially appraising performance and supporting improvement, a culture of indifference about the quality of instruction in each classroom dominates.

Our research confirms what is by now common knowledge: tenured teachers are identified as ineffective and dismissed from employment with exceptional infrequency. While an important finding in its own right, we have come to understand that infrequent teacher dismissals are in fact just one symptom of a larger, more fundamental crisis—the inability of our schools to assess instructional performance accurately or to act on this information in meaningful ways.

This inability not only keeps schools from dismissing consistently poor performers, but also prevents them from recognizing excellence among top-performers or supporting growth among the broad plurality of hard-working teachers who operate in the middle of the performance spectrum. Instead, school districts default to treating all teachers as essentially the same, both in terms of effectiveness and need for development.

Of course, as teachers themselves are acutely aware, they are not at all the same. Just like professionals in other fields, teachers vary. They boast individual skills, competencies and talents. They generate different responses and levels of growth from students.

In a knowledge-based economy that makes education more important than ever, teachers matter more than ever. This report is a call to action—to policymakers, district and school leaders and to teachers and their representatives—to address our national failure to acknowledge and act on differences in teacher effectiveness once and for all. To do this, school districts must begin to distinguish great from good, good from fair, and fair from poor. Effective teaching must be recognized; ineffective teaching must be addressed.

Recently, President Obama spoke in bold terms about improving teacher effectiveness in just this way, saying, “If a teacher is given a chance or two chances or three chances but still does not improve, there is no excuse for that person to continue teaching. I reject a system that rewards failure and protects a person from its consequences. The stakes are too high. We can afford nothing but the best when it comes to our children’s teachers and the schools where they teach.”² We could not agree more. It is our hope that the recommendations contained in this report will outline a path to a better future for the profession.

A teacher's effectiveness—the most important factor for schools in improving student achievement—is not measured, recorded, or used to inform decision-making in any meaningful way.



EXECUTIVE SUMMARY

Suppose you are a parent determined to make sure your child gets the best possible education. You understand intuitively what an ample body of research proves: that your child's education depends to a large extent on the quality of her teachers. Consequently, as you begin considering local public schools, you focus on a basic question: who are the best teachers, and where do they teach?

The question is simple enough. There's just one problem—except for word of mouth from other parents, no one can tell you the answers.

In fact, you would be dismayed to discover that not only can no one tell you which teachers are most effective, they also cannot say which are the least effective or which fall in between. Were you to examine the district's teacher evaluation records yourself, you would find that, on paper, *almost every* teacher is a great teacher, even at schools where the chance of a student succeeding academically amounts to a coin toss, at best.

In short, the school district would ask you to trust that it can provide your child a quality education, even though it cannot honestly tell you whether it is providing her a quality teacher.

This is the reality for our public school districts nationwide. Put simply, they fail to distinguish great teaching from good, good from fair, and fair from poor. A teacher's effectiveness—the most important factor for schools in improving student achievement—is not measured, recorded, or used to inform decision-making in any meaningful way.

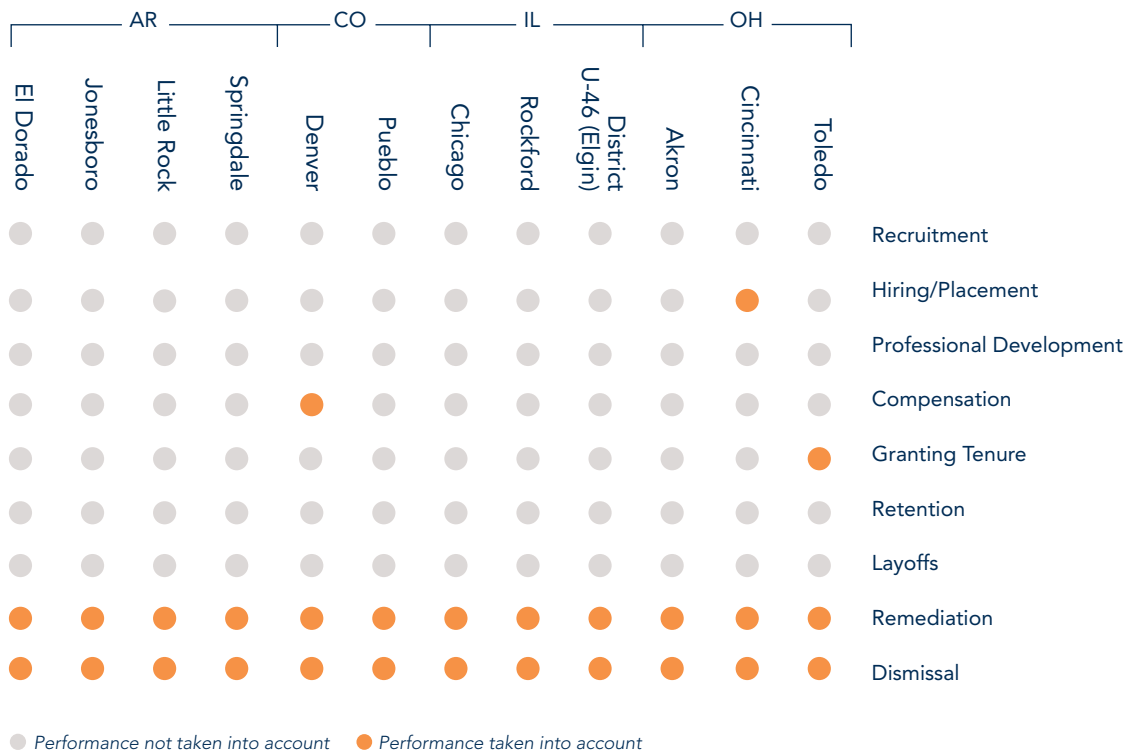
The Widget Effect

This report examines our pervasive and longstanding failure to recognize and respond to variations in the effectiveness of our teachers. At the heart of the matter are teacher evaluation systems, which in theory should serve as the primary mechanism for assessing such variations, but in practice tell us little about how one teacher differs from any other, except teachers whose performance is so egregiously poor as to warrant dismissal.

The failure of evaluation systems to provide accurate and credible information about individual teachers' instructional performance sustains and reinforces a phenomenon that we have come to call the **Widget Effect**. The Widget Effect describes the tendency of school districts to assume classroom effectiveness is the same from teacher to teacher. This decades-old fallacy fosters an environment in which teachers cease to be understood as individual professionals, but rather as interchangeable parts. In its denial of individual strengths and weaknesses, it is deeply disrespectful to teachers; in its indifference to instructional effectiveness, it gambles with the lives of students.

Today, the Widget Effect is codified in a policy framework that rarely considers teacher effectiveness for key decisions, as illustrated below.

Where Is Performance a Factor in Important Decisions About Teachers?*



The fact that information on teacher performance is almost exclusively used for decisions related to teacher remediation and dismissal paints a stark picture: In general, our schools are indifferent to instructional effectiveness—except when it comes time to remove a teacher.

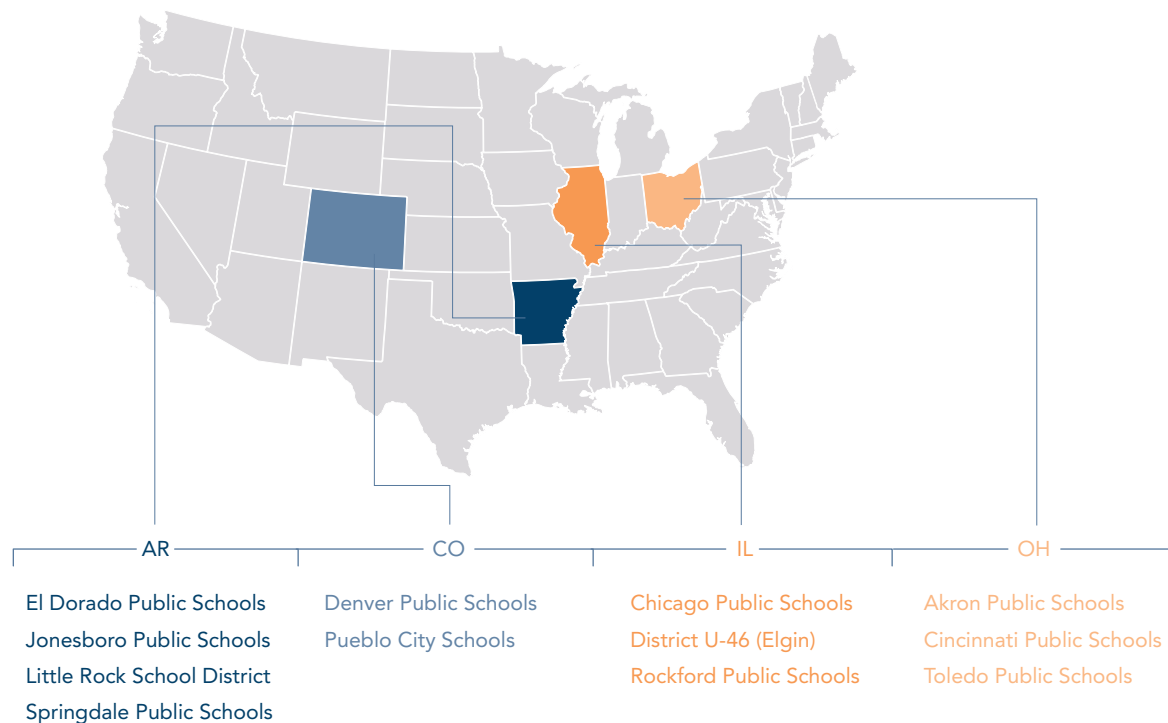
* See “Policy Implications of the Widget Effect” for additional information

Study Overview

This report is the product of an extensive research effort spanning 12 districts and four states. It reflects survey responses from approximately 15,000 teachers and 1,300 administrators, and it has benefited from the insight of more than 80 local and state education officials, teachers union leaders, policymakers and advocates who participated in advisory panels in each state, shaping the study design, data collection instruments, and findings and recommendations.

The four states included in the study, Arkansas, Colorado, Illinois and Ohio, employ diverse teacher performance management policies. The 12 districts studied range in size, geographic location, evaluation policies and practices and overall approach to teacher performance management. Jonesboro Public Schools, the smallest district studied, serves approximately 4,450 students; Chicago Public Schools, the largest, serves 413,700. All 12 districts employ some formal evaluation process for teachers, but the methods and frequency of evaluation differ. The outcomes, however, are strikingly similar.

Study Sites*



*For more information on the study sites, please see *Methodology*.

Characteristics of the Widget Effect in Teacher Evaluation

The Widget Effect is characterized by institutional indifference to variations in teacher performance. Teacher evaluation systems reflect and reinforce this indifference in several ways.

All teachers are rated good or great

In districts that use binary evaluation ratings (generally “satisfactory” or “unsatisfactory”), more than 99 percent of teachers receive the satisfactory rating. Districts that use a broader range of rating options do little better; in these districts, 94 percent of teachers receive one of the top two ratings and less than 1 percent are rated unsatisfactory.

Excellence goes unrecognized

When all teachers are rated good or great, those who are truly exceptional cannot be formally identified. Fifty-nine percent of teachers and 63 percent of administrators say their district is not doing enough to identify, compensate, promote and retain the most effective teachers.

Inadequate professional development

The failure to assess variations in instructional effectiveness also precludes districts from identifying specific development needs in their teachers. In fact, 73 percent of teachers surveyed said their most recent evaluation did not identify any development areas, and only 45 percent of teachers who did have development areas identified said they received useful support to improve.

No special attention to novices

Inattention to teacher performance and development begins from a teacher’s first days in the classroom. Though it is widely recognized that teachers are least effective in their beginning years, 66 percent of novice teachers in districts with multiple ratings received a rating greater than “satisfactory” on their most recent performance evaluation. Low expectations characterize the tenure process as well, with 41 percent of administrators reporting that they have never “non-renewed” a probationary teacher for performance concerns in his or her final probationary year.

Poor performance goes unaddressed

Despite uniformly positive evaluation ratings, teachers and administrators both recognize ineffective teaching in their schools. In fact, 81 percent of administrators and 57 percent of teachers say there is a tenured teacher in their school who is performing poorly, and 43 percent of teachers say there is a tenured teacher who should be dismissed for poor performance. Troublingly, the percentages are higher in high-poverty schools. But district records confirm the scarcity of formal dismissals; at least half of the districts studied did not dismiss a single non-probationary teacher for poor performance in the time period studied (ranging from two to five years in each district).

Flaws in Evaluation Practice and Implementation

The characteristics above are exacerbated and amplified by cursory evaluation practices and poor implementation. Evaluations are short and infrequent (most are based on two or fewer classroom observations, each 60 minutes or less), conducted by administrators without extensive training, and influenced by powerful cultural forces—in particular, an expectation among teachers that they will be among the vast majority rated as top performers.

While it is impossible to know whether the system drives the culture or the culture the system, the result is clear—evaluation systems fail to differentiate performance among teachers. As a result, teacher effectiveness is largely ignored. Excellent teachers cannot be recognized or rewarded, chronically low-performing teachers languish, and the wide majority of teachers performing at moderate levels do not get the differentiated support and development they need to improve as professionals.

Reversing the Widget Effect

The Widget Effect is deeply ingrained in the fundamental systems and policies that govern the teachers in our public schools. Better evaluation systems may offer a partial solution, but they will not overcome a culture of indifference to classroom effectiveness. Reversing the Widget Effect depends on better information about instructional quality that can be used to inform other important decisions that dictate who teaches in our schools.

01 | Adopt a comprehensive performance evaluation system that fairly, accurately and credibly differentiates teachers based on their effectiveness in promoting student achievement. Teachers should be evaluated based on their ability to fulfill their core responsibility as professionals—delivering instruction that helps students learn and succeed. This demands clear performance standards, multiple rating options, regular monitoring of administrator judgments, and frequent feedback to teachers. Furthermore, it requires professional development that is tightly linked to performance standards and differentiated based on individual teacher needs. The core purpose of evaluation must be maximizing teacher growth and effectiveness, not just documenting poor performance as a prelude to dismissal.

02 | Train administrators and other evaluators in the teacher performance evaluation system and hold them accountable for using it effectively. The differentiation of teacher effectiveness should be a priority for school administrators and one for which they are held accountable. Administrators must receive rigorous training and ongoing support so that they can make fair and consistent assessments of performance against established standards and provide constructive feedback and differentiated support to teachers.



03 | Integrate the performance evaluation system with critical human capital policies and functions such as teacher assignment, professional development, compensation, retention and dismissal. Even the best evaluation system will fail if the information it produces is of no consequence. An effective evaluation system must be fully integrated with other district systems and policies and a primary factor in decisions such as which teachers receive tenure, how teachers are assigned and retained, how teachers are compensated and advanced, what professional development teachers receive, and when and how teachers are dismissed. Only by attaching stakes to evaluation outcomes will teachers and administrators invest in the hard work of creating a truly rigorous and credible evaluation system.

04 | Adopt dismissal policies that provide lower-stakes options for ineffective teachers to exit the district and a system of due process that is fair but efficient.

If the evaluation system is implemented effectively, unsatisfactory ratings will not be anomalous, surprising or without clear justification. Likewise, the identification of development areas and the provision of support will be continual. As in other professions, teachers who see significant, credible evidence of their own failure to meet standards are likely to exit voluntarily. Districts can facilitate this process by providing low-stakes options that enable teachers to leave their positions without being exiled. For teachers who must be officially dismissed, an expedited, one-day hearing should be sufficient for an arbitrator to determine if the evaluation and development process was followed and judgments made in good faith.

Our recommendations outline a comprehensive approach to improving teacher effectiveness and maximizing student learning. If implemented thoroughly and faithfully, we believe they will enable districts to understand and manage instructional quality with far greater sophistication. Improved evaluation will not only benefit students by driving the systematic improvement and growth of their teachers, but teachers themselves, by at last treating them as professionals, not parts.



THE PROBLEM: TEACHERS AS INTERCHANGEABLE PARTS

Teaching is the essence of education, and there is almost universal agreement among researchers that teachers have an outsized impact on student performance. We know that improving teacher quality is one of the most powerful ways—if not *the* most powerful way—to create better schools. In fact, a student assigned to a very good teacher for a single school year may gain up to a full year’s worth of additional academic growth compared to a student assigned to a very poor teacher. Having a series of strong or weak teachers in consecutive years compounds the impact. Give high-need students three highly effective teachers in a row and they may outperform students taught by three ineffective teachers in a row by as much as 50 percentile points.³

The lesson from these decades of research is clear: teachers matter. Some teachers are capable of generating exceptional learning growth in students; others are not, and a small group actually hinders their students’ academic progress.

This simple premise—that teachers matter—has driven The New Teacher Project’s prior research and continues to drive our work today. Our 2003 report, *Missed Opportunities: How We Keep High-Quality Teachers Out of Urban Classrooms*, documented how vacancy notification policies, rigid staffing rules and late budget timelines caused urban districts to hire too late to capture the highest-quality teacher applicants. Our 2005 report, *Unintended Consequences: The Case for Reforming the Staffing Rules in Urban Teachers Union Contracts*, illustrated how contractual staffing rules, built around the assumption that any teacher could fill any vacancy, forced schools to hire teachers they did not want and teachers to take positions for which they might not be a good fit.

Each of these reports in its own way documented a flawed assumption that has pervaded American educational policy for decades—the assumption that teachers are interchangeable parts. We have come to call this phenomenon the **Widget Effect**. In the presence of the Widget Effect, school systems wrongly conflate educational access with educational *quality*; the only teacher quality goal that schools need to achieve is to fill all of their positions. It becomes a foregone conclusion that, so long as there is an accredited teacher—any teacher—in front of the classroom, students are being served adequately.

While the Widget Effect pervades many aspects of our education system, it is in teacher evaluation that both its architecture and its consequences are most immediately apparent. In this report, we examine the central role that the design and implementation of teacher evaluation systems play in creating and reinforcing the Widget Effect; how teacher and administrator beliefs about evaluation illustrate the Widget Effect at work; and how the Widget Effect fuels a policy framework that ignores both strong and weak teacher performance. In the absence of meaningful performance information, teacher effectiveness is treated as a constant, not a variable, and school districts must instead rely on other considerations—many of them unrelated to student academic success—to make critical workforce decisions.

In the presence of the Widget Effect, school systems wrongly conflate educational access with educational quality.

CHARACTERISTICS: THE WIDGET EFFECT IN TEACHER EVALUATION

The Widget Effect is rooted in the failure of teacher evaluation systems to produce meaningful information about teacher effectiveness. In theory, an evaluation system should identify and measure individual teachers' strengths and weaknesses accurately and consistently, so that teachers get the feedback they need to improve their practice and so that schools can determine how best to allocate resources and provide support. In practice, teacher evaluation systems devalue instructional effectiveness by generating performance information that reflects virtually no variation among teachers at all.

This fundamental failing has a deeply insidious effect on teachers and schools by institutionalizing indifference when it comes to performance. As a result, important variations between teachers vanish. Excellence goes unrecognized, development is neglected and poor performance goes unaddressed.

All Teachers Are Rated Good or Great

The disconnect between teacher evaluation systems and actual teacher performance is most strikingly illustrated by the wide gap between student outcomes and teacher ratings in many districts. Though thousands of teachers included in this report teach in schools where high percentages of students fail year after year to meet basic academic standards, less than one percent of surveyed teachers received a negative rating on their most recent evaluation.⁴

This is not to say that responsibility for a failing school rests on the shoulders of teachers alone, or that none of these teachers demonstrated truly high performance; however, there can be no doubt that these ratings dramatically overstate the number of exemplary teachers and understate the number with moderate and severe performance concerns. These data simultaneously obscure poor performance and overlook excellence, as the value of superlative teacher ratings is rendered meaningless by their overuse.

To a large degree, teacher evaluation systems codify this whitewashing of performance differences, beginning with the rating categories themselves. Five of the ten districts in this study with available teacher evaluation rating data⁵ use a binary rating system for assessing teacher performance;

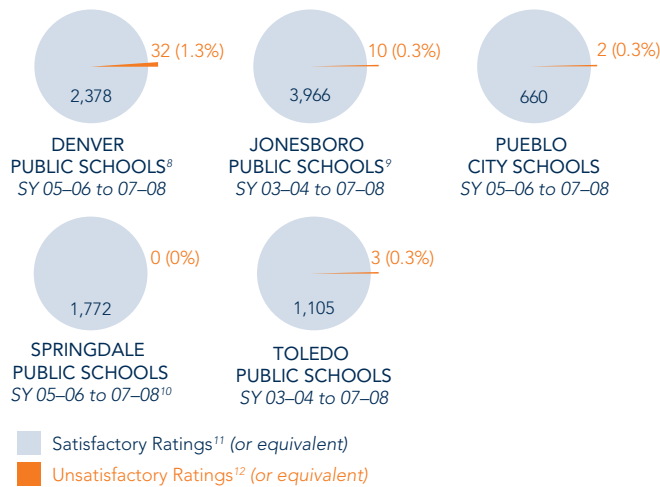
“Poorly performing teachers are rated at the same level as the rest of us. This infuriates those of us who do a good job.”

—Akron Public Schools Teacher

teachers are categorized as either “satisfactory” or “unsatisfactory.”⁶ There are no shades of gray to describe nuances in performance.

As *Figure 01* illustrates, in districts that use binary ratings, virtually all tenured⁷ teachers (more than 99 percent) receive the satisfactory rating; the number receiving an unsatisfactory rating amounts to a fraction of a percentage. In these districts, it makes little difference that two ratings are available; in practice only one is ever used.

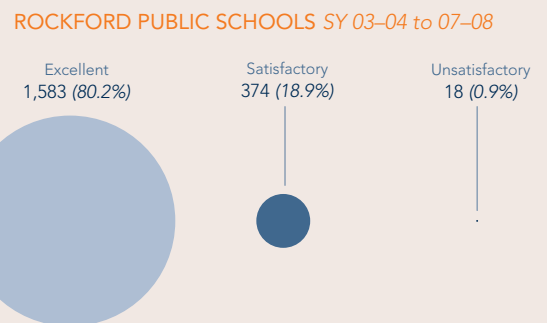
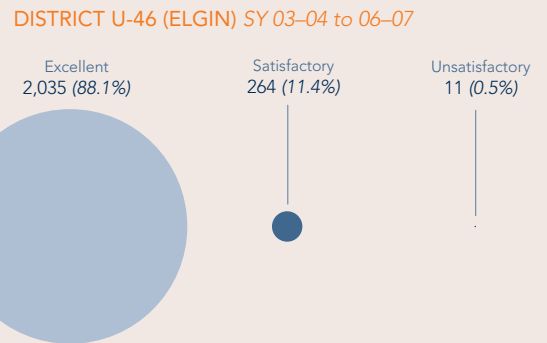
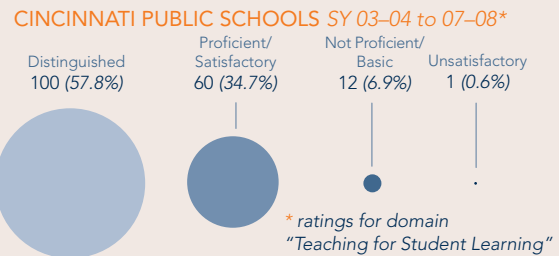
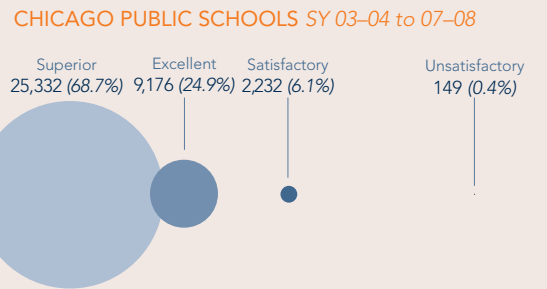
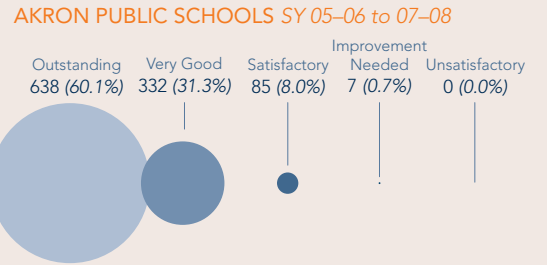
FIGURE 01 | Evaluation Ratings for Tenured Teachers in Districts with Binary Rating Systems*



One might hope that teacher evaluation systems that employ a broader range of rating options would more accurately reflect the performance differences among teachers. However, even when given multiple ratings from which to choose, evaluators in all districts studied rate the majority of teachers in the top category, rather than assigning the top rating to only those teachers who actually outperform the majority of their peers. As illustrated in *Figure 02*, in the five districts with multiple teacher evaluation ratings for which data were available,¹³ 70 percent of tenured teachers still received the highest rating.¹⁴ Another 24 percent received the second-highest rating.

While districts using multiple rating systems do show some additional variability in teacher evaluation beyond those using binary rating systems, districts with four or more ratings still assign tenured teachers the lowest two rating options in one out of 16 cases.¹⁵ In each case, the basic outcome remains true: almost no teachers are identified as delivering unsatisfactory instruction.

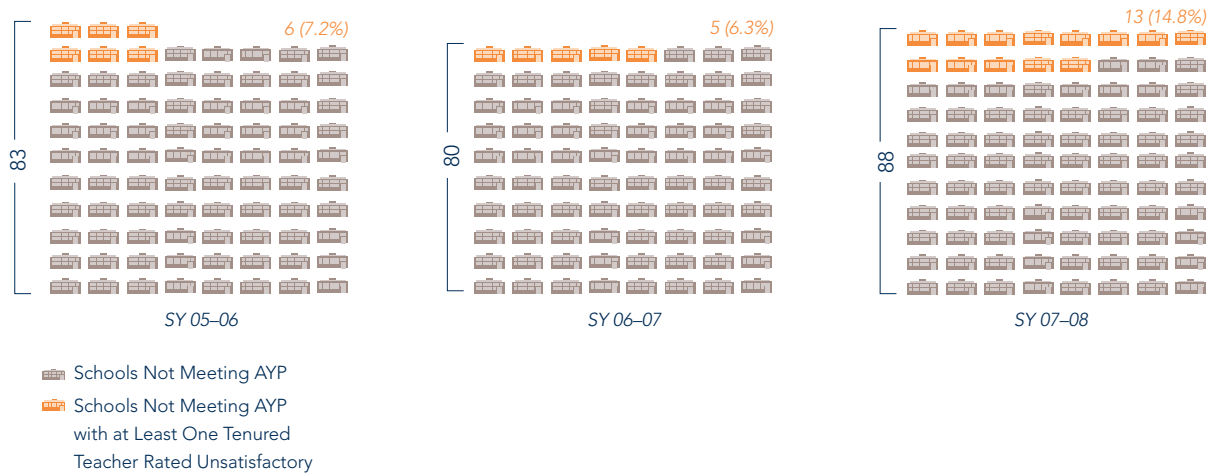
FIGURE 02 | Evaluation Ratings for Tenured Teachers in Districts with Multiple-Rating Systems*



*Note: Evaluation rating data in *Figures 01* and *02* were collected from each district. Data are as accurate as the records provided to TNTP for this study.

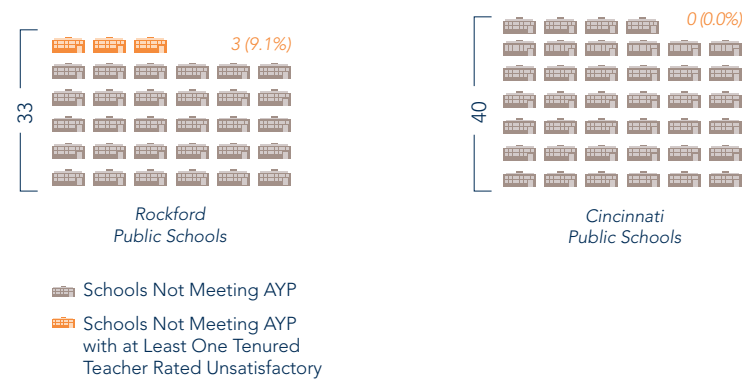
These data often stand in sharp relief against current levels of student achievement. For example, in Denver schools that did not make adequate yearly progress (AYP), more than 98 percent of tenured teachers received the highest rating—satisfactory.¹⁶ On average, over the last three years, only 10 percent¹⁷ of failing schools issued at least one unsatisfactory rating to a tenured teacher.

FIGURE 03 | Frequency of Unsatisfactory Ratings in Denver Public Schools that Did Not Meet AYP¹⁸



These findings are consistent with a one year snapshot of data from other districts. Less than 10 percent of Rockford’s failing schools rated a tenured teacher unsatisfactory in 2007–08, and none of Cincinnati’s failing schools did.

FIGURE 04 | Rockford Public Schools & Cincinnati Public Schools AYP Data (SY07-08)¹⁹



Moreover, it is important to note that performance simply goes untracked for a subset of teachers. In some cases, this is systemic. One of the 12 districts studied does not centrally track or record any evaluation data at all. However, in many other cases, it reflects the perfunctory nature of the evaluation system itself, as 9 percent of teachers surveyed appear to have missed their most recent scheduled evaluation.²⁰

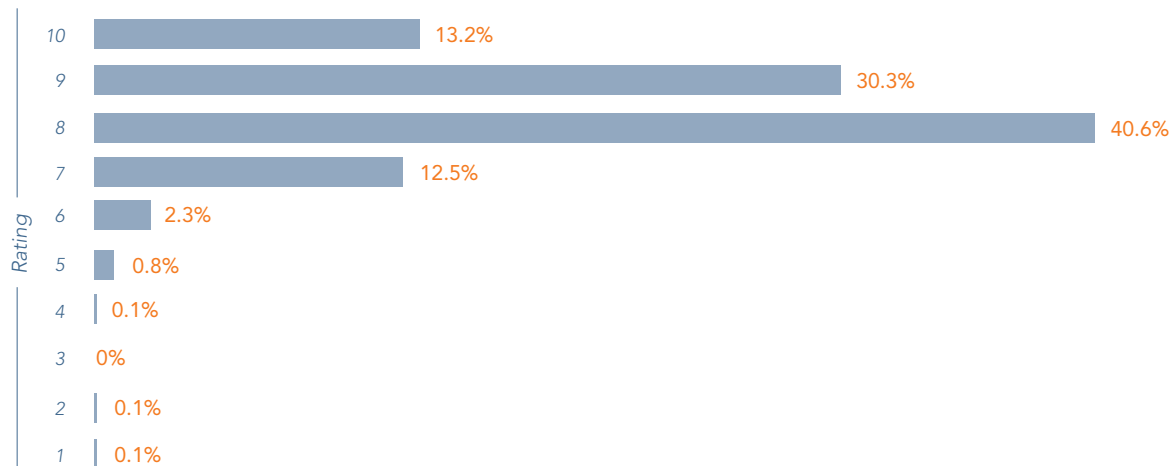
Excellence Goes Unrecognized

In a world where all teachers are rated as good or great, the truly outstanding teachers—those who are realizing life-changing academic success for their students—cannot be formally identified. And if they are not formally identified, schools cannot prioritize their retention or leverage them to develop and improve their colleagues.

In theory, districts should be able to identify their top performers by awarding them the highest rating on the evaluation scale, but as previously illustrated, the highest rating is awarded to many more teachers than can possibly fall into this category. The dilution of the highest rating category is reflected in teacher and administrator perceptions about how this category is defined. Nearly a quarter of administrators (24 percent) and nearly a fifth of teachers (18 percent) equate their district’s highest rating with a teacher who is merely effective or even somewhat effective, rather than seeing that rating as reserved for those who are truly exceptional.²¹

In the absence of a mechanism for identifying and rewarding outstanding performers, the average effort becomes the bar for the mark of excellence. In a subset of districts²² where teachers were asked to rate their instructional performance on a scale from 1 to 10, more than 43 percent rated themselves a 9 or higher (see *Figure 05*). These teachers are not irrationally inflating their estimate of their teaching performance; they are simply responding to an environment in which all are assumed to be superior performers.

FIGURE 05 | Teacher Assessments of Their Own Instructional Performance
ON A SCALE OF 1 TO 10, HOW WOULD YOU RATE YOUR INSTRUCTIONAL PERFORMANCE?



If districts could systematically identify which teachers perform at the highest level, they could use this information to inform teaching assignments, target teachers for teacher leader positions, and prioritize the retention of these teachers. In the absence of this information, however, excellence cannot be recognized or rewarded. As in other areas studied, there is broad agreement among teachers and administrators that this is a problem. Fifty-nine percent of teachers and 63 percent of administrators from the four study sites where we surveyed more deeply on the topic report their district is not doing enough to identify, compensate, promote and retain the most effective teachers.²³

“There is no recognition for teachers who are doing an exemplary job.”

-Chicago Public Schools Teacher

Development Is Limited

The damage of ignoring differences in teacher effectiveness is not isolated to the limited recognition of excellence; an equally troubling consequence is that teachers rarely receive meaningful feedback on their performance through the formal evaluation system. In the 12 districts studied, development areas were identified for only 26 percent of teachers during their most recent evaluations.²⁴

In other words, nearly 3 of 4 teachers went through the evaluation process but received no specific feedback about how to improve their practice. This is true even for novice teachers who are most in need of actionable feedback as they learn their craft—only 43 percent of teachers in their first four years had any development areas identified. It is inconceivable that 74 percent of teachers, and 57 percent of teachers in their first four years, do not require improvement in any area of performance.

Some may argue that administrators prefer to give teachers critical feedback outside the formal evaluation process. However, 47 percent²⁵ of teachers report not having participated in a single informal conversation with their administrator over the last year about improving aspects of their instructional performance. In addition, of the relatively small group of teachers who had a performance area identified as in need of improvement or unsatisfactory, 62 percent said they were not aware of performance concerns before their evaluation.²⁶

This suggests that many administrators do not regularly or proactively offer feedback on instructional performance outside of the formal evaluation process.

While districts often fail to identify areas where teachers are in need of improvement, they also fail to provide targeted support to the subset of teachers who have had development areas identified. Less than half (45 percent²⁷) of teachers across all districts who had development areas identified on their most recent evaluations said they received useful support to improve those areas.

Constructive feedback that specifies areas for development is a critical facet of any performance evaluation, even for strong performers. In theory, even if virtually all teachers are rated as good or great, their evaluations could provide them with valuable feedback they could use to improve their instructional practice. However, that theoretical potential currently goes unrealized and teachers are too often denied both the knowledge and the opportunity to improve.

As a result, it is not surprising that so many teachers believe that the current evaluation system, and the absence of meaningful feedback it produces, does them a disservice. Only 42 percent of teachers agree that evaluation allows accurate assessment of performance and only 43 percent of teachers agree that evaluation helps teachers improve.²⁸

“The evaluation process should have teacher development as the primary goal, not just assigning a number on a rubric. As it is set up now, there is no immediate feedback to the teacher in any constructive format. Scores are based on rigid, often meaningless recitations. It is the epitome of poor teaching methods to give a score without discussion.”

—Cincinnati Public Schools Teacher

Novice Teachers Receive No Special Attention or Scrutiny²⁹

One could argue teacher ratings are so high and development is so limited because probationary teachers undergo a rigorous screening process through which weak performers are weeded out. According to this line of argument, all the poorly performing teachers were effectively ushered out while they were still novices. Yet as illustrated in *Figure 06*, our research found no evidence that teachers are subject to a rigorous screening process during their probationary periods; only a fraction of teachers are “non-renewed” by the districts when they have the opportunity to do so.

As a result, though the awarding of tenure status has the potential to recognize effective teaching and to transition out teachers who are unable to reach a reasonable performance standard, in practice there is no observable rigor applied to the tenure decision. It is not surprising that many administrators (41 percent) report that they have never non-renewed a teacher in his or her final probationary year because they found that teacher’s performance unworthy of tenure. Moreover, 76 percent³⁰ of novice teachers express confidence that they will receive tenure even before they have completed the probationary period, often because they have consistently received superlative ratings—even as first-year teachers.

This lack of rigor also leads to a limited focus on development for novice teachers. Though it is widely recognized that teachers are less effective in their first years in the classroom, differences in performance tend to go unremarked from the very beginning of a teacher’s career. Novice teachers begin receiving the highest rating when they start their career or within a few years of being hired, with 66 percent of novice teachers in districts with multiple ratings receiving a rating greater than “satisfactory” on their most recent performance evaluation.³¹ By giving novice teachers high ratings from the day they begin teaching, schools communicate inattention to and low expectations for instructional performance. Furthermore, they miss a critical window of opportunity to focus new teachers on their instructional strengths and

“New teachers are given so little support in my district that sometimes they are simply doomed to fail. Yet, no one notices and they finish their probationary status without a negative evaluation.”

-Denver Public Schools Teacher

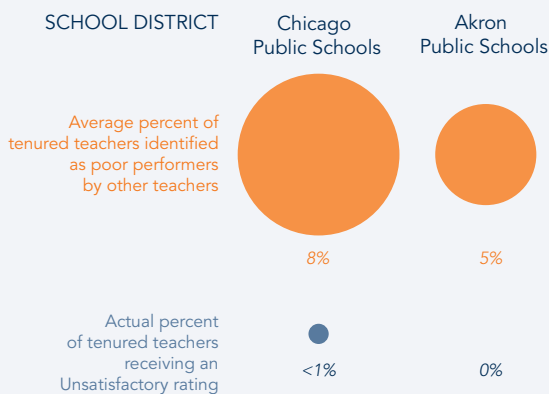
FIGURE 06 | Non-renewal Patterns of Probationary Teachers³²

SCHOOL DISTRICT	Number of non-renewals for performance in 5 years	Average percent of probationary teachers non-renewed for performance each year
Denver ³³ Public Schools	132	3.0%
Jonesboro Public Schools	0	0.0%
Chicago ³⁴ Public Schools	29	0.1%
District U-46 (Elgin)	28	0.9%
Toledo Public Schools ³⁵	7	0.9%
Cincinnati Public Schools	7	0.1%

“I think it gives the hard working, honest teachers a bad reputation being lumped together with a group of sub-par teachers. What’s even worse is that our principal does absolutely nothing about any of this.”

-Akron Public Schools Teacher

FIGURE 07 | Percent of Poor Performers Teachers Observe in Their Schools vs. Percent of Teachers Given an Unsatisfactory Rating³⁶



weaknesses during a formative point in their careers. Instead of getting meaningful feedback about what they are doing right and wrong in their instructional practice, new teachers mostly get the message that their actual performance has little bearing on how they are rated.

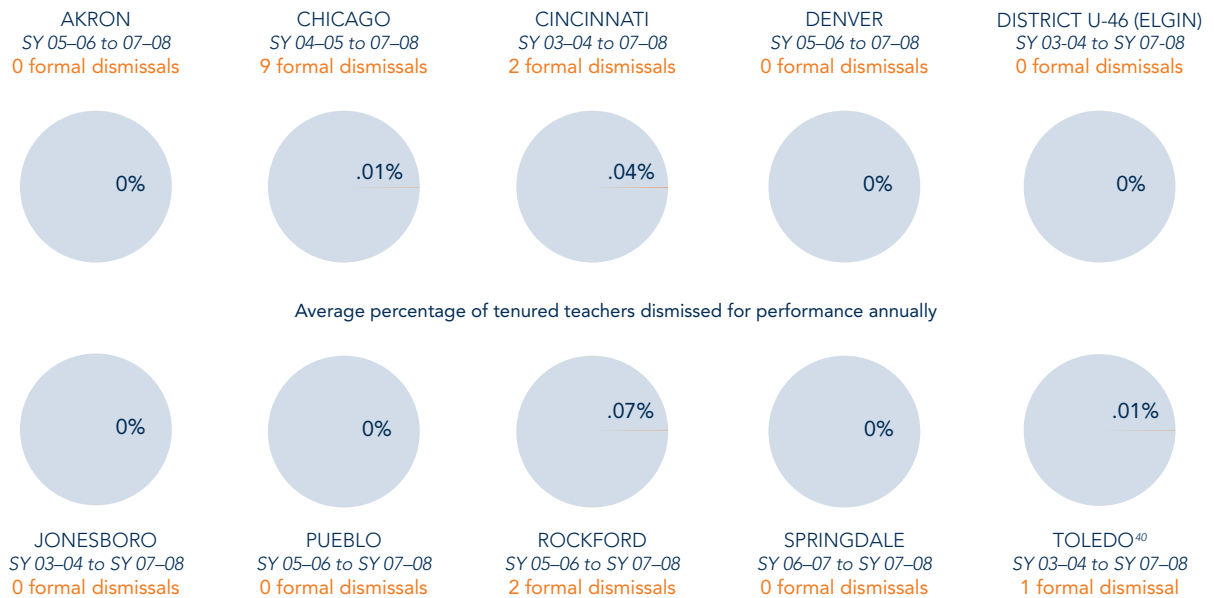
Poor Performance Goes Unaddressed

It goes without saying that teacher dismissal has become a polarizing issue in the education community; however, we found that teachers and administrators broadly agree about the existence and scope of the problem and about what steps need to be taken to address poor performance in schools. In fact, an overwhelming majority of both teachers (68 percent) and administrators (91 percent) agree or strongly agree that dismissing poor performers is important to maintaining high-quality instructional teams. This may seem self-evident, but it suggests a consensus that teacher performance management should entail accountability, not just development.

In the four districts where we surveyed more deeply, teachers and administrators agree that there is a small but significant subset of teachers who perform poorly, with 81 percent of administrators and 57 percent of teachers reporting that there is a tenured teacher in their school who delivers poor instruction.³⁷ In *Figure 07*, we examine the levels of poor instructional performance teachers observe in their schools and compare it to the actual number of unsatisfactory ratings given in Chicago and Akron.³⁸ The data confirm what teachers and school administrators report—the number of teachers identified as unsatisfactory is miniscule and far lower than the percentage of poor performers observed by their colleagues.

Moreover, 43 percent of teachers across all districts believe that there is a tenured teacher in their school who should be *dismissed* for poor instructional performance but has not been. Yet experienced teachers are almost never actually dismissed for poor performance. Most administrators have not initiated the dismissal of a single tenured teacher in the past five years.³⁹ In fact, the number of dismissals for performance in each district studied can be counted in the single digits, if at all.

FIGURE 08 | Frequency of Tenured Teacher Dismissals for Performance

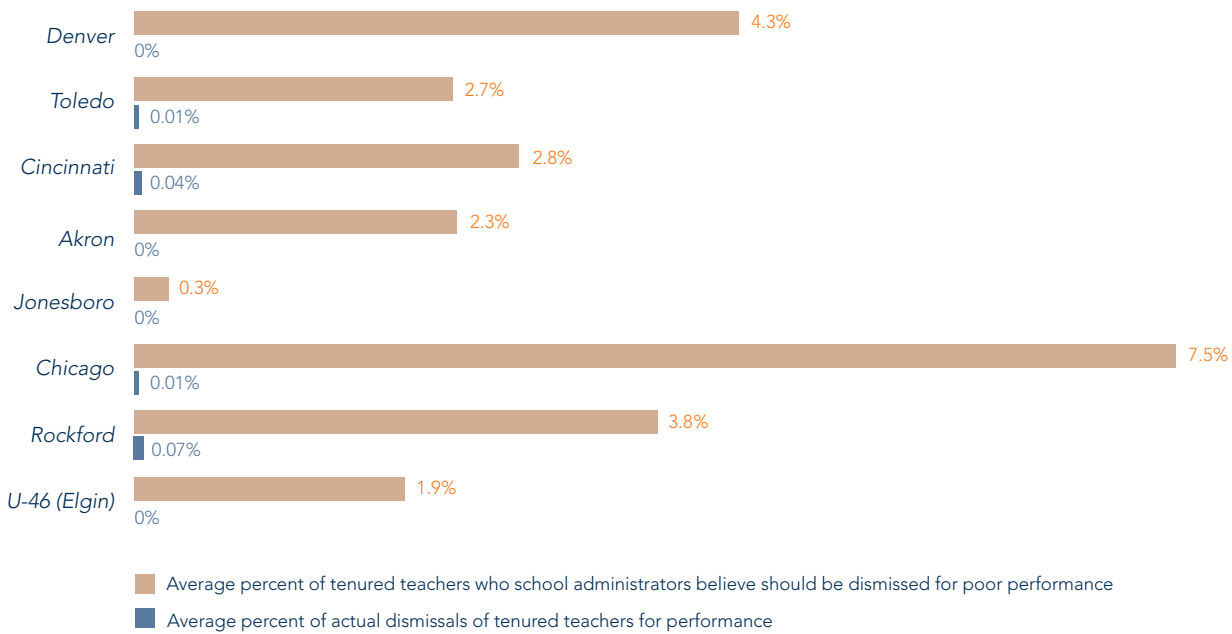


Note: Teacher dismissal for performance data was collected from ten districts, representing some combination of school years 2003-04 through 2007-08.⁴¹

It is not surprising then that most teachers (68 percent⁴²) believe that poor performance is overlooked by administrators. This is essentially confirmed by administrators themselves, 86 percent⁴³ of whom say they do not always pursue dismissal even if it is warranted. School administrators appear to be deterred from pursuing remediation and dismissal because they view the dismissal process as overly time consuming and cumbersome, and the outcomes for those who do invest the time in the process is uncertain. Even for the small number of administrators that actually do attempt the process, fully half report that it yielded an outcome other than dismissal.

While all of the districts studied share the goal of an evaluation system that can identify instances of ineffective performance so administrators can properly intervene, the data make clear that this does not occur. Despite the fact that teachers and administrators report that poor performance is commonplace, intervention appears to be extremely rare when compared to the scope of the problem (see Figure 09). We are left to conclude that current systems for managing teacher performance fail to function on the most basic level—addressing poor instructional performance.

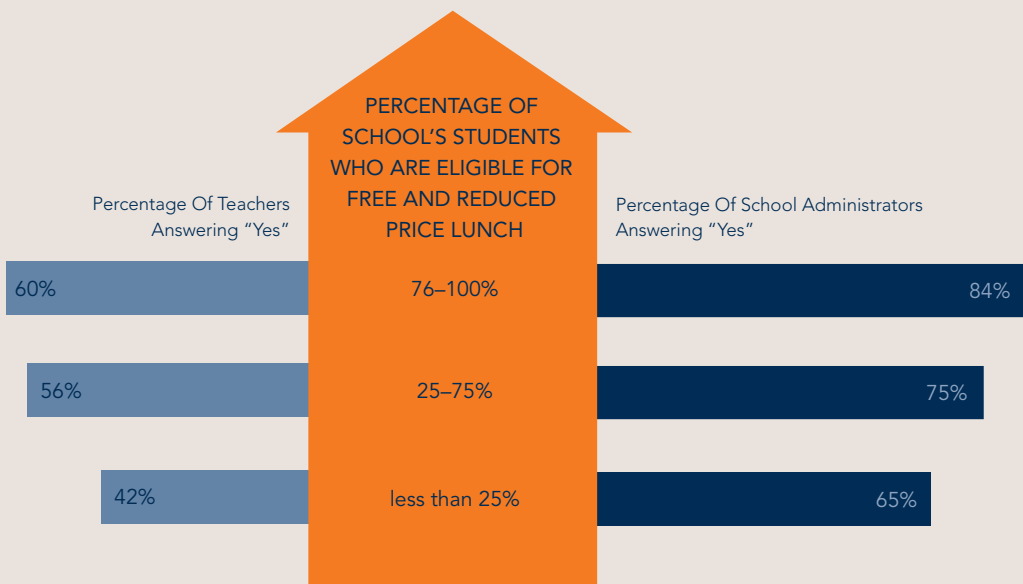
FIGURE 09 | Perceived Need for Dismissals vs. Actual Dismissals, by District



The Impact on High-Need Schools

Though poor performance goes unaddressed in most schools, our data indicate that the problem is most acute in the highest-need schools. These data are consistent across multiple districts⁴⁴ and with research that reflects that poor and minority children, who have the greatest need for effective teachers, are least likely to get them.

FIGURE 10 | In your opinion, are there tenured teachers in your school who deliver poor instruction?



FLAWS IN EVALUATION PRACTICE AND IMPLEMENTATION

While most teacher evaluation systems espouse grand intentions for teacher development, assessment and improvement, the data above show that all too often the outcome fails to equal the intent. Instead, the process becomes devalued. Evaluations are perfunctory, school districts do not invest in administrator capacity to provide meaningful feedback, and teachers come to expect that they will receive only positive feedback.

Teacher Evaluations Are Perfunctory

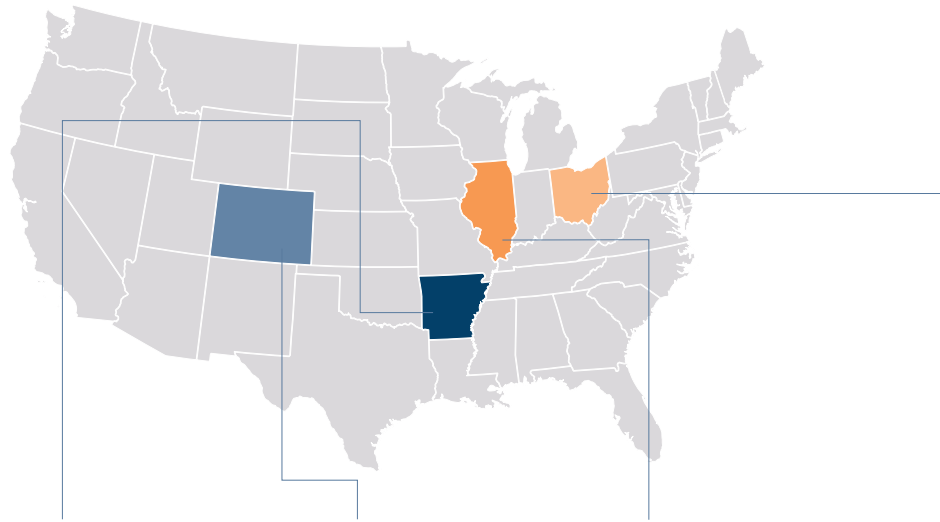
The current evaluation process reflects and codifies the assumption underlying the Widget Effect—that all teachers are essentially interchangeable. Operating under a belief system that one teacher is as good as any other, schools invest very little time or effort in evaluating teachers. Instead, they apply a perfunctory process, at best designed to capture a snapshot of a teacher’s instructional performance at a moment in time. Across the four states studied, all probationary teachers must be evaluated annually; however, tenured teachers may not be required to be evaluated at all, or only once every few years.

“It’s the easiest thing for administrators to do. It’s the path of least resistance. They don’t have time or often, even the authority, to coach or correct ineffective teachers. The good teachers remain unrewarded for doing fantastic jobs, while bad teachers get to coast along.”

—Little Rock Public Schools Teacher



FIGURE 11 | State Teacher Evaluation Requirements in Brief



	AR		CO		IL		OH	
	Probationary	Tenured	Probationary	Tenured	Probationary	Tenured	Probationary	Tenured
Evaluation Frequency	no requirement	no requirement	1 per year	1 every 3 years	1 per year	1 every 2 years	2 per year	no minimum
# of Observations Required	3 per year	no minimum	2	1	1 (2 per year in Chicago only)	1 (2 per year in Chicago only)	2	2
Duration of Observations	no requirement	no requirement	no requirement	no requirement	no requirement	no requirement	30 minutes or more	30 minutes or more

Moreover, only five of the districts studied track evaluation results electronically, a step that would at least provide the opportunity to easily monitor and use evaluation information to inform decision-making at a school and district-wide level. Other districts record evaluations in paper files, typically housed at the central office.

Not surprisingly, school administrators spend very little time on what is a largely meaningless and inconsequential evaluation process. Most teacher evaluations are based on two or fewer classroom observations totaling 76 minutes or less. Across all districts, 64 percent of tenured teachers were observed two or fewer times for their most recent evaluation, for an average total of 75 minutes.⁴⁵ Probationary teachers receive little additional attention despite their novice status; 59 percent of probationary teachers were observed two or fewer times

for their most recent evaluation, for an average total of 81 minutes, a mere six additional minutes. Clearly, effective evaluation amounts to far more than how much time an administrator spends in a teacher’s classroom, but the infrequency and brevity of administrator observations underscores their inattention to performance.

Equally important, evaluators spend no more time to observe or give feedback to the small number of teachers identified as mediocre or poor performers than they spend with highly rated teachers. Teachers receiving lower than the highest rating report the same number of observations as their more highly rated colleagues and the same amount of informal feedback.

FIGURE 12⁴⁶ | Number of classroom observations by evaluator, prior to evaluator assigning final evaluation rating(s).

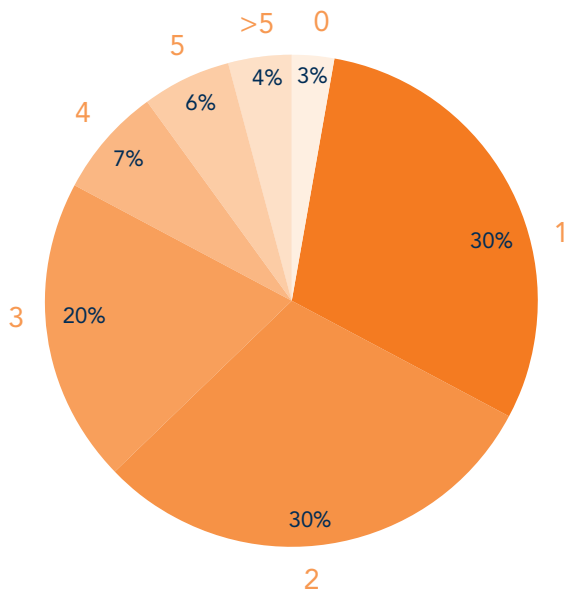
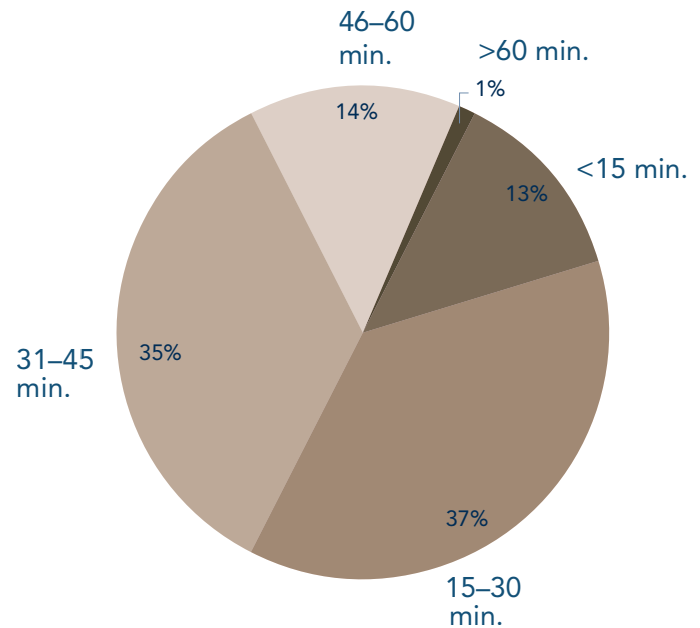


FIGURE 13⁴⁷ | Average minutes of a classroom observation, prior to a teacher being assigned a final evaluation rating(s).



65 percent of the lower-rated teachers and 62 percent of highest-rated teachers report 2 or fewer observations during their last evaluation cycle.⁴⁸

58 percent of lower-rated teachers receive informal feedback as compared to 56 percent of highest-rated teachers.⁴⁹

Even when performance is clearly an issue—as represented by the small number of teachers who received the lowest rating on their last evaluation—evaluators fail to invest significant time monitoring instruction. Among the small number of teachers receiving the lowest rating, 74 percent report that they were observed three or fewer times despite significant concerns about their performance.

“I do not feel adequately trained to conduct a teacher evaluation. There are evaluation tools, but no one reviews them with you. We are not trained on the process. As a first year principal, you try it and you move through the process because it has to be done.”

—Toledo Public Schools Principal

School Administrators Receive Limited Training

Given the low priority assigned to teacher evaluation, it comes as no surprise that school districts invest minimally in evaluation training for school administrators. In many districts, evaluation training is a one-time endeavor provided either when an administrator is new in his or her position or when the district implements a revised teacher evaluation system. Consequently, school administrators are ill-equipped to evaluate teachers effectively.

Background conversations conducted with district staff suggest that, in many of our study sites, school administrators receive varying levels of training on how to conduct an effective teacher evaluation. For example, in the Cincinnati Public Schools, evaluation training can be provided upon request. In Chicago Public Schools and District U-46 (Elgin), training may occur once a year for a limited number of principals, but not all. In other districts, including El Dorado Public Schools and Akron Public Schools, it simply does not occur.

As a result, across all study sites, 51 percent of school administrators describe their level of training in how to conduct an effective evaluation as “very extensive” or “extensive”⁵⁰ and school administrators with more evaluation training are more likely to report that they enforce a high standard for instructional performance.

Yet, it is important to note that extensive training alone did not produce a significant change in evaluation outcomes. School administrators with more extensive training report increased percentages of teachers enrolled in remediation or dismissed for delivering poor instruction than school administrators with less training. Yet even among those who report “very extensive” training, only 36 percent have recommended dismissal of a tenured teacher for poor instruction in the last five years.

Teacher Expectations Are Skewed

It is tempting to believe that simply requiring more frequent and thorough evaluations would result in more rigorous and accurate assessments of teacher performance and increase teachers’ confidence in and esteem for the evaluation process. However, we believe these reforms, while necessary, would be insufficient because the minimal nature of the process speaks to a far deeper problem in the culture of schools: the assumption that not only are all teachers the same, but that they are all performing at a high level.

Our research reflects that there is a strong and logical expectation among teachers that they will receive outstanding performance ratings. While the vast majority of teachers receive the highest rating, those teachers who do not receive it tend to believe that the higher rating was warranted.

In the six districts with multiple-rating scales for which survey data were available,⁵¹ 49 percent of probationary teachers and 77 percent of tenured teachers indicated that they believe they should have received the highest rating on their most recent evaluation. In the four districts with binary rating scales for which survey data were available,⁵² 99 percent of probationary and 100 percent of tenured teachers think they should have received the highest rating (Satisfactory) on their most recent evaluation.

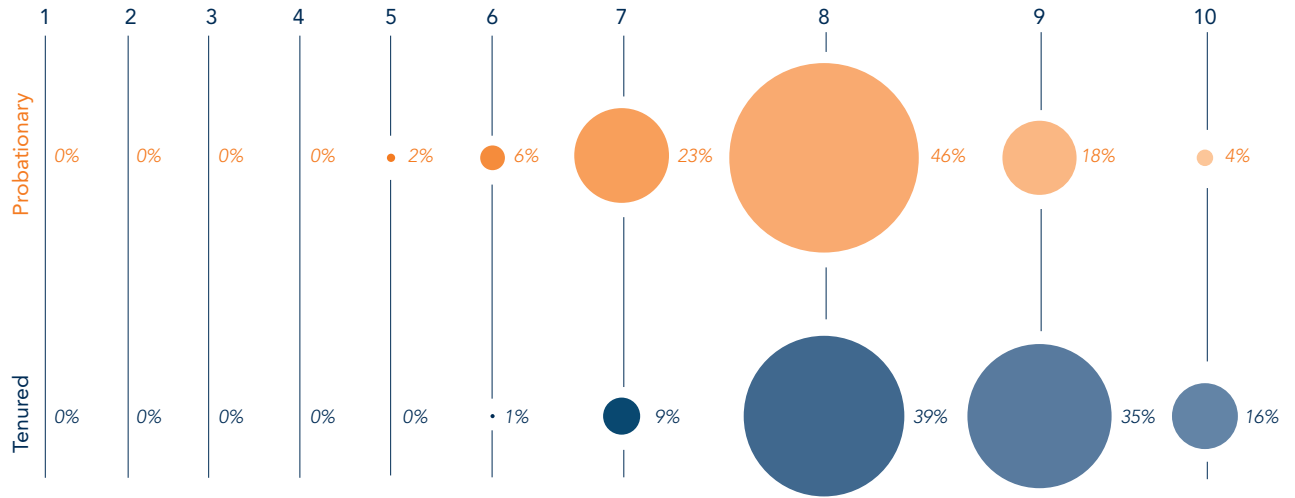
Even teachers who are just beginning their careers believe they deserve the highest performance ratings and are dissatisfied if they are rated good, not great. This inflated sense of performance is evident in the self-assessment ratings of novice teachers. In a subset of districts⁵³ where teachers were asked to assess their own instructional performance on a scale of 1 to 10, 69 percent of novice teachers rated their instructional performance an 8 or higher.

“Many teachers are accustomed to receiving a ‘superior’ rating and simply do not accept anything lower. It also seems to be an easier way out for the administrators, rather than have a confrontation with the teacher.”

—Chicago Public Schools Teacher

FIGURE 14 | Teachers' Self Assessments of Instructional Performance

ON A SCALE OF 1 TO 10, HOW WOULD YOU RATE YOUR INSTRUCTIONAL PERFORMANCE?



In a system where negative or even less than perfect performance ratings are given only rarely, teachers naturally develop an expectation that they will be among the large majority considered top performers. In this context, teachers perceive low or negative ratings not in terms of what they communicate about performance but as a personally-directed insult or attack. The response is understandable in the context of the current system, where so few teachers get critical feedback of any kind. When their evaluation does include criticism, they feel as though they have been singled out while other examples of poor performance go unaddressed.

This creates a culture in which teachers are strongly resistant to receiving an evaluation rating that suggests their practice needs improvement. Schools then find themselves in a vicious cycle; administrators generally do not accurately evaluate poor performance, leading to an expectation of high performance ratings, which, in turn, cause administrators to face stiff cultural resistance when they do issue even marginally negative evaluations. The result is a dysfunctional school community in which performance problems cannot be openly identified or addressed.

