



# Reforming Teacher Evaluation: One District's Story

Morgaen L. Donaldson and John P. Papay December 2012



Center for American Progress



# Reforming Teacher Evaluation: One District's Story

Morgaen L. Donaldson and John P. Papay December 2012



# Contents

- 1 Introduction and summary**
- 5 The study setting and the evaluation program**
- 9 Origins of the reform**
- 17 Reform: A process and a product**
- 25 Implementing the teacher evaluation program**
- 29 Outcomes of the teacher evaluation program**
- 35 The current state of the teacher evaluation program and areas for growth**
- 37 Conclusion**
- 41 About the author**
- 43 Acknowledgements**
- 45 Endnotes**



# Introduction and summary

In recent years policymakers have seized on teacher evaluation as a primary lever for improving schools. Of all school factors—from expanded school calendars to smaller class sizes to community and family engagement programs—teachers contribute the most to student achievement.<sup>1</sup> Policymakers reason that evaluating teachers based on their students' performance will lead to the removal of underperforming educators and an improvement in the overall quality of the teacher workforce. Furthermore, there is growing evidence that effective evaluation systems may prompt all educators to improve.<sup>2</sup>

In the past few years, nearly all states have passed legislation revising teacher evaluation. Through Race to the Top, Teacher Incentive Fund grants, and No Child Left Behind waiver requirements, the federal government has pushed states to use student achievement as a major component in teacher evaluations. More than 40 states and Washington, D.C., have responded by changing their laws related to education policy.<sup>3</sup> Teacher evaluation reforms and, in particular, efforts to assess teachers on the basis of student achievement have sometimes resulted in confrontations between teachers and school districts. Chicago's 2012 teachers' strike is only the most recent example of bitter standoffs related to teacher evaluation.<sup>4</sup>

Yet confrontation and conflict are not the dominant themes in all districts seeking to reform teacher evaluation. Some districts have prioritized a collaborative approach to developing new evaluation systems and can provide valuable lessons for others to follow. But what exactly can be learned from such places?

This report examines one district's efforts to develop and implement a new teacher evaluation system in a cooperative manner with its teachers. For reasons of confidentiality, we refer to this district as the Studyville School District, a medium-sized school system in a northeastern urban center with roughly 20,000 students and 1,600 teachers. It has been recognized at both the national and state level for the collaboration that has characterized the development and implementation of the district's new teacher evaluation program and for the use of student

achievement as a benchmark in its the evaluation system.<sup>5</sup> Leaders from the teachers union and district have worked together to adopt a framework for the new evaluation system through the collective bargaining process. They've designed and implemented the new system through joint labor-management committees and are working collaboratively to oversee the program. Teachers and school-based administrators have been central to the reform's design and implementation throughout the process.

To date, Studyville's evaluation program has garnered support from key stakeholders, including teachers, school and district leaders, the teachers union, and the city school board. A key element of the reform is that it has real consequences: Some teachers have been recognized as exemplary while others have left the district because of poor performance. Furthermore, teachers report that the reform focuses their efforts more directly on student performance.<sup>6</sup> While the evaluation program's direct effects on teachers' instruction and student achievement have not yet been examined, stakeholders' favorable views of the program thus far suggest that the reform may be gaining traction in the district and that it could improve these key outcomes.

In this paper we present an in-depth case study charting the evaluation program's progress—from concept to initial design to implementation to the program's current state. We examine aspects of policy and practice that have facilitated or impeded the program's success and pay particular attention to developments in labor-management relations at the school and district level affecting the reform's development.

Based on our interviews with key district and union leaders and with a cross-section of teachers and school leaders, we are able to make the following observations:

- **Economic, political, and policy factors have facilitated the teacher evaluation program's development and acceptance.** These factors help to explain the reform's favorable review among many stakeholders, including teachers, school-based administrators, and district and union leaders.
- **Collaboration has been at the heart of the teacher evaluation program's creation and development.** The program was conceived through the district's 2009 collective bargaining agreement with its teachers union. District leaders and union leaders worked together to create the evaluation program's framework and oversee its implementation. Teachers and school administrators also collaborated to develop many of the key components of the reform.



- **The teacher evaluation program represents both a process and a product.**

The collective bargaining agreement laid out a framework for the teacher evaluation program and articulated a process through which the district would formalize the program details. This process led to a product, the evaluation system, which satisfied at least some of the needs of all constituents. Both the process and the product of the reform are important factors in the reform's progress to date.

- **The teacher evaluation program's progress reflects strong leadership**

**coupled with broad input.** Leaders at the city and district level were willing to work to bring the program to fruition, but they didn't do it alone. They relied on teacher- and school-administrator working groups to develop many of the details of the program. The program therefore addressed the priorities of district and union leadership and reflected the needs of teachers and administrators.

Some of what we learned from Studyville's experience with its teacher evaluation program is summarized in the following recommendations:

- **Get out in front of a wave of reform, but be authentic to district needs.**

Our study district has made strides with its teacher evaluation program in part because it seized on momentum to improve teacher evaluation and staked out its ground with a collaborative approach to this area of reform. The school district's status as an early implementer led to recognition from prominent government and labor leaders, which likely reinforced the reform's value to key local stakeholders and encouraged them to continue to search for productive compromise. Moreover, by getting out in front of teacher evaluation reform, the district was able to set its own course for change and design its own system that responded to the district's context and needs.

- **Invest in collaboration, but understand its challenges.** Collaboration played a major role in the progress of this reform. While collaboration was important to the acceptance of the evaluation program, it is not the only ingredient driving reform progress. In the case of the Studyville School District, the context shaped collaboration among the various stakeholders, encouraging them to compromise and find mutually beneficial solutions. The lesson here is that investing in collaboration can lead to a better outcome for all, but this degree of collaboration is hard earned and depends on both the context and the people involved. Leaders of both management and labor should take stock of these factors when seeking to engage in collaborative reform.

- **Pay attention to process and product.** The evaluation program's progress to date is as much due to the process through which it was developed as it is to the structural elements that comprise the reform. As a process, the program was developed collaboratively and over time. While the framework for the evaluation program was laid out in the collective bargaining agreement, the details were developed through joint labor-management committees. This process cultivated trust and understanding among parties and has been critical to the teacher evaluation program gaining traction in the district. This is even more important given that the evaluation program seeks to advance a common vision of effective teaching. The more time teachers and school leaders spend discussing this vision in a structured way, the more likely it is to affect instruction both within and outside of the boundaries of the teacher evaluation program.

# The study setting and the evaluation program

Our study school district, which we refer to as the Studyville School District, is a medium-sized district in a northeastern urban center. More than 70 percent of the district's approximately 20,000 students receive free or reduced-price lunches, and the district performs in the bottom 10 percent of districts on state achievement tests. The teacher evaluation program seeks to increase the instructional quality of its approximately 1,600 teachers.

Implemented in the fall of 2010, the evaluation program requires that school leaders evaluate teachers annually and provide more frequent, informal coaching to improve instructional practice. These evaluations are based on the following three components:

- Student progress towards specific performance-growth goals
- Standards-based observations that follow the district's teacher-evaluation rubric
- Professional conduct

Based on these three components, the evaluator provides a summative rating for teachers: 1 (needs improvement), 2 (developing), 3 (effective), 4 (strong), or 5 (exemplary).

The system requires evaluators to observe and conference with teachers regularly throughout the school year. The evaluator uses standardized rubrics to evaluate a teacher's instructional practice and his or her professional values. Additionally, the evaluator is expected to visit classrooms on a regular basis and provide ongoing informal feedback to teachers.

Student performance comprises approximately 50 percent of a teacher's evaluation score, with the exact weight varying based on teachers' scores on the other components of the evaluation program. Each fall a teacher meets with his or her evaluator to set at least two performance goals (called student-learning objectives) for the academic year. Each goal is based on student growth in key skills

and knowledge related to the subject or grade taught. Elementary teachers in the fourth through eighth grades must create at least one goal based on the state standardized test. Teachers and evaluators again meet at the midpoint of the year to discuss teachers' progress toward meeting their goals. Goals may be modified slightly at that time if both parties agree to the change.

Below we present typical goals that teachers might adopt as part of the evaluation program.

## Teacher goal setting

**Example 1: High school media arts teacher.** In arts and technology classes, where state standardized tests do not exist, students often develop portfolios of work showcasing their mastery of various skills. A media arts teacher might therefore set one goal of having 90 percent of his or her students complete a portfolio. For the second goal, the teacher might use the district's 21st century skills standards and assessments to establish that 90 percent of students should advance one performance level from their fall baseline score on the rubric of essential skills in the strand of communication and collaboration by the end of the school year.

**Example 2: Elementary teacher.** In elementary grades that administer the state test, it is possible to set performance goals that attempt to measure growth on the test. But since state test results aren't published until July, many teachers use district assessments, which are aligned with the state test, to set their goals. In the fall teachers in tested grades administer pretests and set goals based on those results. A teacher could, for example, set a goal of moving students' average score on the state English language arts standardized test from 50 percent to 70 percent "correct" over the course of the academic year.

Based on evidence that they gather in the fall, district evaluators are required to identify potential exemplary and underperforming teachers by early November of the school year. These identifications trigger independent observations by third-party evaluators who are external to the district. These outside observers evaluate teachers concurrently but independently of school leaders, thereby verifying the leader's assessment of a teacher. This serves to "validate" teacher ratings. If both the school leader and the external evaluator rate a teacher as 5 (exemplary) on the instructional practice segment of his or her evaluation, and if the other portions of the evaluation (student growth and professional conduct) are sufficiently high,

the teacher earns an overall exemplary for a summative rating. These highly rated teachers become eligible for leadership positions. On the other hand, teachers scored as needs improvement, earning a rating of 1 from both internal and external evaluators, receive intensive support. If they remain in the needs improvement category by year's end based on the three components of their evaluation, these individuals can be dismissed.

The new teacher evaluation program, therefore, departs from the district's previous system in its mandate that teachers be evaluated annually, its effort to increase the quality and frequency of feedback (both formal and informal), the incorporation of student achievement as a benchmark, and its clear positive and negative consequences for teachers based on their performance.

---

## Study methodology

This paper presents a case study describing the teacher evaluation program's development and implementation, which was guided by the following questions:

- How did the district's teacher evaluation program come to be?
- What has facilitated the program's development?
- What has impeded its success?

---

## Data source and study sample

To gain insight on the teacher evaluation program's development and implementation, we conducted in-depth, semi-structured interviews with a sample of 95 participants: 10 principals, 10 assistant principals, and 72 teachers, including four teachers who were union stewards for their schools. We also interviewed three district-level leaders: the director of the teacher evaluation program, the president of the teachers union, and the assistant superintendent. Finally, we interviewed teachers and administrators at a purposive sample of 10 schools. Half were schools where teachers reported the most positive assessment of the evaluation program in district surveys conducted in 2011, and half were schools where in the same surveys teachers, on average, reported the most negative assessment of the evaluation program. Four sample schools were high schools and six were kindergarten through eighth-grade schools. At each school, we interviewed the principal, assistant principal, and 20 percent to 25 percent of the school's teachers. We selected

the teacher sample based on years of teaching experience and by subject and grade level taught. This allowed us to maximize teacher-rating variation.

We interviewed participants during the 2011-12 school year, the second year of the district's teacher evaluation program. We also conducted interviews at the start of the 2012-13 school year. Interviews were 45 minutes to 60 minutes long; they were audio-recorded and then transcribed. We asked participants to describe the origins of the evaluation program and that factors that they thought facilitated or impeded its success. We also asked specific questions about the role the teachers union played in the creation and implementation of the program. Lastly, we inquired into the role of teachers and administrators (at both the district and school level) in developing and implementing the evaluation program.<sup>7</sup>

# Origins of the reform

Although Studyville’s teacher evaluation program was implemented in the fall of 2010, and the collective bargaining agreement that laid the groundwork for the reform was inked in the fall of 2009, the story of the evaluation program begins much earlier. Studyville had been building a foundation for evaluation reform and labor-management collaboration for more than a decade.

---

## An orientation toward collaboration

The district’s teacher evaluation program grew out of a larger collaborative climate among key stakeholders, including the city’s mayor, the superintendent, and the president of the teachers union. The mayor, who came to office in the early 1990s, has administrative authority over the school district. In addition to being a member of the school board, the mayor appoints the board’s seven other members. In the past decade the mayor has made education a primary focus of his leadership. He secured funds for a major school reconstruction project, an effort that has led to all of the city’s schools being newly constructed or renovated during the past 10 years.

The superintendent of the school district assumed his post a few years before the mayor came to office, so he was on the job for nearly two decades when the teacher evaluation program was adopted. During his tenure, the superintendent has overseen successful initiatives to increase the involvement of parents and caregivers in the education of their children. He also put an end to social promotion—the practice of promoting students to the next grade level based on their age rather than their demonstrated skills—in the district.

Unlike his long-tenured counterparts, the president of the district’s teachers union, a longtime educator in the system who worked as a teacher and an instructional coach for 28 years, came to office in 2007. The president was elected in a tight vote, after losing a close race for the union presidency in 2001. He promised

when elected to take a more “activist” role, enforcing the teachers’ contract more strictly and focusing attention on supporting less-experienced teachers.

These three leaders cultivated personal relationships with each other over time. One union steward, commenting on the relationship between the three leaders, said, “They actually get along fairly well, and they argue but they seem to always come to some kind of commonality, which is pretty good. Otherwise nothing would happen.”

This relationship bore fruit during the 2008-09 school year, when the district and union began to negotiate a new collective bargaining agreement. In the end both sides agreed on not only “bread-and-butter” issues such as salaries and benefits, but also on a major school-reform initiative that, along with the new teacher evaluation program, included an effort to increase college enrollment rates for the district’s students and a provision of differentiated levels of autonomy for schools based on their performance. The union steward credited the success of the negotiations to the relationship between these three men in these negotiations, stating that the contract’s success “is very dependent on the people. It’s very dependent on the union guys, the top-level administration guys who are on the [school] board, a superintendent, and his couple of underlings really being together for an extended period of time.”

### Why focus on teacher evaluation?

As contract negotiations commenced, leaders on both sides agreed to reform teacher evaluation as one of three focal points in the effort to improve student performance in the district. They made this decision because all parties—the district, school leaders, and teachers—realized the evaluation system in place at the time wasn’t working and needed to be improved. The director of the district’s new teacher evaluation program recounted that under the prior evaluation system, “tenured teachers only had to be observed once every five years.” He added that it was “unheard of to terminate a tenured teacher for incompetence.” The union president agreed the old system was broken, saying that teacher evaluation had been “mixed bag.” He explained that “some principals were very thorough” but that many of the classroom observations were no more than “drive-bys” and “superficial at best.” He pointed out that teachers “wanted reform of evaluation,” seeing it as an opportunity to “grow professionally.” To sum up he said, “We all recognized this needed to be done. Teacher evaluation was broken.”



Widely held frustration with the teacher evaluation process in the district was brought into high relief with the hiring of an external firm to survey educators about their experiences with teacher evaluation. The results were overwhelmingly negative. Neither teachers nor principals thought teacher evaluation was working. At the same time, publications such as “The Widget Effect” were advancing teacher evaluation as a key lever for school reform.<sup>8</sup> These factors help to solidify the district’s decision to focus on improving teacher evaluation as a vehicle for improving student achievement.

---

## Roots in collective bargaining

Although the district could have created a new evaluation system with little union involvement, district leaders made a deliberate choice to set down the roots for the new teacher evaluation program in the 2009 collective bargaining agreement between the school district and the teachers union. The contract with teachers and the negotiations that produced it were significant for two reasons. First, it visibly signaled a stronger partnership between labor and management than had previously existed in the district, reflecting a shift toward a more collaborative and experimental approach to reforming schools on the part of the city, school leaders, and teachers. Second, in negotiating the teachers’ contract, district and union officials focused on the principles underlying teacher evaluation reform and embedded these in the final contract. The details of the district’s eventual teacher evaluation program were resolved outside of the formal contract negotiations. This approach prevented the program’s development from becoming bogged down by unrelated concerns and reduced the possibility of the evaluation program being held hostage to bread-and-butter issues such as wages and benefits.

While the contract received largely positive reviews as an important step forward in teacher evaluation reform, several news editorials argued that the reform was not as dramatic as it at first seemed to be because it was not embedded in the terms of the teachers contract. The president of the teachers union said some roundly criticized the final contract, asking, “But where’s the TEP [teacher evaluation program]?” The assistant superintendent said others disparaged the contract, calling it “an agreement to agree.”

District and union leaders, though, believed that this perceived weakness was actually a key strength. The union president argued that attempting to hammer out the details of evaluation reform during negotiations would have been too

cumbersome. Instead it was decided to root the teacher evaluation program in the contract through a side-letter agreement. This agreement was developed concurrent to the negotiations over wages and working conditions, but, as the assistant superintendent described, “at an entirely separate table” and in a “separate room.” This physical and symbolic division of “bread-and-butter” issues from reform concerns was “important,” according to the assistant superintendent. Relatively free from the strictures that govern collective bargaining, representatives, at what was termed the “reform table,” were able to have, in the words of the assistant superintendent, “a freewheeling discussion of reform issues.” In addition to district and local union leaders, state-level union leaders, representatives from the mayor’s office, and lawyers with expertise in labor relations were seated at the “reform table.”

Shortly after the conclusion of contract negotiations, the president of the teachers union faced the daunting challenge of securing agreement from his membership. According to the union president, it was clear in discussions leading up to the vote on the proposed contract that some teachers were skeptical about the idea of teacher evaluation. Several union stewards mentioned that some teachers thought it was what they termed a “salary dump”—a convenient and easy way to get rid of higher-paid veteran teachers and reduce the district budget during tight economic times. As one union steward observed, “It’s just very scary, especially with the economy the way it is. I mean, teachers who have been here for a while pretty easily could be replaced by three teachers who would cost the district as much money as that one longstanding, tenured teacher.” The union president admitted that his membership also had concerns about being fired if their “students did lousy on the test.” Eventually, union leaders convinced the skeptics and calmed fears, winning membership support for the contract and the new teacher evaluation program. The union president credited his members’ buy-in to the fact that “teachers felt more confident because they had input.”

Had the teachers rejected the proposed contract, the district and union would have entered into a binding arbitration process to resolve the contract—an outcome neither side wanted. An external arbitrator would have reviewed the various demands and made a binding judgment regarding how negotiations would be settled. This threat of binding arbitration pushed teachers and administrators toward accepting the contract, as union and district leaders felt that any alternative imposed by an external arbitrator would have been worse. Ultimately, teachers voted by an overwhelming margin of approximately 22 to 1 to ratify the contract, which gave them what they sought—the maintenance of their health benefits and annual pay raises over the four-year life of the contract.

Administrators also received much of what they wanted—a commitment to the teacher evaluation program and an agreement that the process for creating the evaluation system would be developed in a joint labor-management committee and spelled out in a side-letter agreement to be appended to the contract. The details of the teacher evaluation program would subsequently be developed through committees of teachers and administrators. The president of the teachers union stated that the inclusion of the teacher evaluation program in the contract via a side-letter agreement was significant: “This time it [the evaluation program] was in the contract, whereas other reforms weren’t.” He said this meant that the district “couldn’t walk away” from the program.

---

## Facilitating factors

The teacher evaluation program, and the contract that created it, represented a significant change for the school district. It required leaps of faith from teachers, school administrators, and district and union leaders. Not only did it establish a whole new way of evaluating teachers and leaders, but it also advanced a new process for reform through collaboration. If the district or union had backed out, or if teachers or principals had lost faith in the effort, the reform effort could have folded. What enabled the program to grow and move steadily forward?

Studyville’s investment in the teacher evaluation program was facilitated by a number of things, most significantly the economic context, the policy context, and the political context. One union steward argued that these three factors, along with the contract negotiations themselves, spurred the district to create the evaluation program:

*The mayor wanted to do something different. The union contract was coming. The president [President Barack Obama] had it on his table with education, and the economy was an issue at the same time. ... those four things kind of came together to allow [Studyville] to really put this together and jump out of the box. And if it was not a union negotiating year, I don’t think we’d have seen the same results. If it was not a bad economy, we might not have the same results.*

## The economic context

Study participants agreed that the economic context—the prevailing economic winds of the late 2000s—played a large role in the evaluation program’s adoption. The stock market plunged in 2008 and bottomed out in March 2009.<sup>9</sup> The local economy was struggling as well. The mayor was projecting large budget shortfalls and warned that layoffs might occur. A union steward explained the economy’s influence on the teachers’ willingness to support the evaluation program.

*“We were afraid to lose our jobs. In other towns, [teachers unions] were going out and demanding arbitration, losing arbitration ... and not voting for their new contracts. ... as bad as the economy is right now [fall of 2011], it was much scarier two years ago when all this was going on. The banks were failing. Most people were saying: ‘We’re entering a great depression.’ So I think most people in the union were like—‘I’m working. I’m going to work today.’ When we took the vote [on the contract], there were people that dissented but ... it was a very small minority.”*

The union president echoed this steward’s sentiment, suggesting that the economy forced teachers and administrators toward compromise, which played out in the contract negotiations and later in support of the reform.

---

## The policy context

The policy context—the landscape of education reform in 2009—gave Studyville stakeholders reasons to collaborate on teacher evaluation reform. Nationally, policymakers were focused on reforming teacher evaluation systems that were widely perceived as broken. In July 2009 President Obama launched the Race to the Top initiative, which included teacher evaluation as a major focus of reform and provided incentives to states to permit teachers to be evaluated based on student test scores. Moreover, it encouraged districts and unions to work together to improve schools. The mayor, superintendent, and president of the teachers union recognized that their proposed teacher evaluation program fit several of the specifications of Race to the Top. The mayor wanted the school district to adopt a collaborative teacher evaluation reform that would make it more competitive in seeking federal education grants, including Race to the Top. In fact, when the state applied for Race to the Top later in 2009, it included the district’s new contract with its teachers in its application.

---

## The political context

Lastly, the political context made Studyville ripe for reform. First, the mayor was up for re-election and wanted to make education a focal point of his campaign. Being involved in a high-profile struggle with the teachers union could have put a damper on his re-election bid. Second, the mayor wanted the school district to be seen as a model of collaborative school reform. As a result, he was willing to compromise to ensure that reform happened and that the process went relatively smoothly.

Furthermore, there was growing rhetoric on the national stage that was critical of teachers—and of their unions in particular. As one building principal noted:

*[T]he message has been very clear you know, from Washington and throughout the country: “Teacher, you are the problem and we are going to face it. We’re going to hold you accountable to the nth degree for student learning. We don’t want to hear about excuses ... and we have the instruments to be able to measure that [student learning].”*

Just like in Studyville, teachers and their unions across the country were looking for good press to combat negative portrayals in the media. Some district teachers and union leaders saw the teacher evaluation program as a way to demonstrate their openness to reform while also preserving a role for themselves in its design. This need also led to involvement from state and national union leaders, who provided support and encouraged a collaborative outcome that could be a model for agreements across the country. For this reason, leaders from the state and national union affiliate were present in the negotiations that led to the development of Studyville side-letter agreement.



# Reform: A process and a product

With the teacher evaluation program ensconced in the contract via the side-letter agreement, the district and teachers union went about designing the details of the reform. They began by developing a set of common beliefs to guide discussions.

---

## A collaborative process

The design process leading to the development of the district’s teacher evaluation program was explicitly collaborative and deliberate. In late 2009 the district formed three committees to carry out the district’s broader reform agenda: a citywide reform committee consisting of three teachers, three administrators, two parents, and the assistant superintendent (*ex officio*); a joint committee devoted to developing the teacher evaluation program composed of six teachers and six administrators; and a committee devoted to developing teacher and caregiver surveys composed of four teachers and four administrators. Teacher and administrator working groups supported the work of these committees.

The teacher evaluation program joint committee met on alternate weeks. During the “off weeks,” members of the joint committee met with working groups to develop specific pieces of the reform model and also to solicit input on ideas from the joint committee from their constituencies. This iterative process continued throughout the winter and spring of 2010, when the teacher evaluation program committee made recommendations to the school board regarding the program. The program’s development was therefore explicitly and deliberately collaborative. The director of the evaluation program observed, “Collaboration has been a huge part of our process and development of our process. I don’t think another district could just adopt the evaluation program. A large part of why it works is because of the conversations we’ve had along the way.”

Furthermore, the process ensured that teachers played a central role in the reform's development. "Teachers had tremendous input," said the program's director. "Any teacher could participate in the development of the program," she added. The union president described attending joint evaluation reform meetings every other week and meeting with a working group of teachers at the union office on the off weeks. Some 40 or 50 teachers would sometimes attend the meetings to hear updates from representatives of the working groups and to work on issues that arose in these sessions.

Even beyond the involvement of individual teachers, the union was deeply engaged in the process. "The union was very heavily involved from the beginning and they continue to be," said the evaluation program's director. A principal stated, "I think they [the union] were great. I mean, there were union people on the committee that I was on. ... it was a collaborative effort and I think everyone realized that—everyone realized that this is what's right."

In addition to teachers, administrators also participated in working groups and had their input incorporated into the process. One principal who was a member of a working group emphasized the collaborative aspect of the evaluation program's creation and implementation: "We met many, many, many times. And it was definitely a collaborative effort." Discussing the teacher evaluation instrument, she added, "I'm very happy with it—just the ability to have your meetings that you are making the goals together. It's not top down."

---

### The benefit of an extended development timeline

Participants in our study emphasized the benefits of time spent developing the teacher evaluation program. They said that taking the time to develop the system collaboratively had three main benefits: enhanced trust between labor and management, the opportunity to communicate the purpose and details of the evaluation program to important constituencies, and ultimately, a more robust evaluation system.

The time spent developing the teacher evaluation program permitted the designers to consider a broader range of possible logistical glitches and unintended consequences, thus preventing them from arising. According to the union president, the investment of time on the front end led to a better system. "There were lots of work and meetings—tons of input. "We argued about it. We thought about all the



hypotheticals.” He noted that because of the initial effort, the actual implementation of the program “went smoothly.”

The time spent in the development stage also provided the opportunity for administrators and teachers to gain a better understanding about what constitutes good instruction. The union president explained, “We knew we wanted to improve instruction but didn’t have a good handle on what good instruction was. ... there were 48 principals and 48 different conceptions.” To develop a shared framework, he said, “principals and teachers worked side by side defining effective teaching.” This work drew on Charlotte Danielson’s “Framework for Teaching,”<sup>10</sup> as well as on the personal insights of teachers and administrators about effective teaching. Importantly, it also included a rubric. Although the district had previously used Danielson’s work, the inclusion of a rubric and subsequent discussions of what the rubric meant in terms of concrete practices made the district’s instructional expectations more explicit this time around.

The time spent also allowed relationships to develop. “We spent one school year working on [the teacher evaluation program] collaboratively,” noted the union president. “That year spent is the best thing we did. It couldn’t be done in three months.” He went on to explain that because of the initial groundwork that went into developing the reform, its rollout and implementation has been less problematic. “We never got completely stuck because of the preparation we did in year one.” As a result, when issues arose with the program, the process never broke down. Instead the union and district leaders have continued to communicate relatively openly to resolve any differences.

Lastly, the time spent up front enabled leaders to communicate the purpose and details of the evaluation program to teachers and school administrators as they were worked out. This proved to be very important because when the system was implemented in the fall of 2010, some details were still being decided. “I spent a year messaging, keeping everyone apprised,” said the union president. This messaging continued in the program’s first year, the 2010-11 school year. One assistant principal recalled that her building leader relayed messaging constantly to ease teachers’ concerns. “The principal kept saying, ‘This is new. They’re introducing it. Don’t panic. It’s not meant to hurt anyone.’”

She added that the union was communicating deliberately as well. “One of the things that our union kept doing, which I loved, was saying to the teachers—‘We were part of this. We agreed to this. ... this is something that was really thought

out. It's not meant to hurt.” Another principal offered a similar observation, saying that “real clear messaging from the union” at the beginning of the program’s second year served to “clarify any other misconceptions” teachers were holding onto from the first year. “I think we opened this year [the second year] crystal clear,” he said.

While the new evaluation system inevitably encountered bumps in the road, the time that was spent developing the teacher reform led to a smoother implementation. Clearly, none of the challenges that arose derailed the process or compromised the support from teachers and administrators, in large part because these groups felt that their voices had been heard in program design. They believed that the program had been designed with care, that all parties had forged strong relationships, and that the details of the system’s key elements had been clearly communicated to teachers and administrators.

---

## System design

In addition to the collaborative process used in developing the teacher evaluation program, the district made six key decisions in designing the system that facilitated the reform’s acceptance and perceived success.

First, leaders on both sides decided to emphasize the reform’s potential to affect teachers’ practice through coaching rather than its potential as a mechanism to dismiss underperforming teachers. Both sides recognized, however, that some underperforming teachers needed to leave the classroom, and the evaluation system provided for that outcome. The district wanted to recognize exceptional teachers, but the bulk of the district’s attention has remained on improving the instructional practices of teachers. In the words of the evaluation program’s director, “The most powerful aspect of the program is the conferences and conversations [between teachers and administrators], not the ratings [of the teachers].”

Although this comment could be construed as proverbial “spin” to make the new system more palatable to the union membership, the district’s actions actually lend further credence to its authenticity. The district has not focused on evaluating the rigor of goals, but has instead invested in the coaching relationship between teachers and administrators. As the evaluation program’s director noted, “We haven’t checked goals [student-learning objectives] ... because in some ways that would detract from the teacher/administrator relationship and ownership over

goals.” This decision, she said, was made deliberately in order to ensure that teachers and evaluators could set goals that were most likely to bring about instructional improvement.

A second key design decision was to incorporate student achievement into teacher evaluation by using student-learning objectives rather than value-added measures, which are based solely on external calculations using standardized test scores. This was a deliberate choice on the part of designers. The decision was based in part on practical concerns about value-added measures. The director of the evaluation program noted that all teachers can develop student-learning objectives, “but VAMs [value-added measures] can’t apply to everyone.” This is because many teachers don’t teach in subjects and grades that have adequate student testing. Additionally, the director noted that the real value of student-learning objectives comes from the fact that, unlike value-added measures, they are jointly developed between teachers and administrators, which gives teachers’ a sense of ownership. She said that the district’s teacher evaluation program “is not a hard measure of student achievement. Its strength is that teachers are more invested in their goals.” The assistant superintendent elaborated, saying, “In districts using value-added, there is a lot of discussion about validity and accuracy and consistency ... [and] the primary goal is sorting teachers.” By contrast, he said, “our priority is the formation of that professional learning culture” based on “professional conversations and feedback loops.” He added, “we’re far from perfect,” but “we realized that it [the teacher evaluation program] didn’t have to be perfect to be good.”

Although their measurement properties can be questioned, student-learning objectives support the designers’ focus on promoting conversations and self-reflection on the part of teachers. It is also worth noting that the use of student-learning objectives reflects the notion that teachers would be more motivated to improve their practice if they believed in the importance of their goals and understood what it would take to meet them, two areas in which districts using value-added measures have struggled.<sup>11</sup>

Third, the district’s teacher evaluation program required annual evaluations for teachers and administrators. Participants—both teachers and administrators—said that it was critical that teachers be evaluated more frequently than was the case with the earlier evaluation system. Teachers are now officially observed several times during the year, and evaluators are expected to conduct informal classroom visits and provide feedback even more regularly. Annual evaluations, though time-consuming to conduct, aligned with some of the policy recommen-

dations receiving vocal support in 2009.<sup>12</sup> Furthermore, they signaled to teachers and administrators the importance that the district placed on evaluation.

Fourth, the school district instituted evaluations for school- and district-level administrators that paralleled the teacher evaluation program. The district also created surveys for teachers and community members that factored into the evaluations of school administrators. This effort to promote “top to bottom accountability” was cited by administrators and teachers as a key aspect of Studyville’s larger reforms that facilitated the teacher evaluation program’s success. As one principal noted, “It’s not just the teacher. The principal is [also] being looked at. It’s what I’m being rated on as well, so it’s us. It’s not the administration [on one side] and the teachers [on the other]. It’s us together moving the children.” Administrators also said that the more comprehensive and inclusive evaluation system had the added benefit of giving school leaders the opportunity to model how to be evaluated, allowing them to cultivate trust and teamwork with their faculty. One principal said, “One of the things I did was when I self-evaluated for my evaluation, I shared it with them [teachers]. So they saw that I gave myself a couple of 2s [in areas] that I was developing on. So that really built trust once I did that.”

A fifth key feature was the inclusion of an external validator in cases where teachers were identified as potentially receiving a 1 (needs improvement) rating or 5 (exemplary) rating. This validation process increased teachers’ trust in the evaluation program. Providing another layer of teacher confidence is the fact that union leadership reviews the list of external evaluators and can remove anyone it does not want in the position. Administrators and validators observe the same lesson and rate it independently. For teachers in jeopardy of being rated 1, this validation must occur for three separate lessons. Teachers can only receive a 1 or 5 designation if, at the end of the evaluation cycle, both the evaluator and external validator give the teacher’s instruction the same rating and the other portions of the evaluation are sufficiently low or high, respectively.<sup>13</sup> Teachers felt the inclusion of the outside validator was critical, particularly in the case of low-rated teachers who are at risk of losing their jobs. “I think the greatest thing in the teacher evaluation program is the validator,” said one teacher who had initially received a rating of 1.

Lastly, a key aspect of the design was the creation of rubrics to guide observation of instruction. The program rubric described effective instruction more specifically than the district had ever described it before. Because observations were becoming more high-stakes, the school district had to take steps to standardize observations within and across observers. Relying on these instructional rubrics

made this work easier: Evaluators could point to specific practices rather than rely on broader judgments of a teacher's performance. But the creation of rubrics also promoted ongoing learning and development on the part of teachers. As one principal said, "I definitely was a cheerleader when it came to the rubric that teachers were able to take to look at and say 'Okay. Well, this is where I am and this is where I want to be.'" Participants reported that the rubrics allowed them to understand what was expected of teachers.



# Implementing the teacher evaluation program

The new teacher evaluation program represented a substantial change from the prior system. Despite the time spent developing it, implementation was not simple. It required school-level administrators, teachers, and district-level administrators to adopt different behaviors related to evaluation, and the learning curve was steep for many of these educators. As a result, several challenges arose in the program's first year.

One major challenge related to the training of administrators to carry out the program. Like many evaluation reforms, Studyville's program hinges on the will and skill of administrators and requires them to devote much more time to evaluation than many had been used to devoting. Not only does the program require administrators to rate teachers using a standards-based rubric, it also requires more conferences with teachers than the previous evaluation system mandated. Moreover, the district's new teacher evaluation program relies on administrators working with teachers to set "rigorous and reasonable" student-learning objectives. Both teachers and administrators were uncertain about what "rigorous and reasonable" meant or how to operationalize it going into the program.

It was a challenge to provide sufficient training to administrators to allow them to implement the program fully in 2010, its first year. "No one had been trained by July," recalled the union president. Summer trainings were held, but "training was limited and compressed," according to the president, and some administrators could not attend the sessions. Administrators and teachers reported that administrators did not have a consistent understanding of the process or specifics of the evaluation program when it launched in the fall of 2010. As a result, the union president observed, "lots of administrators weren't implementing it correctly in the first year. They defaulted back to their old practices."

The district recognized this inconsistency as problematic and hired a full-time director of teacher evaluation early in the 2010-11 school year. Since then the director has worked to increase the standardization of practices across evaluators,

ensured that trainings happened, communicated about the system to teachers and evaluators, and provided support to those implementing the system. The person appointed to the position had served as a longtime teacher and administrator in one of the city's high schools and, according to participants, was well-regarded by all stakeholders. This helped her achieve credibility from the start and be accepted as a fair manager by all sides.

A second challenge concerned data management. With administrators rating teachers' instruction at regular intervals, setting and revising student-learning objectives with teachers throughout the year, and considering other forms of data simultaneously, along with keeping track of the data, became a major concern. In the program's first year there was no good system in place. Case in point: Administrators routinely emailed Microsoft Word documents with key details and teacher ratings back and forth to central office. This process was inefficient and unnecessarily time-consuming for evaluators, and it also raised privacy concerns.

According to one principal, the teacher evaluation program's "minutia was difficult." Her assistant principal provided the example of the difficulty of tracking program data in the first year: "[Teachers] would type something on their teacher evaluation [document] then I'd have to get it on mine or we would have to replace mine. At the end they were giving me their data and I was actually typing it in for them." She added that the process was stressful for teachers and that she offered to enter their data because, as she told them, "This should not make you crazy." She said everyone quickly realized that this was not an effective or efficient way to manage data. In the program's second year, the district put a new online system in place to track educator performance.

A third challenge related to the limitations of the evaluation program from the perspective of school administrators. Despite efforts from district and reform program leaders to frame the evaluation system in terms of teacher development, some school administrators perceived the system as an opportunity to remove persistently under-performing teachers without following the necessary district policies and procedures. Like the previous system, however, the teacher evaluation program has an established process that administrators must follow and safeguards to protect teachers from unwarranted efforts to remove them. One union steward said:



*A lot of the principals came right out of the box labeling [teachers] 1s on previous stuff, not on what was going to happen in the teacher evaluation program. It was kind of was like housecleaning. Our particular principal came in and said [to one teacher], 'You're a 1.' She didn't even read the goddamn teacher evaluation program. ... she just didn't like the teacher out of the box. And so she [the teacher] wasn't a team player within the first month, so automatically she wasn't scored well on her evaluation and from then on, it was game on. So as a union rep, I spend a lot of time defending that position and trying to implement the evaluation program correctly.*

The district's director of the evaluation program voiced a similar concern, albeit less forcefully. She reported that some administrators have been disappointed that teachers they perceived as underperforming ultimately remained in district schools. For the lowest- and highest-performing teachers, student performance counts for more than 50 percent of the final rating. As a result, the administrator's assessment of instruction does not represent the teacher's final evaluation score. This practice limits the impact of administrators' subjective views of teachers' practice. The director explained the effect of this, saying, "For a lot of teachers [in danger of dismissal], quite honestly, student-learning growth saves them."



# Outcomes of the teacher evaluation program

The purposes of the evaluation system are to raise student achievement in the district through several mechanisms: by providing feedback to all teachers so they may improve, by recognizing excellent educators, and by providing a clear, efficient, and fair method for removing teachers who consistently underperform. Studying the overall effect of the teacher evaluation program on student achievement is challenging because the district adopted several reforms at the same time, which makes it hard to isolate the contribution of the evaluation program. There is strong evidence, however, that the evaluation program affected each of these mechanisms.

The clearest outcomes of the program are in the number of teachers who have been validated as 5 (exemplary) and the number of teachers who have left the district for performance reasons. In Table 1 below, we present these outcomes during the program's first two years. As seen in the top panel, 36 teachers were validated as exemplary during the 2010-11 school year, and 66 teachers were validated as exemplary the following school year. Interestingly, evaluators initially identified many more teachers as exemplary in 2011-12 than in 2010-11, but a smaller proportion of these teachers were validated as exemplary by the end of the year.

**TABLE 1**  
**Teacher outcomes in the teacher evaluation program**

	2010-11	2011-12
<b>Total number of teachers evaluated</b>	<b>1846</b>	<b>1457</b>
Exemplary teachers		
• Initially identified as exemplary (5)	40	95
• Validated as exemplary (5)	36	66
Underperforming teachers		
• Initially identified as needs improvement (1)	75	58
• Validated as needs improvement (1)	38	19
• Validated as developing (2) or effective (3)	29	20
• Resigned mid-year	8	19
Resignations due to performance		
• Total resignations	34	28
• Tenured teachers	16	17
• Nontenured teachers	18	11

It's not surprising that the teacher evaluation program has led to a substantial uptick in teacher departures for performance reasons. As seen in the second panel of Table 1, 75 teachers in 2010-11 and 58 teachers in 2011-12 were identified as potential 1s (needs improvement) at the beginning of the year. Some of these teachers improved and were validated at a higher performance level by year's end. Others saw the proverbial writing on the wall and resigned from the district before the final evaluations were completed. Of those who were validated as 1 (needs improvement) at the end of the year, many were notified that they would be dismissed. A few teachers were given a second year to demonstrate improvement, and most of these did indeed improve their instructional practice.

All of the educators who were notified that they would be dismissed ultimately left the district voluntarily. In total, district leaders reported that 34 teachers, including 16 tenured teachers, left the district because of their evaluation ratings in 2010-11. This represented approximately 1.3 percent of tenured teachers and 2.8 percent of nontenured teachers in the district. Twenty-eight teachers (1.9 percent) left the district after the 2011-12 school year. Four principals or assistant principals left in the program's first year, and after the program's second year, four administrators departed for performance reasons. The exit of tenured teachers for performance reasons is a marked departure from the district's record under its previous evaluation system. As noted above, the evaluation program's director said, "In terms of tenured teachers, it was probably unheard of that anybody was terminated for incompetence [prior to the new evaluation system]."

Interestingly, participants noted that the evaluation system has made it more challenging to non-renew nontenured teachers. According to the director, prior to the program, “it was pretty easy to non-renew someone. You just basically had to meet the deadlines and have evaluated them.” The union president concurred. “Under the previous system nontenured teachers had no rights pre-April 1,” of their first year, which is when their legal protections increased. “Administrators could just tell them they were fired. ... now administrators have less leeway to non-renew.” The director of the evaluation program said that tenure and evaluation “law hasn’t changed but we’re treating them the same as we’re treating tenured teachers.” As a result, she explained, nonrenewal “is not just automatic.” This has caused “frustration for principals,” the director noted. But the process has become fairer for both nontenured and tenured teachers, and it is now aligned closely to the evaluation system.

Largely because the evaluation process was clear, providing teachers with direct information about their performance, and because the union worked with the district to oversee the program, underperforming teachers chose to leave voluntarily. “We have not gone to termination hearings on anyone,” reported the evaluation program’s director. The district has instead encouraged teachers to sign separation agreements that offer departing teachers a few extra months of pay and avoid legal proceedings. The director explained that such agreements “allow people to resign with dignity and to really make a decision on their own in terms.” She added, “It’s better to pay somebody for a couple extra months than to fight litigation on a termination.”

District and labor leaders have viewed these results differently. While asserting that the teacher evaluation program is “consequential” based on the number of teacher departures, the district continues to place more emphasis on the program’s capacity to improve instruction. District leaders regularly cite the number of teachers who were identified as potentially in the needs improvement category early in the school year but showed sufficient improvement to be rated higher by the end of the year.

Union leaders have framed the number of teachers who have departed the district for performance reasons as a positive to their members. Teachers have been generally supportive of these moves because they recognized that some teachers needed to be dismissed. They also saw that the administrators were not using the evaluation system capriciously to fire teachers. The union president said that teachers felt the number of teachers dismissed was lower than many feared. “At the end of year one the district didn’t fire 300 teachers. Teachers realized ‘He’s [the union president] right, they didn’t fire everybody.’” The union president concluded, “there is much more acceptance [of the evaluation program] now.”

In short, the increased identification of teachers as highly effective educators, the streamlined pathway out of the classroom for underperforming teachers, and the focus on instructional practice throughout the district has led the program to earn the support of a range of stakeholders.

---

## Facilitators

Probably the most influential facilitator of the teacher evaluation program's continued acceptance and success has been the involvement of teachers and their union in the reform's initial design and ongoing implementation. Many teachers said that the involvement of teachers and the union helped them have faith in the reform. One high school science teacher said that she "liked the idea that it was put together by the union, the teachers, the district—that everybody participated in the process." When asked whether she received less resistance to this program than to the prior evaluation system, one elementary school teacher responded with a resounding "Yes." She also said that when teachers expressed nervousness about the new evaluation program she would respond:

*It's just what we've all signed up to do and your union voted on it. Your union was actually probably more involved in the process than the administrators' union in crafting it. ... it's your document. You guys came up with it. You approved it. ... it wasn't an administrator-created document like the past.*

The director of the evaluation program said that this aspect of the reform was especially important when challenging and potentially divisive decisions were being made. "It's made a huge difference to be able to tell teachers, 'This is what you as a group decided you want.' When we have those difficult discussions it's really important."

More broadly, the ongoing relationship between the teachers union and district leadership has facilitated the evaluation program's success. Several district and union leaders noted that stakeholders had begun the process of developing the teacher evaluation program by drafting a set of shared understandings. The program's director recounted that it "started from a common set of beliefs and using them as a touchstone. When it gets heated, we come back to agreements. That's helped." She added, "We make a conscious effort to keep collaborating. It's more time-consuming. It's easier to just make decisions sometimes, but it's important to have the conversation first."

Both the president of the teachers union and the director of the evaluation program reported that there have been no grievances with the evaluation process filed to date. “We try to work it out first with the administrator,” said the president of the teachers union. If this doesn’t solve the problem, the issue goes to the union president and teacher the director of the evaluation program. The president concluded, “Most issues are resolved at the building level” rather than through the district’s grievance procedure.

Representatives of both labor and management pointed out that discussions could at times get contentious, and sometimes both sides needed to suspend discussion. “We collaborate but we don’t always agree,” explained the union president. “It’s not all sweetness and light,” added the assistant superintendent. Critically, both sides were able to return to the initial guiding principles, and discussions usually commenced again within days.

Key to this ongoing collaborative relationship was the willingness to compromise on both sides. Even though, as the teacher evaluation director said, “there’s not always complete happiness” between the district and union, “without [the union’s] support we would not be moving on [dismissing] teachers.” The union’s willingness to support the district’s efforts to remove ineffective teachers has meant that the district has had to be especially diligent about holding school administrators accountable for carrying out the evaluation program fairly and with fidelity. The program’s director said that she has been forced to have “tough conversations with principals” about sticking to component weights and rating a teacher lower than his or her student-performance indicators and classroom instruction would indicate.

The district has also taken extra steps to make sure that the evaluation process is fair and transparent. As the director of the evaluation program explained:

*I think we’ve built in mechanisms to ensure that it’s fair. I think the validation goes a long way with that. I also think that when we looked at those hard cases at the end of the year, we almost bend over backwards to ensure that the process has been followed with fidelity [and] if there’s any question about it not being fair that we err on the side of the teacher.*





# The current state of the teacher evaluation program and areas for growth

Though most study participants queried viewed the district's teacher evaluation program positively, many suggested concrete ways in which the system could be improved. "I think we've made a lot of progress. There are a lot of positive outcomes," said the evaluation program's director. She said, however, that "there's still a lot of work we've got to do." Many participants said that there needed to be more standardization of the process and the ratings among administrators. The director, for example, named "firming up the process" as a key area of improvement. By this she meant standardizing the process across all administrators. One way to accomplish that is to focus on calibration in observation and student-learning objective goal setting across administrators, which is seen as a key area of weakness, according to both district leaders and teachers. "We need to get better at that. Sometimes goals [student-learning objectives] are too high or too low," said the program's director. She added that having more consistency across evaluators will help to facilitate even greater acceptance of the evaluation ratings.

Another area of concern noted by district leaders involved the possibility that there might have been some inflation of ratings in the first two years of the evaluation program. Speaking about the number of teachers in the 4 (strong) category on summative ratings, the program's director said, "It's great if there actually are so many [strong teachers]. But it may be skewed upwards. Maybe goals aren't as rigorous as they could be. It doesn't jibe with what we see in student achievement." Continuing, she said, "We expected to see more teachers in the 'developing' category given the number of new teachers every year. I think we're going to see, as administrators get better calibrated in instructional practice and goals, maybe a downward correction." Creating the needed rating balance will also involve resetting teachers' expectations, acknowledging that being rated as "developing" is not a bad thing, particularly for early-career teachers.

Additionally, the district is working to provide administrators with more time to observe teachers and discuss instruction. As the union president said, this involves rethinking assumptions about time use. "Do you really need a \$125,000 adminis-

trator to do bus duty?” he asked. Furthermore, the evaluation program’s director has worked to streamline the evaluation process to make the reporting burden less cumbersome. The director noted that the new online data management system has made it easier for evaluators to report data and for her to track the process at the teacher, administrator, or school level. This also allows her to provide more support and accountability for school administrators.

Finally, the district is attempting to shift away from the practice of labeling teachers using a numbers system to a rating system that assigns teachers performance labels. “Psychologically, [a numerical label] can be very damaging,” said one union steward. Hearing this complaint, the district is making an effort to move to a qualitative labeling model using terms such as “effective” and “exemplary” instead of numbers. Although the district has always had qualitative labels to accompany numerical ratings, administrators and teachers have tended to use numerical ratings in practice. The district is now making a more concerted effort to use only the labels in all presentations and discussions about the program within and outside the district.

# Conclusion

With the Studyville School District's evaluation program well into its third year this fall, the question is: What lies ahead?

Participants in our study were optimistic about the program's continued existence. The president of the teachers union said that the evaluation program's relatively smooth first two years were a good sign for its continuation. "I was pretty confident that things would go well, but who really knew because [it was a] brand new system in a school district that's not huge but not small. There [were] capacity issues. We have money problems like everybody else. It's kind of surprising that it did go so well."

When asked whether he thought the evaluation program would survive the next contract negotiation set to begin the following school year (2012-13), one union steward said that he thought the program would remain in place. "Now that it's there, it's almost going to be impossible to get out. And we've been on national television and our superintendent and our union president have been down to Washington to talk about the successes of educational reform." He added, "The general public is out here saying, 'Schoolteachers all make too much money. Schoolteachers only work half a year. Schoolteachers this, school that.' It is at least some way to hold teachers accountable, so I don't see it going away." Moreover, study participants generally felt positive about the reform, and as such most thought it should continue to exist in the district.

In sum, our findings indicate the following:

- **Economic, political, and policy factors facilitated the development and success of the teacher evaluation program.** These factors played a large role in our study's school district in creating the will of key stakeholders to adopt teacher evaluation reform and invest in it over time. These factors help explain the reform's favorable review from teachers, school-based administrators, and district and union leadership.

- **Collaboration was at the heart of the teacher evaluation program's creation and development.** The teacher evaluation program in the Studyville School District was conceived through the district's 2009 collective bargaining agreement with its 1,600-member teachers union. Teachers and school administrators have worked together to develop many of the key components of the reform. Both district and union leaders created the framework through which the program was created, and they continue to oversee it in a generally collaborative manner. The program therefore represents collaboration between both teachers and administrators and the union and district leadership. It is worth noting that the one district in the country where teacher evaluation has been shown to improve student achievement, Cincinnati, also initiated its teacher evaluation system through a collective bargaining agreement.<sup>14</sup> Moreover, like Studyville's teachers union, the Cincinnati Federation of Teachers has been integral to the development and implementation of its teacher evaluation system.
- **The teacher evaluation program represents both a process and a product.** Both the process and product of the evaluation system are aspects of the reform that rely on each other, and they are both important factors in the reform's progress to date. The program was developed collaboratively and over time. The time spent developing the program in joint labor-management committees cultivated trust and understanding among parties. Many contingencies were also worked out in these meetings, and the product of the reform reflects different participants' viewpoints. Participants argued that this process has been integral to stakeholders' support for the program.
- **The teacher evaluation program's progress reflects strong leadership coupled with broad input.** Leaders at the city and district level were willing to lead, to push the envelope, and to compromise. But they did not design the program alone and impose it on their constituents. They relied instead on teacher and administrator working groups to develop many of the details of the program. The program was therefore framed to address the priorities of district and union leadership. The specifics, however, reflected both teachers' and administrators' needs.

To date, Studyville's teacher evaluation program has garnered support from key stakeholders: teachers, school and district leaders, the teachers union, and the city school board. The reform has led to consequences, with some teachers being recognized as exemplary and others leaving the district for performance reasons. There is some indication that teachers feel the reform focuses them on student performance.<sup>15</sup> While the program's direct impact on teachers' instruction or

student achievement has not yet been examined, stakeholders' favorable views of the program suggest that the reform may be gaining the traction in the district that would allow it to affect these key outcomes. So what can other school districts learn from the experience of teacher evaluation reform in our study district?

- **Get out in front of a wave of reform, but be authentic to district needs.**

Studyville has made strides with its teacher evaluation program by responding early to calls to improve teacher evaluation. This has allowed Studyville to stake out its ground in this area of reform. Studyville's status as an early implementer and its collaborative stance to evaluation reform led to recognition from prominent government and labor leaders, which likely reinforced its value to key local stakeholders and encouraged them to continue to search for productive compromise. Moreover, by getting out in front of teacher evaluation reform, Studyville was able to set its own course for change and design its own system. To some extent, then, it was able to influence state policy rather than be confined by it.

Obviously, however, not all districts can get out in front of reforms. What can districts that engage in the process of reform later than Studyville learn from its example? Even those who are not the first to engage in change can seize the opportunity to design creative solutions to key problems. When leaders see a wave of reform building, then can use this pressure to convene stakeholders to devise productive plans for moving their system forward.

- **Invest in collaboration but know its challenges.** Collaboration played a major role in the progress of this reform. But one should not conclude that collaboration, in and of itself, is the key ingredient to a reform's progress based on this study alone. In the case of our study district, the context shaped collaboration among the parties, encouraging them to compromise and find mutually beneficial solutions. The lesson here is that investing in collaboration can lead to a better outcome for all, but that collaboration is hard earned and dependent on context and people. Both management and labor leaders should take stock of these factors when seeking to engage in collaborative reform.

- **Pay attention to process and product.** The teacher evaluation program's progress to date is as much due to the process through which it was developed as it is to the structural elements that comprise the reform. The time spent developing the program has been critical to its gaining traction in the district. Specifically, the time spent by teachers and administrators and district and labor leaders in designing the program helped to develop trust and understanding among

the parties. It also facilitated one of the main goals of the reform: to develop a common vision of effective teaching. The more time teachers and school leaders discuss this vision in a structured way, the more likely it is to affect instruction within and outside of the boundaries of the program.

To truly improve teacher evaluation, changes must occur not only in the structure of teacher evaluation but also in the culture that surrounds and supports it. Teachers must become more comfortable having others, whether they are administrators or teachers, observing their instruction and offering constructive feedback on how to improve. Attention to the process of developing a new teacher evaluation system may have greater success in facilitating these cultural changes than an exclusive focus on crafting the structural elements of an evaluation system.

On the whole, our study participants reported that the first two years of the teacher evaluation program in the Studyville School District have been challenging but productive. In the challenges and successes of Studyville we see opportunities for what teacher evaluation might become, with teachers becoming more skilled through the process and students benefiting in the end.

---

## About the author

**Morgaen L. Donaldson** is an assistant professor of educational leadership at the University of Connecticut's Neag School of Education and a research affiliate at the university's Center for Education Policy Analysis. She is also a research associate at the Project on the Next Generation of Teachers at Harvard University. As a researcher, she studies policies and practices related to teacher quality, teacher evaluation, school and district human capital development, and teacher unions. Her recent publications include "Strengthening Teacher Evaluation: What District Leaders Can Do" in *Educational Leadership* (2012); "Teach for América: The Latinization of U.S. Schools and the Critical Shortage of Latino/a Teachers" in the *American Educational Research Journal* (2012); and "The Price of Misassignment: The Role of Teaching Assignments in Teach for America Teachers' Exit from Low-Income Schools and the Teaching Profession" in *Educational Evaluation and Policy Analysis* (2010). A former public high school teacher, Donaldson was a founding faculty member of the Boston Arts Academy, Boston's public high school for the arts. She also served as a project manager in an initiative funded by the Gates Foundation to replicate the practices of small schools that effectively served low-income and minority students. Donaldson holds a doctorate and a master in education degree from the Harvard Graduate School of Education and a bachelor's degree from Princeton University.

**John Papay** is an assistant professor of education and economics at Brown University. His research focuses on teacher policy, the economics of education, and teacher labor markets. He has published on teacher value-added models, teacher evaluation, teacher preparation programs, teacher compensation, high-stakes testing, and quantitative program evaluation methodology. He recently co-authored "Redesigning Teacher Pay: A System for the Next Generation of Educators" with Susan Moore Johnson. His current research focuses on teacher development throughout the career and the role of the school context in teachers' work. He is a research affiliate with the Project on the Next Generation of Teachers at Harvard University. A former high school history teacher, he earned his doctorate in quantitative policy analysis from the Harvard Graduate School of Education.





---

## Acknowledgements

We are grateful to the assistant superintendent, the president of the teachers union, the director of the teacher evaluation program, and the teachers and school leaders who shared their experiences with us. We would also like to thank the district leaders who granted us access to the district and facilitated the contacts with schools that enabled us to conduct this research. We thank Glenda Partee and Kate Pennington of the Center for American Progress for comments on earlier drafts. All errors are our own.



---

## Endnotes

- 1 See for example: Daniel Aaronson, Lisa Barrow, and William Sander, "Teachers and student achievement in the Chicago public high schools," *Journal of Labor Economics*, 25 (1) (2007): 95–135.
- 2 E.S. Taylor and J.H. Tyler, "The Effect of Evaluation on Teacher Performance," forthcoming from the *American Economic Review*.
- 3 S. Headden and E. Silva, "Lessons from D.C.'s Evaluation System," *JSD*, 34 (6) (2011): 40–44.
- 4 Monica Davey and Steven Yaccino, "Teachers Union in Chicago to Extend Strike Into 2nd Week," *The New York Times*, September 16, 2012, available at <http://www.nytimes.com/2012/09/17/education/chicago-teachers-union-extends-strike.html?pagewanted=all>.
- 5 All proper names in this paper are pseudonyms.
- 6 Morgaen L. Donaldson, "Teachers' Perspectives on Teacher Evaluation Reform" (Washington: Center for American Progress, 2012).
- 7 To analyze the data, we coded the interview transcripts using open, axial, and selective coding. Using thematic summaries, categorical matrices, and analytical memos, we conducted cross-case analysis to identify emerging themes across participant experiences.
- 8 Daniel Weisberg and others, "The Widget Effect" (New York: The New Teacher Project, 2009).
- 9 E.S. Browning, "Bears Are Wary as Bull Returns," *The Wall Street Journal*, March 27, 2009, available at <http://online.wsj.com/article/SB123811041706752903.html>.
- 10 Charlotte Danielson, "The Framework for Teaching," The Danielson Group, 2012, available at <http://www.danielsongroup.org/article.aspx?page=frameworkforteaching>.
- 11 Eva L. Baker and others, "Problems with the use of student test scores to evaluate teachers" (Washington: Economic Policy Institute, 2010).
- 12 Weisberg and others, "The Widget Effect."
- 13 Technically, teachers could receive a 1 or a 5 at year's end if they had not been through the validation process, but these ratings would not carry any consequences and these teachers would need to have their ratings validated the following year.
- 14 Taylor and Tyler, "The Effect of Evaluation on Teacher Performance."
- 15 Donaldson, "Teachers' Perspectives on Teacher Evaluation Reform."

---

The Center for American Progress is a nonpartisan research and educational institute dedicated to promoting a strong, just, and free America that ensures opportunity for all. We believe that Americans are bound together by a common commitment to these values and we aspire to ensure that our national policies reflect these values. We work to find progressive and pragmatic solutions to significant domestic and international problems and develop policy proposals that foster a government that is “of the people, by the people, and for the people.”

---

Center for American Progress

