RUTGERS
Graduate School of Education

January, 2014

New Jersey's Pilot Teacher Evaluation Program:
Year 2 Final Report

William A. Firestone, Timothy L. Nordin,

Anton Shcherbakov, Dessi Kirova, and Cynthia L. Blitz

Introduction

In March 2011, the New Jersey Educator Effectiveness Task Force proposed a plan for teacher evaluation and recommended a new state educator evaluation system. This was the first step in a process that was followed by the New Jersey Legislature's 2012 passage of the TEACHNJ Act, which revised laws governing teacher evaluation, and the State Board's adoption of revised teacher evaluation regulations in September 2013.

New Jersey's evaluation system aims to be a major tool for improving student achievement and promoting equity in New Jersey by providing a firmer basis for awarding teacher tenure and providing guidance for improving teaching practice (New Jersey Department of Education, 2011). The program is also part of a much broader national effort. More than a decade of federal and state-level legislation mandating teacher assessment and accountability policies, including Race to the Top competition requirements, provides the backdrop for the current reforms to enhance teacher quality and preparation. While there is broad agreement that high-quality, effective teaching contributes to student learning, assessing teachers' effectiveness and the quality of their teaching has proven to be very challenging. Researchers, politicians, and practitioners are grappling with fundamental questions about what constitutes effective teaching, which aspects of teaching are most likely to improve student learning, how to best measure teachers' effectiveness, and how student assessment data should be used and for what purposes. Not surprisingly, for most of the recent past, teacher evaluation has not been particularly systematic. The use of student data to assess teachers only began to be seriously considered in the late 1990s (Wright, Horn, & Sanders, 1997). In the past, personnel decisions were often linked to formal credentials—degrees, credits, and years of experience—for a variety of reasons, including weak measures of teaching quality (Podgursky & Springer, 2007).

During the 2011–2012 academic year, the New Jersey Department of Education (NJ DOE) launched a pilot teacher evaluation program to help further the State Educator Evaluation System. Expanded in 2012–2013, this program is part of NJ DOE's commitment to "elevating the teaching profession, recognizing classroom excellence, and providing support to educators needing assistance" (New Jersey Department of Education, 2012, p. 5). Expanded in 2012–2013, the pilot teacher evaluation program has two primary elements: measures of student growth and observational assessments of teaching practice. All school districts in New Jersey will implement the new program beginning in the 2013–2014 academic year.

To support this effort, the NJ DOE contracted with the Rutgers University Graduate School of Education (RU GSE) to conduct an external review of the pilot teacher evaluation program. The scope of this review included reporting on the implementation of new evaluation practices, documenting participants' perceptions of the pilot programs, and identifying factors that influenced the implementation process. The RU GSE research team collected data from various sources but primarily used administrator and teacher surveys, site visit interviews, and focus group data to understand the implementation of the pilot teacher evaluation program. This report summarizes the evaluation's findings.

The RU GSE evaluation research of the pilot focused on three major questions:

1. What were administrators' and teachers' perceptions of the implementation of the new teacher evaluation practices?
2. What were teachers' and administrators' orientations or beliefs about new teacher evaluation procedures?
3. What factors are perceived to be barriers and facilitators to program implementation?

The answers to each of these questions are discussed in the following sections. A critical issue that could not be examined was the relationship between measures of teacher effectiveness based on classroom observation and student achievement because student growth scores were not available at the time this research was conducted.

The pilot was crucial for testing New Jersey's new evaluation system and identifying potential operational problems before all districts in the state were required to follow new regulations in evaluating their teachers. While a small program could achieve that goal, it is important to recognize that, due to the small size of the pilot, the use of perceptual data from interviews and surveys, and the fluid nature of the implementation, these findings might not be representative of all that is likely to be observed about teacher evaluation as the system becomes operational statewide in the 2013–2014 academic year. We hope that this report will be useful to school districts in New Jersey as they implement and improve their teacher evaluation practices, and to educators and policy makers in other states as they develop and refine their own approaches to teacher evaluation.

Methods

The Year 2 assessment was a continuation of the work reported on in the first-year report (Firestone, Blitz, Gitomer, Gradinarova-Kirova, Shcherbakov, & Nordin, 2013). The evaluation began when RU GSE and NJ DOE finalized the first Memorandum of Understanding on January 26, 2012. The RU GSE team worked closely with the NJ DOE leadership to refine the assessment strategy for the second year within the overall scope of the project.

Participating Districts

The ten school districts selected by the NJ DOE to participate in the first year of the pilot program were joined by fifteen additional districts for Year 2 of the pilot. The Cohort 2 districts began their work in August 2012. Table 1 provides information on each district's location, district factor group, and student enrollment. Lenape Valley and Stanhope, as well as Audubon, Collingswood, and Merchantville, entered the pilot program as consortia. The districts vary substantially in enrollment and geographic distribution. Although the districts are not a representative sample of all districts in the state, they are quite diverse. Eight districts are from northern New Jersey, nine are from southern New Jersey, and eight are from central New Jersey. In addition, a mixture of low and high socioeconomic school districts is represented in the pilot sample as indicated by the distribution from A to J district factor grouping (DFG)[1] in Table 1.

---

[1] The NJ DOE introduced the District Factor Grouping (DFG) system in 1975. This system provides a means of ranking school districts in New Jersey by their socioeconomic status with 'A' districts being the poorest in the state and 'IJ' districts being the wealthiest (NJ DOE web site, http://www.state.nj.us/education/finance/sf/dfgdesc.shtml).

The three lowest DFG categories make up 31% of the districts in the state, and they constitute 24% of the pilot districts. The middle three DFG categories make up 46% of the state's districts and represent 60% of the sample districts. The Year 2 cohort also included two districts from the two highest categories (I and J), which make up 24% of the state's districts and are 12% of the sample. Middlesex County Vocational was not assigned a DFG. In terms of enrollment, the pilot contained a variety of districts ranging from very small and suburban (e.g., Alexandria and Stanhope) to rural and mid-sized (e.g., Pemberton and Bordentown) to very large and urban (e.g., Elizabeth).

Table 1 – *Participating School Districts in Pilot Teacher Evaluation*

District	County	Location	2000 DFG	Enrollment
Alexandria	Hunterdon	Central	GH	477
Audubon	Camden	Southern	DE	1,489
Bergenfield	Bergen	Northern	FG	3,510
Bordentown	Burlington	Central	FG	2,495
Collingswood	Camden	Southern	FG	1,862
Cranford	Union	Northern	I	3,877
Elizabeth	Union	Northern	A	23,988
Freehold Borough	Monmouth	Central	B	1,481
Gloucester City	Camden	Southern	B	2,027
Haddonfield Borough	Camden	Southern	J	2,482
Lenape Valley	Sussex	Northern	GH	775
Middlesex County Vocational	Middlesex	Central		1,886
Merchantville	Camden	Southern	DE	374
Monroe	Middlesex	Central	FG	6,042
Ocean City	Cape May	Southern	DE	2,084
Pemberton Township	Burlington	Central	B	4,994
Piscataway	Middlesex	Central	GH	7,361
Red Bank	Monmouth	Central	CD	1,134
Rockaway Township	Morris	Northern	I	2,426
Secaucus	Hudson	Northern	DE	2,199
Stanhope	Sussex	Northern	GH	356
Teaneck	Bergen	Northern	GH	3,729
West Deptford	Gloucester	Southern	DE	3,004
Woodbury City	Gloucester	Southern	B	1,503
Woodstown-Pilesgrove	Salem	Southern	FG	1,597

Primary Sources of Information for the External Assessment

Data came from surveys and site visits which included administrator interviews and teacher focus groups.

Surveys. The RU GSE assessment team designed two online surveys in Year 2—one targeted toward all teachers participating in the pilot program and the other toward administrators in pilot districts. Survey questions for teachers and administrators overlapped a great deal, although certain questions were specific to each group and to members of each cohort. These surveys were modeled on the Year 1 surveys and modified to address the updated goals of the assessment. Preliminary work included reviewing prior state evaluations of other teacher practice instruments, as well as testing individual survey questions and the entire surveys. Once the RU GSE assessment team finalized the questions and response categories, surveys were created in Qualtrics, an online survey tool. Both the administrator and teacher surveys were approved by Rutgers University's Institutional Research Board, ensuring that data collection efforts complied with federal and university requirements for the protection of human subjects. The surveys were successfully administered as a result of the ongoing efforts of our assessment team in cooperation with district administration and NJ DOE staff.

Administrator survey. We worked with the pilot Teacher Evaluation project directors in all 25 pilot districts to identify respondents for the administrator survey who had direct involvement in the evaluation system. The survey was distributed to these administrators between March and April 2013. The response rate for the survey was 62% with 280 responses. Cohort 1 had a 61% response rate with 155 responses and Cohort 2 had a 63% response rate with 125 responses. Some analyses in this report compare second-year data from 25 districts with first-year data from the 10 districts in Cohort 1[2]. In both years, the surveys asked administrators about the numbers and quality of the observations they completed, in addition to their experiences with and perceptions of the implementation of the new teaching practice evaluation instruments in their districts.

We generated questions for the administrator surveys to reflect the main components of the implementation process: we asked about the choice of teacher evaluation framework; experiences with training; the number and quality of observations; and perceptions of the quality, fairness, usefulness, and ease of use of the selected teacher evaluation framework. Please refer to Appendix B for the Year 2 administrator survey.

Teacher survey. The teacher survey was initially distributed electronically in May 2013 and was available through June 2013. The overall response rate was 39%, with 2,926 responses. District response rates ranged from 11% to 100%. Most districts fell in between, with a median response rate of 36%. The response rate for a similar survey given to Cohort 1 districts in May through June 2012 was 59% with a total of 2,495 respondents. Although this represents a significant drop in response rate, we believe this was due to other surveys being conducted in the school districts that composed our sample. Numerous respondents replied that they had already completed our

---

[2] During the first year of the pilot, we surveyed administrators in the spring and again in late summer. The spring survey had a 60% response rate with 154 completed surveys, and the second-year administrator survey had a 54% response rate with 134 completed surveys.

survey when we first sent them out. Despite the drop in response rate, there is still good reason to believe that our response rate of 39% adequately captured a representative sample of the population. Indeed, Keeter, Kennedy, Dimock, Best, and Craighill (2006) demonstrated that surveys with a 25% response can provide comparable results to those with a response rate of 50%. In their study, 77 out of 84 comparisons were statistically indistinguishable between the two surveys, and the differences in proportions ranged from four to eight percentage points.

Our assessment team organized questions in the teacher survey to parallel the topics covered by the administrator surveys so that we could compare the information that is common to both populations. The questions followed the main components of the pilot implementation: we asked about the selected teacher evaluation framework, about training on the district-selected framework for teacher evaluation, and about experiences with the pilot teacher evaluation program in general. Additionally, we asked participants about their perception of the quality, fairness, usefulness, and ease of use of the selected teacher evaluation framework. The teacher survey appears in Appendix C.

Site visits. Site visits to 11 school districts (one consortium of two districts) between December 2012 and May 2013 allowed for collection of interview and focus group data from key stakeholders in piloting districts, including both teachers and administrators.

Sample of school districts for site visits. The RU GSE assessment team conducted site visits in 11 of the 25 participating school districts, including two in one consortium. Budgetary limitations did not allow visits to all pilot districts. In selecting districts to visit, we secured representation of districts from different regions (northern, central, and southern New Jersey), as well as those of different student enrollment sizes and DFGs. The site visits included four districts from Cohort 1 visited in Year 1 and seven districts from Cohort 2.

Targeted participants in piloting districts. We developed a list of key individuals in pilot districts who were directly involved with the implementation of the pilot teacher evaluation program. The list included:

- the superintendent,
- the key instructional leader in charge of curriculum and instruction,
- the district Teacher Evaluation project director,
- the director of professional development,
- the director of student data,
- the president of the local teachers' association, and
- two principals (elementary and secondary).

We worked with the pilot teacher evaluation program project directors to schedule site visits and, for all but one district, we met with all individuals on the list. For one district, because of scheduling constraints, the director of professional development was unavailable. Interview times ranged from 20–60 minutes. In addition, we conducted two focus groups consisting of teachers in each district. One focus group was comprised of elementary teachers and the other of secondary teachers. Districts were asked to provide a range of content specialties, grade levels,

and special educators for the focus groups. Focus groups consisted of 4–10 teachers and lasted from 30–60 minutes.

Interview and focus group topics. Members of our assessment team conducted all interviews and facilitated all focus groups according to the site visit guide protocols (please refer to Appendix A). Interview guides were developed to ensure common lines of questioning across visits while granting the interviewer/facilitator flexibility to prompt interviewees for further detail, elaborations, and clarifications as needed. The interview guides focused on the following topics organized around the two types of data used to evaluate teachers:

1. The Teacher Practice Instrument

    a) Choice of framework for teacher evaluation
    b) Training on the new framework
    c) Collection of teacher observation data
    d) Quality of observations
    e) Impact of the pilot program on professional development, professional collaboration, and school culture

2. Student Assessment Data

    a) Tested subjects
    b) Untested subjects

The focus group protocol focused on the following items:

1. Teacher Observations

    a) Compare and contrast current observations with past observations
    b) Sources of knowledge about teacher observation data
    c) Teacher practice instrument and planning, supervision, and professional development
    d) Expertise of observers
    e) District collaboration

2. Student Assessment Data

    a) Tested subjects
    b) Untested subjects

Data Analysis

Analysis of collected data began as results from the administrator survey became available and continued as other data sources provided more information about the ongoing implementation of the teacher evaluation pilot program. The mix of various data sources—interviews with participants, focus group data, survey data, and artifacts of the teacher evaluation process—allowed us to triangulate many of the results derived from these separate analyses and put together a clearer picture of the pilot teacher evaluation program implementation.

Survey data. Data from the administrator and the teacher online surveys were exported from Qualtrics for analysis in SPSS. Descriptive statistics were generated in order to facilitate comparisons across roles (teachers vs. administrators) and cohorts.

Interview and focus group data. The RU GSE assessment team developed a site visit guide for analysis of qualitative data for Year 2 of the pilot program. The site visit guide provided guidance for consistent data collection, including a list of procedures for site visitors to follow: types of information about each of the pilot sites visited, the types of data collected at each site, documents needed after each visit, procedures for handling consent forms, procedures to be followed for taping of interviews/focus group recordings, and the timeline for generating a site visit report following the visit. Site visit reports were compiled from the recorded interviews as a means of condensing and summarizing the progress noted in each district. This method promoted rapid turnaround of analysis and synthesis across districts and cohorts, ensuring that burgeoning themes and insights were not lost during data collection.

Spreadsheets were used to compare districts. As suggested by Miles and Huberman (1994), these matrix displays allowed the RU GSE assessment team to summarize information about events or expressed sentiments around particular issues in a format that facilitated understanding and comparison. This allowed for comparisons across both roles (teachers vs. administrators) and cohorts and facilitated comparing site visit findings with those from the surveys.

The above-described methodology combining survey and site visit data allowed our assessment team to address the questions that guided the pilot teacher evaluation program in 2012–2013. The study findings are presented in the following three main sections:

1. Perceptions of implementation of the teacher evaluation systems: the instrument chosen, the quantity of observations completed, and progress in developing student growth measures, including districts' progress in implementing and using Student Growth Percentiles (SGPs) and Student Growth Objectives (SGOs).

2. Teachers' and administrators' beliefs about teacher practice instruments and student growth objectives: how teachers and administrators assessed the fairness and accuracy of the evaluation instruments and the incentive quality of the evaluation program for teachers.

3. Perceived barriers to and facilitators of implementation: state guidance for pilot districts, the training provided to teachers and administrators, the time taken to conduct observations, and continuing issues involving the data-management tools.

All sections explore the differences in Cohort 1 districts from Year 1 to Year 2 of the pilot as well as differences across Cohorts 1 and 2. The report concludes by addressing possible future challenges that New Jersey will face as the state moves to full implementation of the new teacher evaluation requirements in the 2013–2014 school year.

Implementation

Districts from Cohort 2 were notified that they had been awarded Teacher Evaluation pilot grants in August 2012. Many Cohort 2 districts had already begun to incorporate elements of the pilot requirements prior to officially being notified of the award and thus were prepared to begin formal implementation early in September. Cohort 1 districts began to refine the processes related to teacher observation and worked to develop student growth measures during the second year of participation. This section first describes the teacher practice instruments adopted by each district, then reports on numbers of observations completed during the 2012–2013 academic year before turning to progress made in implementing Student Growth Objectives.

Teacher Practice Instruments Adopted

Each district had latitude to select among a set of state-approved teacher practice instruments. Districts also selected a data-management system that observers would use to record in-class observations, write and store reports, and communicate with observed teachers and other administrators. The data-management systems were also used to generate post-observation reports, teachers' responses to those reports, and summary evaluations. Typically, the data-management system and teaching practice evaluation instrument were bundled together. As the New Jersey pilot teacher evaluation program was coming online, some providers of teacher practice instruments were upgrading their instruments and changing relationships with providers of data-management systems. Most notably, the Danielson Framework for Teaching redefined some of its observation categories in 2012. After that time, it required new users of the framework to use Teachscape as its data-management system. Cohort 1 districts that used the framework typically continued to use protocols and data systems selected in Year 1. Table 2 details the observation protocol and data-management system for each of the participating pilot districts.

Table 2 – *Districts' Selected Teacher Evaluation Framework and Data-Management Systems*

District	Teacher Evaluation Framework	Data-Management System
Cohort 1		
Alexandria	James Stronge	Oasys
Bergenfield	Danielson	In-House
Elizabeth	Danielson	iObservation
Monroe	Marzano	iObservation
Ocean City	Danielson	iObservation
Pemberton	Danielson	Teachscape
Red Bank	Danielson	Teachscape
Secaucus	Danielson	Teachscape
West Deptford	McREL	McREL
Woodstown-Pilesgrove	McREL	McREL
Cohort 2		
Bordentown	Danielson	Teachscape
Collingswood/Audubon/Merchantville	Danielson	Teachscape
Cranford	Danielson	Teachscape
Freehold Borough	Danielson	Teachscape
Gloucester City	Danielson	Teachscape
Haddonfield	Danielson	Teachscape
Lenape Valley/Stanhope	Danielson	Teachscape
Middlesex County Vocational	McREL	McREL
Piscataway	Danielson	Teachscape
Rockaway Township	Danielson	Teachscape
Teaneck	Danielson	Teachscape
Woodbury City	Marzano	iObservation

Each protocol addresses multiple facets of practice and guides observers in making a rating on each of a set of dimensions identified in the protocol. While the protocols vary in their particulars, recent research has demonstrated strong correlations among protocol results when judging the same lessons (Bill and Melinda Gates Foundation, 2013). A brief summary of the teacher evaluation frameworks used by pilot districts follows.

Purpose. All four teacher evaluation frameworks used during the pilot (Danielson, Marzano, McREL, and James Stronge) have a dual purpose—one aspect is the use of the system for improvement of teacher accountability through teacher evaluation, and the second aspect is to improve teaching quality by offering targeted professional development for individual teachers, based on their performance.

Research-based. All four teacher evaluation frameworks reflect research-based standards of teaching quality. The McREL evaluation instrument and accompanying process is based on elements of a 21st-century education and a set of rigorous research-based standards (www.mcrel.org), as is the Danielson model, which is aligned to the Interstate Teacher Assessment and Support Consortium (INTASC) standards (Danielson, 2011). Marzano (2011) claims to have evidence of a causal link between the teaching characteristics observed in his model and increased student achievement. The James Stronge framework is based on seven practice-tested teacher performance standards (Stronge, 2006).

Framework content. As can be seen in Table 3, all four teacher evaluation frameworks consist of multiple domains/standards designed to address all aspects of teaching.

Table 3 – *Categories Used in the Four Teacher Practice Instruments*

Danielson	Marzano	McREL	James Stronge
Domain 1. Planning and Preparation	Domain 1. Classroom Strategies and Behaviors	Standard 1: Teachers demonstrate leadership.	Standard 1: Professional Knowledge
Domain 2. Classroom Environment	Domain 2. Planning and Preparing	Standard 2: Teachers establish a respectful environment for a diverse population of students.	Standard 2: Instructional Planning
Domain 3. Instruction	Domain 3. Reflecting on Teaching		Standard 3: Instructional Delivery
Domain 4. Professional Responsibilities	Domain 4. Collegiality and Professionalism	Standard 3: Teachers know the content they teach.	Standard 4: Assessment of/for Learning
		Standard 4: Teachers facilitate learning for their students.	Standard 5: Learning Environment
			Standard 6: Professionalism
		Standard 5: Teachers reflect on their practice.	Standard 7: Student Progress

Reported Success in Conducting Required Numbers of Observations

The Notice of Grant Opportunity (NGO) to which the pilot districts responded each year specified the number of observations participating districts would be required to complete, much as the regulations written for the whole state would specify the number of observations all districts must complete beginning in the fall of 2013. The NGOs also specified the numbers of certain types of observations—e.g., the number to be at least 30 minutes long or to be preceded by a preconference or to be unannounced—that would be required for different types of teachers each year. These minimum specifications of different types of observations for different teachers was one way the State of New Jersey sought to regulate quality of teacher practice assessment.

The required number of observations for teachers in Year 2 was altered from the first year of the Teacher Evaluation Pilot, increasing the number of formal observations for teachers in the core content areas. The number of observations required also varied with teachers' tenure status. Non-tenured core content teachers were required to be observed five times, while a tenured, non-core content teacher was required to be observed only twice. Regulations also spelled out the variety of observational forms to be used, including dual-observed sessions, unannounced observations, and length of observations, in the *2012–2013 Teacher Evaluation Pilot Toolkit* (State of New Jersey Department of Education, 2012). Informal observations of teachers were not a requirement for Year 2 of the pilot. The formalization of all observations resulted in an increase in total observations for Year 2. Table 4 shows required observations for both pilot years and the regulations for full implementation in the 2013–2014 school year. In addition to the changes shown in in the Table 4, NJ DOE adjusted the length of required observations from year to year

and modified the acceptable qualifications for those conducting teacher observations. The final requirements are somewhat less demanding than those set for Year 2.

Table 4 – *Minimum Number of Required Formal Observations for Teachers of Different Status in Different Years*

Teacher Status	School Year								
	2011-2012		**2012-2013**				**2013-2014**		
	Non-Tenured	Tenured	Non-Tenured Core	Non-Tenured Non-Core	Tenured Core	Tenured Non-Core	Non-Tenured Years 1–2	Non-Tenured Years 3–4	Tenured
with **post-conference input and feedback** (post-conference must occur within 10 days)	3	2	5	3	4	2	3	3	3
requiring a **pre-conference**	3	2	1	1	1	1	1	1	1
unannounced	0	0	2	1	2	1	1	1	1
conducted by an **external evaluator**	0	0	2	1	2	1	0	0	0
that must be a **minimum of 30 minutes**	3	2	2	1	2	1	2	1	0
double-scored	0	0	1	0	1	0	0	0	0
informal observations	2	2	0	0	0	0	0	0	0

Administrators were surveyed between March and April 2013 to assess their confidence that they could complete the required number of observations. Data from the project directors show that Cohort 1 project directors were overall much more optimistic than Cohort 2 project directors. Table 5 shows that project directors in both Cohort 1 and Cohort 2 were most optimistic about their ability to satisfy requirements for observations that provide feedback to teachers (with 70% of Cohort 1 project directors saying that the district is likely to satisfy those requirements and 46% of Cohort 2 saying the same).

Table 5 – *Project Directors' Perceptions in March and April 2013 Regarding Their Likelihood of Satisfying Requirements of the Teacher Evaluation Pilot*

			Likely to satisfy	Possible to satisfy	Unlikely to satisfy
Cohort n=11	1	Number of observations per teacher	70%	10%	20%
		Number of double-scored observations	40%	40%	20%
		Number of observations with an external evaluator	60%	30%	10%
		Number of observations that provide feedback	70%	20%	10%
		Number of observations that are a minimum of 30 minutes	70%	20%	10%
Cohort n=13	2	Number of observations per teacher	23%	38%	38%
		Number of double-scored observations	38%	38%	23%
		Number of observations with an external evaluator	38%	23%	38%
		Number of observations that provide feedback	46%	46%	8%
		Number of observations that are a minimum of 30 minutes	46%	46%	8%

Data collected by the NJ DOE suggest that in spite of this lack of confidence by project directors during the spring of 2013, most districts did complete the required number of observations. The six of 10 Cohort 1 districts that provided information on the number of completed observations averaged the required three observations per teacher and the seven of 15 Cohort 2 districts that reported their number of observations averaged 3.3 observations per teacher. Although the NJ DOE report concluded that these requirements were met during 2012–2013, the requirements were reduced for 2013–2014 with statewide implementation of teacher evaluation in recognition of the strain that the number of required observations caused (Evaluation Pilot Advisory Committee, 2013).

Progress in Implementing Student Growth Objectives

During site visits, many districts reported significant progress in developing SGOs. SGOs were reported to be in development in all visited districts. Site visits were conducted from December 2012 through May 2013. Some districts reported continuing confusion around the requirements for SGOs; these were primarily districts visited earlier in the year. The NJ DOE released new SGO guidelines in April 2013. In at least two districts, teacher evaluation leaders reported at the time of the site visit that they were preparing for SGO development but were awaiting the release of these guidelines. Administrators in all districts visited after the release of the SGO guidelines reported finding the new guidelines useful and claimed to have made significant progress on

creating SGOs. In two districts, interviewees noted that they had had to change their SGOs based on the new guidelines but were making good progress on this adaptation.

Almost uniformly, districts reported using teacher-made tests for SGOs but the formats of these assessments varied substantially from multiple-choice tests to portfolios. In one district, administrators required teachers to shift from essay exams to multiple-choice tests when a previously-established system of midterms and final exams was repurposed to serve as SGOs. In other districts, teachers had more leeway in choosing the format as well as the content of the test. Three districts reported using assessments other than teacher-made tests. In one district, the central office developed tests and distributed them to teachers. Two districts used commercial tests with known psychometric properties to assess a combination of mathematics and language arts mastery in the early grades.

Teachers' concerns about designing SGOs varied among districts visited during site visits. In one district, SGOs would essentially be developed centrally, so the task did not burden teachers, and they were comfortable with having the central office do this. Two of the remaining nine districts where teachers would design the SGOs were comfortable with developing their own SGOs and reported being clear about what was required. In the seven other districts, a variety of issues was raised. In two, teachers said they were unclear about what the task was. This statement was different from two other districts where teachers said they didn't know enough—meaning that both the task was unclear and that they lacked adequate background—and they were waiting for training. In three districts, teachers worried about how to achieve comparability of SGOs across grades and subject areas, and in another, teachers were not clear about how to develop SGOs that would provide them with useful information on their teaching.

Roughly half of all teachers responding to the Year 2 survey reported that they had been personally involved in developing their districts' SGOs; 48% responded yes when asked. Table 6 shows the response rate for this question by district.

Table 6 – *Percent of Teachers Reporting Involvement with Developing Student Growth Objectives (SGOs) (n=1980)*

District	Percent Yes	Percent No
Cohort 1		
1	81%	19%
2	40%	60%
3	37%	63%
4	24%	76%
5	53%	48%
6	51%	49%
7	24%	76%
8	29%	71%
9	11%	89%
10	93%	8%
Cohort 2		
11	75%	25%
12	25%	76%
13	51%	50%
14	27%	73%
15	98%	2%
16	63%	37%
17	88%	12%
18	87%	14%
19	69%	31%
20	54%	46%
21	17%	83%
22	90%	10%

The percent of teachers reporting involvement in developing SGOs ranged from 11% to 98%, as shown in Table 6. More teachers reported involvement with SGO development in Cohort 2 than in Cohort 1. It is difficult to know why reported district percentages of involvement in SGO development varied so greatly. Site visit data indicate that in some districts, administrators encouraged groups of teachers—usually departments or grade levels—to develop common assessments. In one district, for instance, previously existing end-of-course tests were redesigned to become part of an SGO to which quarterly assessments were added so growth could be assessed over time. In another district, the central office played a greater role in designing assessments and required teachers to administer them. Examination across site visits suggests that in some cases, districts with more staff who knew how to design assessments and analyze the results may have been more ready to dictate procedures to teachers, but this variation of capacity for working with assessment data does not seem to have differed by cohort.

In any case, although surveys and site visits were able to track the process of SGO development, help us to better understand what decisions districts were making, and even to collect some artifacts of local assessments, the RU GSE research team was not able to collect data that would have facilitated determination of the quality of the actual SGOs developed.

Summary

Pilot districts continued to work on implementing the many parts of the teacher evaluation program. Key findings about perceptions of implementation that are important to consider as New Jersey enters full adoption in 2013–2014 are:

- All districts adopted teacher practice instruments and related data-management systems.

- In the survey, about half the project directors questioned whether they would be able to complete the high number of observations required for 2012–2013. While data collected by NJ DOE indicate that the 13 districts for which information is available averaged three observations per teacher, the required number of observations was reduced for full implementation in 2013–2014. This reduction will increase the likelihood that New Jersey school districts will be able to complete the required number of observations.

- Pilot districts reported making significant progress in developing SGOs for their schools. These would largely be teacher-made tests. Some teachers and administrators voiced questions about how to develop effective SGOs—sometimes about requirements and sometimes about technical design issues. Development was substantially bolstered by the new guidelines from NJ DOE, however.

Orientations

This section explores how teachers and administrators evaluate the teacher evaluation program from two perspectives. First, we ask how teachers and administrators assess the fairness and accuracy of the evaluations that are being conducted. It is worth exploring these issues because their judgments may influence how they respond to the new program and whether they see the resulting evaluations as a legitimate source of information about their practice.

Second, we examine how educators see two goals of teacher evaluation changing their relationship to their work. The summative goal is to improve management of New Jersey's educational workforce, or, as the New Jersey Educator Effectiveness Task Force says, "to vastly improve personnel decisions, such as the awarding of tenure and the setting of compensation levels, and drive significant improvements in student learning" (New Jersey Educator Effectiveness Task Force, 2011, p. 4). Although better decisions about who should teach could significantly increase the quality of the teacher work force, addressing this goal could be a source of stress for teachers. In the extreme, it might create a teacher evaluation equivalent of teaching to the test, which some have argued has become especially prevalent with the adoption of high states achievement testing required by No Child Left Behind (Brown & Clift, 2010; Hamilton, Stecher, & Yuan, 2008).

The formative goal is to provide learning opportunities for teachers as indicated in the notice of grant opportunity (NGO) to which the pilot districts responded. This says that teacher evaluation

is intended to provide "teachers with targeted professional development opportunities aligned to assessment and feedback to support their growth" (New Jersey Department of Education, 2011, p. 4). By helping teachers develop their skills and increase their capacity to match their instructional approaches to their students' needs, this goal could increase teachers' sense of efficacy and satisfaction with their work.

Accuracy and Fairness

The people who use the teacher evaluation tools must believe that these data provide an accurate picture of teachers' contributions to student learning and that they are fair. Generally, administrators were more convinced of the fairness and accuracy of the observation data generated than teachers. Table 7 presents survey data comparing how many teachers and administrators agreed with one statement that "the new teacher evaluation protocol can be used to generate accurate assessments" and another that "the teacher evaluation protocol for assessing teachers is fair." From 68% to 83% of administrator survey respondents agree with this statement as opposed to just over 30% of teacher survey respondents. Similarly, about 80% of administrators agreed with the statement that the teacher evaluation protocol is fair, as opposed to between 29% and 30% of teachers.

The overall sample results includes some rather substantial district-to-district variation. The percent of administrators agreeing with a position could range from about two fifths to all administrators in a district. District-to-district variation was not so large for teachers, perhaps because each district had so many more. Still, responses to all items ranged from 30 to 40 points across districts.

Table 7 – Median Percentage of Teachers and Administrators in Each District That Agree With Statements Regarding the Accuracy and Fairness of the Teacher Evaluation Protocol

	Administrators (n=275)		Teachers (n=2555)	
	Cohort 1	Cohort 2	Cohort 1	Cohort 2
The new teacher evaluation protocol can be used to generate accurate assessments.	83%	68%	31%	33%
District range:	60–100%	50–100%	13–46%	16–51%
The teacher evaluation protocol for assessing teachers is fair.	81%	79%	29%	30%
District range:	57–100%	38–100%	16–47%	12–53%

Note: Agree and Strongly Agree response options were combined.

Table 7 does not distinguish between those who disagree with the statements made in the survey and those who currently have no opinion. Figure 1 combines data from both cohorts to show the percent that agree, disagree, and neither agree nor disagree. It shows that in most districts, teachers are divided into three groups on both questions--those who agree, those who disagree, and those who have not yet made up their minds. However, in the typical district almost four fifths of administrators agree with both statements and most of the rest neither agree nor disagree.

Figure 1 – The median percentage of teachers and administrators in each district agreeing or disagreeing with statements regarding the accuracy and fairness of the teacher evaluation protocol.
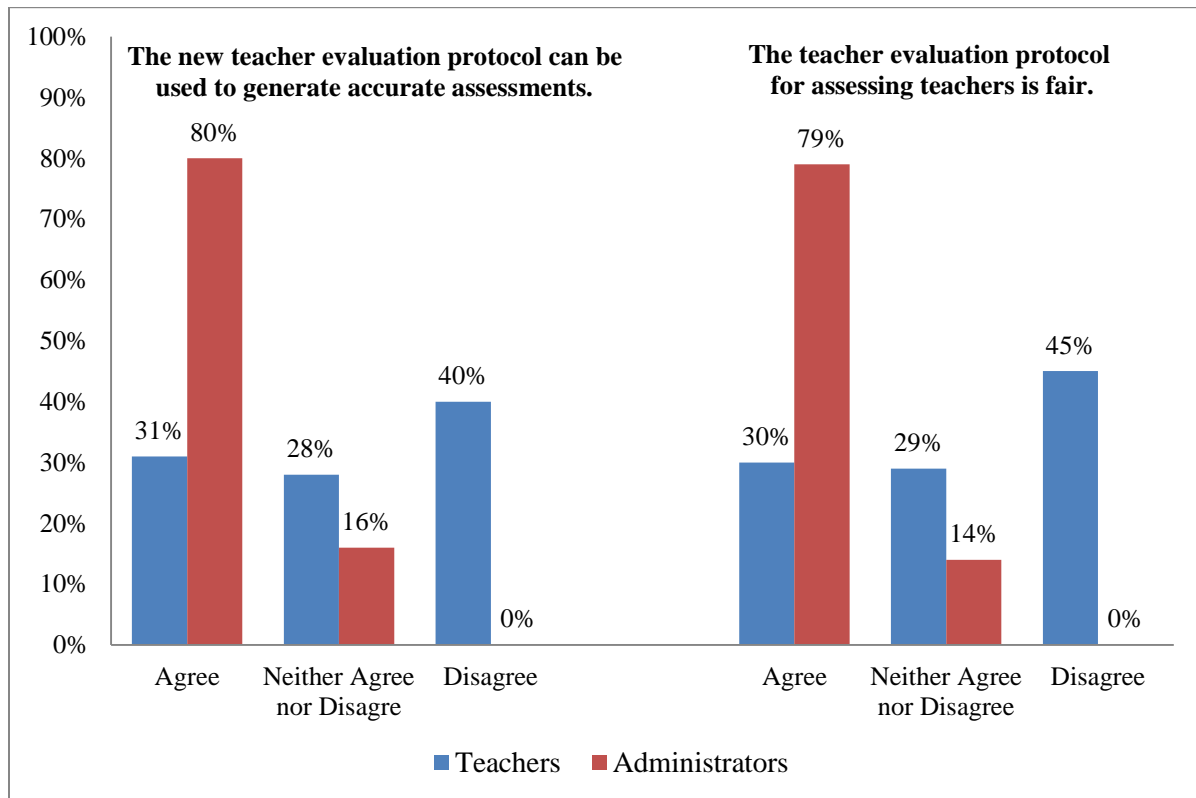
Several factors affected perceptions of accuracy and fairness. One issue recognized by the NJ DOE and those who design teacher evaluation systems is the need for consistency among raters (Kane & Staiger, 2012). During site visits, inconsistency was mentioned in teacher focus groups in four of the 10 districts visited in the second year. Teachers explicitly noted that different observers would rate the same teacher differently. While there might be a variety of reasons for these differences, teachers viewed these discrepancies as a sign of inaccuracy. The level of concern varied among districts, and in a fifth district, teachers actually commented that when two raters went together to do a double observation, their observations were quite consistent. These observations suggest that concerns about inconsistency among raters occurred in several districts, but not all of them.

Most districts took several steps to increase inter-rater consistency. The first step was the training provided to raters, which was intended to help them understand the rating categories in some detail (see data on training in Barriers and Facilitators, below). In districts that used the Danielson Framework for Teaching, training usually included a process to develop and then certify observers' accuracy. Observers' ratings of a videotaped lesson were compared to criterion scores; the observer would have to reach a set level of agreement with the criterion scores to be certified. Most districts that had access to this system used it and provided their observers with several videotaped lessons on which to practice beforehand. Moreover, in many districts,

administrators continued to meet to refine their rating accuracy after initial training was finished (see Training in Barriers and Facilitators below). In addition, at least one district apparently removed a principal whose observations were inadequate—although reasons for this person's dismissal were not given—and also hired an additional person to do observations who was especially expert in doing so. Consistency in ratings appears to be an issue that varies across districts. In four districts, teachers described inconsistencies across raters during focus groups. Whether these inconsistencies reflected differences between raters or the way the same teacher taught at different times is difficult to ascertain from these accounts, but it is clear that teachers voiced a concern. On the other hand, teachers in three districts noted that consistency across raters was reasonably high.

Another concern for teachers was the observer's distance from the classroom. Here the conventional psychometric wisdom is that more distant, unbiased researchers are to be preferred. A principal has several incentives to increase his or her teachers' ratings (Donaldson, 2013; Weisberg, Sexton, Mulhern, & Keeling, 2009). In contrast, teachers believe that some closeness to the classroom is important. In response to an item in the second year survey, 80% of teachers said that it was important that "the person observing me knows my classroom well."

The regulations governing teacher observation during the pilot were designed to vary the amount of distance among observers viewing a teacher's work. Each teacher was supposed to be evaluated both by internal evaluators who should know the classroom situation and external evaluators from outside the building where the teacher worked. The latter were almost always district employees whose knowledge of the particular classroom may have been less extensive and whose relationship to the teacher would have been less close. During focus group interviews, teachers noted instances where they thought their observers did not understand their classroom adequately. Two teachers in separate districts described how they had been criticized by an observer for not appropriately handling a discipline problem with a student. The teachers described how the students had special conditions that required the actions taken. In one case, the observer did not realize that the criticized teacher behavior was what was prescribed by that student's IEP for that student's recurring disruptive activity. While these situations are unusual and can be reduced in some instances through pre-conferences, they illustrate that the same distance that reduces observer incentives to give inflated ratings may also limit observer access to relevant contextual information. After the second year of the pilot program, regulations were changed so that external observations were no longer required.

Orientations toward Program Goals

We explored teachers' sentiments about the two stated goals (i.e., formative and summative purposes) for the teacher evaluation program through several survey questions. We first asked how important teachers thought both goals were. Table 8 shows that the same-size majorities of teachers agreed that both purposes were important.

One concern about the summative goal was that if new teacher evaluation requirements became more stringent, teachers might fear losing their jobs. To find out how prevalent this concern was in pilot districts we asked teachers how they thought the teacher evaluation process would affect their tenure. We asked tenured teachers how worried they were about losing tenure. Table 9 shows that almost three-quarters of them thought it unlikely that they would have a chance of

losing tenure under the new system. Moreover, on this issue, teachers subject to different processes for assessing student growth had very similar responses. When teachers of tested and untested subjects were compared, the percentage difference between them was only 3%. The district-to-district variation on these items was notable—ranging from 20% to 40%—but not as large as on some items described above.

Table 8 – *Teacher Responses to a Question About How Important They Thought Two Purposes of Teacher Evaluation Were in Their District (n = 2678-2681)*

	Agree	Neither Agree nor Disagree	Disagree
To provide information for teachers to improve practice	58%	22%	20%
To make tenure and promotion decisions	57%	32%	12%

Table 9 – *Tenured Teachers' Responses to a Question That Asked How Likely They Thought it Was for Them to Lose Tenure under the New Evaluation Protocol*

	Likely	Neither likely nor unlikely	Unlikely
Math and/or Language Arts Teachers Grades 4–8 (tested) (n=604)	8%	21%	71%
District range:	0–33%	0–41%	53–95%
Other teachers (non-tested) (n=1628)	7%	20%	74%
District range:	0–21%	9–30%	52–91%

We also asked untenured teachers how they thought the new teacher evaluation process would affect their chances of getting tenure (Table 10). Half of them (50%) thought that the new requirements would have no effect on their tenure chances, but more thought the new rules would help than hurt (26% to 7%).

Table 10 – *Untenured Teacher Responses to a Question That Asked How They Thought the New Evaluation System Would Affect Their Chances of Getting Tenure (n=438)*

	Percent
Improve Chances	26%
No Change	50%
Reduce Chances	7%

Note: Greatly Improve and Improve are combined as are Greatly Reduce and Reduce. Don't Know, not included.

We also explored whether the new teacher evaluation process provided teachers with information that they found helpful for improving their teaching. Table 11 shows teachers' views on the helpfulness of the observation feedback. About a third of teachers who participated in the survey agreed that the observation data helped them improve some aspect of their teaching, either their instructional methods or their classroom climate, with another third neither agreeing nor disagreeing. Teachers were asked similar questions about how the student growth data they received helped them improve their teaching, and slightly more teachers reported that these data were helpful. Neutral responses were almost as prevalent. Here again, district variation is notable. Although the variation is not as extreme as on some questions, a substantial majority of teachers in some districts reported that both observer feedback and student growth data were helpful while as few as a fifth to a quarter did in other districts. It is worthwhile to note that since

we did not ask these questions prior to the implementation of the teacher evaluation pilot, we do not know if these perceptions have improved, stayed the same, or decreased since implementation of the new teacher evaluation system. Thus, these should not be interpreted as reflecting the utility of the new evaluation systems as compared to the previous system, but as a standalone appraisal.

Table 11 – *Median Percentage of Teachers in Each District Agreeing with Statements Regarding the Usefulness of Teacher Evaluation Data (n=2601)*

	Agree	Neither Agree nor Disagree	Disagree
... the feedback I have received from **observers** has helped me to improve the learning environment in my classroom.	35%	33%	29%
District range:	19–57%	23–50%	10–45%
... the feedback I have received from **observers** has helped me to improve the quality of my instruction.	38%	32%	29%
District range:	21–58%	18–49%	10–49%
Student growth measures tied to teacher evaluation provide me with information that helps me to improve the learning environment in my classroom.	40%	30%	26%
District range:	23–61%	21–44%	14–45%
Student growth measures tied to teacher evaluation provide me with information that helps improve learning opportunities for my students.	44%	29%	24%
District range:	27–59%	22–43%	11–41%

Note: Strongly Agree and Strongly Disagree combined with Agree and Disagree, respectively.

We also explored administrators' uses of teacher practice data for improvement purposes. In particular, we wanted to know if administrators were going beyond individual coaching to use teacher practice data to plan school or district professional development or even to understand the patterns of strength and weaknesses of instruction in their schools. When discussing barriers and facilitators to the use of evaluation systems for improving teaching, we will present survey data indicating that the data-management systems used with each observation system were generally seen as easier to use for recording observation data and providing immediate feedback for teaching than for collecting data that could be used to plan for professional development. We also asked administrators if they participated in "any effort where several administrators in the district planned for professional development and used teacher observation data in some way."

Fifty percent of the administrators in Cohort 1 said they participated in an activity where they "shared district-wide data;" the comparable response for Cohort 2 was 14%. By contrast, 49% of Cohort 2 administrators said they didn't have enough data yet to share as opposed to 21% of Cohort 1 administrators. Superintendents might be the best informants on whether districts shared observation data when planning for professional development. Examination of this smaller sample confirms the findings from all administrators. Five of 10 Cohort 1 superintendents said they did share data for professional development planning while two of 12 Cohort 2 superintendents gave the same response. These data suggest that sharing data to plan professional development has not necessarily been an activity among pilot districts during the first year of implementation.

Summary

Substantial numbers of administrators report that the teacher evaluation system can generate fair, accurate data. In contrast, teachers divide into three groups: those who agree that teacher evaluation is fair and accurate, those who disagree, and another group who report neither agreeing nor disagreeing. Change over time in these ratings has been minimal, but two factors contributed to teachers' perceptions of observation fairness and accuracy: agreement among observers on how to rate a classroom, and observers who were well informed about the classroom.

Although teachers take diverse positions on the overall fairness and accuracy of teacher evaluation, most are not worried that this policy threatens their job security. Over two-thirds of tenured teachers thought it unlikely that they would lose their positions because of teacher evaluation. Moreover, more untenured teachers thought enhanced teacher evaluation would increase rather than decrease their chances of getting tenure. Teachers evaluated with SGPs were as confident of maintaining tenure as those in untested subjects. These perceptions are grounded in a teacher's clear understanding of the double purpose of teacher evaluation. That is, most teachers recognize that teacher evaluation is intended to guide personnel decisions and provide feedback for their own improvement.

Less than half the teachers reported finding any one source of information coming from teacher evaluation helpful in improving their teaching. Generally, slightly more teachers reported a neutral position about the helpfulness of teacher evaluation data than disagreed that the data were helpful. Their reservations about the helpfulness of observation data may reflect their concerns about its fairness and accuracy.

Barriers and Facilitators

Four factors stood out as barriers or facilitators to implementation of teacher evaluation in New Jersey. First, districts needed guidance and sometimes support from NJ DOE. Guidance for SGOs improved dramatically in April 2013. Second, the length and quality of training provided to administrators and teachers affected their capacity to play their parts in the program. Third, time for administrators to conduct observations continued to be an important issue. Finally, the data-management tools necessary for teacher practice data collection proved challenging to learn but especially helpful for the more routine aspects of collecting data and providing feedback.

State Guidance

A significant factor affecting progress in Year 2 was guidance from NJ DOE. In Year 2, districts especially sought guidance on the student growth component of teacher evaluation. NJ DOE brought out formal guidance documents throughout the 2012–2013 school year. For instance, in October 2012 it distributed the *2012–2013 Teacher Evaluation Pilot Toolkit.* This document

described required policies as they were developed at the time, but also recommended best practices. Its best practices for measuring student growth in untested areas included:

- Be SMART (specific, measurable, attainable, relevant, and timely);
- Be based on available student learning data;
- Align to state standards, as well as to any school and district priorities;
- Represent the most important learning during an interval of instruction (New Jersey Department of Education, 2012, p. 42).

More detailed guidance on developing measures in untested areas was offered in another document in April 2013. This guidance reflected what NJ DOE had learned from its interaction with pilot districts and other advice it sought out. In between, NJ DOE staff were in frequent contact with pilot districts to provide clarification about requirements and learn about district practice. Because NJ DOE staff were learning from district experience (and other sources of information), the state changed some guidelines as the year progressed. This was most apparent in terminology. What had been Student Achievement Goals in October became Student Growth Objectives (SGOs) by April. This process generated diverse district responses. In one district, the data coordinator welcomed the opportunity to provide feedback and tolerated what he called "waves of refinement" but wished that certain issues that affected his work were pinned down. In other districts, people reported less tolerance of this flux. Still, either data coordinators or project directors noted that guidance received from the state made aspects of their work clearer. Because districts still had questions, they would either attend regional conferences on SGOs sponsored by NJ DOE staff or ask NJ DOE staff to come to their districts to brief them and answer questions.

Table 12 shows that when they encountered challenges, administrators in the pilot districts were more likely to turn to the State than any other source of support. Thirty-one percent of administrators named state guidance as their "go-to" source for information when facing difficulties. This was more than any other source, being slightly more than peers in other pilot districts who were presumably encountering similar challenges. Even some of the peer interaction was facilitated by NJ DOE through monthly EPAC meetings that brought together staff of all levels from the pilot districts in contexts that helped educators to learn from each other.

Table 12 – *Administrator Responses to a Question Asking Who They Are Most Likely to Turn to for Support When Facing Challenges in Implementation (n=297)*

	Percent of Respondents	Number of Respondents
Peer districts in pilot	27%	56
State	31%	64
Expert consultants	17%	35
Professional associations: NJEA, NJPSA, NJASA, or other	6%	12
Other	19%	40

Interviewees in many districts noted state guidance as both a barrier and facilitator to effective implementation. NJ DOE's role in developing SGOs was most commonly noted. Prior to April 2013, a common source of frustration during site visits was lack of clarification about SGO

requirements. Administrators and teachers alike reported that they were unsure whether the SGOs they had created conformed to state expectations. Although all districts worked on developing SGOs, one district put its work on hold until further guidance was received.

After NJ DOE released its April SGO guidelines, districts found state expectations easier to understand. Districts used the new guidelines to rapidly develop their SGOs. In several districts, administrators responded to questions about how to develop SGOs by simply stating that doing so was easily accomplished by following state guidelines. One district revisited previously developed student assessments that would be used as SGOs and redesigned them to conform to the new regulations. While some districts, especially teachers within those districts, still expressed some confusion around SGOs, the state guidance on this issue seems to have clarified requirements for districts and facilitated local SGO development.

Training

Teachers and administrators in Year 2 once again reported a wide discrepancy in the amount of training on how to conduct and use teacher observation instruments as well as how to interpret results. Generally, teachers continued to receive significantly less training than administrators. Table 13 shows administrators' and teachers' levels of training for Year 2.

Table 13 – *Self-reported Hours of Training on the New Teacher Evaluation System by Administrators and Teachers*

Hours of Training	C1 Y1 Admins (n=120)	C1 Y1 Teachers (n=2391)	C2 Y2 Admins (n=129)	C2 Y2 Teachers (n=2756)[3]
Less than 25	63%	97%	42%	94%
25–40	27%	2%	31%	5%
Greater than 40	10%	1%	27%	2%

Note: C1 and C2 represent Cohort 1 and Cohort 2 respectively.

Very few teachers (7% of those in Cohort 2), reported receiving more than 25 hours of training on the new system. Fifty-eight percent of administrators had 25 hours or more of training, with just over a quarter of respondents receiving over 40 hours. This trend mirrors the findings of the Year 1 assessment, though slightly more teachers—7% vs. 3%—of Cohort 2 teachers received larger amounts of training. Administrators received more training because they actually conducted the observations. This gap in the levels of training for teachers and administrators may explain some of the difference in perceptions about the training, detailed below.

In Cohort 2 districts, administrators received more training on the system; over half of all administrators in Cohort 2 received greater than 25 hours of training compared to just over one-third of administrators in Cohort 1 in Year 1. Extended training of over 40 hours was reported by one-quarter of administrators in Cohort 2, as compared to only 10% of administrators in Cohort 1. While several factors may account for these differences, during site visits administrators from Cohort 2 suggested that one factor might be because of a decision to start training earlier because they were informed that they would be in the pilot earlier than their Cohort 1 peers.

---

[3] Over 100% due to rounding.

We also asked administrators about their satisfaction with the training experience on several fronts. Table 14 shows the results of these questions comparing administrators from Cohort 1 Year 1 to Cohort 2 Year 2. Overall, administrators continued to be positive about their training experience; most administrators reported that training helped them understand many facets of teacher evaluation. Administrators were especially positive about the training providing understanding of the practice instruments and the ability to assess teachers' instructional practice. Four-fifths of Cohort 1 administrators felt that training had accomplished these goals; roughly three-quarters of Cohort 2 administrators felt the same. Cohort 2 administrators were generally positive about the training experience but slightly less so than their peers in Cohort 1. The factor that administrators were least confident about was whether their training facilitated their ability to assess teachers' planning practices. Fifty-seven percent of Cohort 1 and 46% of Cohort 2 administrators felt training had achieved this goal. This may be because training on the teacher observation protocols usually focused on understanding and rating factors during the observed lesson rather than elements of lesson planning.

Table 14 – *Administrator Perceptions of How Well Their Training on the New Observation Tools Accomplished the Following Tasks*

	Percent Agree Y1 Cohort 1 (n=120)	Percent Agree Y2 Cohort 2 (n=128)
Help you understand your district's rubric for assessing teachers	80%	77%
Help you to assess teachers' planning practices	57%	46%
Help you to assess teachers' instructional practices	80%	73%
Help you to assess teachers' capacity for feedback and self-correction	N/A	54%
Help you provide effective feedback to teachers after observation	72%	63%
Help you reach thorough, well-grounded judgments of teacher quality	70%	64%
Help you be aware of potential biases in the way you evaluate teachers	62%	68%

During site visits, administrators recounted several training techniques that they found to be particularly helpful. Most often mentioned was the opportunity to practice ratings using the new protocol. This was often done via video lessons provided by the training organization; administrators found this practice invaluable. Other administrators worked with their teachers to score a lesson on an informal basis; scores were discussed with the teacher and other administrators to help to crystallize knowledge around how the various levels of the teacher practice assessment criteria were to be applied.

A second technique that was frequently mentioned in interviews was the chance to talk with other observers about scores. This may have occurred after scoring a video lesson or simply in the training sessions in general. In one Cohort 2 district, an administrator independently formed a "study group" of peers to work through and discuss video training elements in preparation for observer certification. This administrator believed that she would not have been successful in the training without this interaction with peers.

Teachers were not asked in the Year 2 survey about their overall satisfaction with training, but instead we looked at whether or not teachers felt they spent too much, or too little, time on

specific topics in specific training formats. Responses did not so much tell whether teachers thought they received too much or too little training as what approaches they found most useful. In general, teachers, like administrators, appreciated time to discuss issues with their peers. This sentiment was supported by teachers in the focus groups, where teachers frequently mentioned that they found both turn-key training—training by peers who had received more in-depth training, often along with administrators—and time to share thoughts and understandings with peers to be valuable. Teachers were generally less positive about the merits of video lectures, especially when the video sessions were assigned to be completed independently rather than as a group.

Teacher focus groups were especially vocal about the methods in which information was transmitted. Teachers recognized that a certain amount of delivery of information was necessary, and they appreciated getting it from the most informed source possible. However, they objected when the primary vehicle for delivering this information to them did not coincide with what the teacher practice rubrics described as best instructional practices they were expected to use. In several districts, teachers noted, for example, that during training on the Danielson Framework, it was ironic that they were trained using primarily lecture sessions, a practice that was given low scores in the Danielson rubric itself. They preferred less lecturing and more access to other modes of learning. For instance, in the focus groups, some teachers reported appreciating the opportunity to score lessons collectively, an activity that was more commonly part of administrator training. This was a use of video during training that teachers appreciated although it was not "video lectures."

Training for Cohort 1 in Year 2 was also an area of interest. Table 15 shows project directors' responses to what types of training were delivered in their districts for Year 2.

Table 15 - *Training Types Delivered in Cohort 1 Districts According to Project Directors (n=9)*

Type of training	Number of project directors saying "yes"
Conducting training sessions for new teachers	8
Conducting training for observers for calibration or inter-rater reliability	8
Conducting training sessions for new observers	6
Conducting training for observers on the data-management system	4
Retraining teachers or observers where necessary	4
Conducting additional training for teachers already trained	3

Most Cohort 1 districts continued to train new teachers and observers. Training methods for new teachers varied among districts; in one district site visit, a new teacher noted that the training for the system she received was included in an optional mentoring program for new teachers. Calibration for observers was another area in which many districts (eight of 10) were active. This may be due to the requirement in Year 2 to conduct double-scored observations, which provides a natural opportunity for rater calibration.

Some Cohort 1 districts moved beyond training on the mechanics of the teacher practice instruments and into focusing on how to improve instruction in criteria measured by those instruments. Though only three of nine project directors reported that their districts were participating in additional training for teachers already trained, districts may be using the evaluation system to provide training for teachers in elements linked to the evaluation program

beyond the mechanics of the system. In two of the four Cohort 1 districts visited, interviewees noted that aspects of the observation protocols—such as questioning strategies and student engagement—had been selected as district-wide focal points for the year. This may indicate a positive impact of the program.

Time

As with other teacher evaluation programs (Milanowski & Kimball, 2003; White, Cowhy, Stevens, & Sporte, 2012), administrators' time has been an important barrier to implementation in some pilot districts. Here we present survey evidence suggesting that the time constraints noted in the first-year report continued into the second year, although their magnitude declined. We also identify some conditions under which this problem was more or less extreme and consequences noted when administrators lacked sufficient time to conduct evaluations.

Administrators have always spent time observing teachers; the question is whether they had to spend more as a result of the new requirements. Table 16 shows the percent of administrators who reported spending more time conducting tasks related to teacher evaluation across cohorts and years. Ninety percent of Cohort 1 administrators said they spent more time doing teacher observations than prior to the pilot, and this percentage only increased in the second year. It was essentially the same for Cohort 2 administrators. The percent of Cohort 1 administrators who reported spending more time writing up observations in the first year of the pilot was also about 90%, but it declined slightly in the second year. Here again, the percent of Cohort 2 administrators who spent more time writing up observations than before was about the same as that for Cohort 1 administrators in their second year. Table 16 also shows that these program-wide averages encompass a fair amount of district-to-district variation. These differences are especially large for Cohort 2. Finally, from 40% to 50% of administrators reported spending more time on other tasks.

Table 16 – *Percent of Administrators Who Report Spending More Time on Various Responsibilities since the Implementation of the New Teacher Evaluation System*

	Year 1	Year 2	
	Cohort 1 (n=141)	Cohort 1 (n=155)	Cohort 2 (n=127)
Doing observations	89%	96%	91%
District range:	83%–100%	80%–100%	33%–100%
Writing up observations	91%	86%	87%
District range:	73%–100%	70%–100%	50%–100%
Other administrative tasks or job responsibilities	46%	50%	41%
District range:	25%–86%	29%–100%	23%–67%

In Year 2 interviews, principals and district staff reported some of the same challenges created by the added time for observations that had been reported in Year 1. One challenge was that their regular work would suffer (or, at best, change). Repeatedly, principals noted that they had to be "visible" within their buildings and that they could not maintain what they thought was the necessary level of visibility if they had to spend too much time both doing the observations and,

even more, writing the documentation for observations. They also said that their regular work suffered, most notably communication with parents and sometimes with students or budget management (among other tasks). Teachers also noted that principals provided less support than they had in the past, particularly with disciplinary issues. Teachers tended to recognize the additional workload that teacher evaluation created for the principals.

Second, the observations could be affected. Even where all observations might be completed, the quality of feedback provided to teachers would suffer. Teachers critiqued and administrators acknowledged that written feedback was sometimes delivered after a great delay and in a form that either lacked detail or otherwise was not helpful. Similarly, post-conferences might be put off to the point that neither the teacher nor the observer could remember well enough what had happened to make a session a useful learning experience.

Finally, the additional work impacted other aspects of their lives. Observers—especially principals—reported reduced personal time, including evenings, weekends, and "family time." While some individuals were able to manage the work without losing too much personal time, a few were finding the strain exceedingly difficult.

The magnitude of these issues appeared to decline in the second year, however. A review of the site visit reports from the Cohort 1 districts suggests that the extent of the time-management problem declined in three of the four districts. However, administrators still described time management as an issue in two of them. While adding time for observations was a new challenge in all six Cohort 2 districts visited, site visit reports suggest that while it added stress, the observation work was actually completed in four of them. Only in one was there consensus that administrators did not have time to complete their regular work and required observations.

Some factors appeared to reduce the problem—or at least the level of anxiety about it. For one thing, unlike the experience of the Cohort 1 districts, the NJ DOE made its awards to the second cohort of districts early enough in the summer of 2012 that those districts had time to prepare for implementation before the school year began. Meanwhile, administrators in Cohort 1 had a year to learn how to do observations. They rarely struggled with the software to the same extent as they had the year before, and they understood the basic procedures and reporting categories better. Some explicitly mentioned that "the learning curve" was past. In addition, a few Cohort 1 districts adjusted their staff to take the new observation requirements into account. For instance, one principal who was reported to be particularly inept at teacher evaluation was not in the same district for the second year of implementation.

Some factors continued to be challenging. For instance, observers agreed that because writing up each observation took an extensive amount of time, reducing the required number of long observations made less of an impact than did changing the total number of required observations. Thus, the increase in required observations from Year 1 to Year 2 kept the amount of observation time required high even though some observations could be shorter in Year 2.

The variation among districts in the extent to which they felt the effects of extended observation time appeared to have been influenced by two factors. The first was available staff. Some districts seemed to have relatively few administrators able to do observations and little prospect

of reallocating time so existing staff appropriate to do observations could take on more of the burden. These districts reported that they were still suffering from recent retrenchments that required them to lay off supervisors. Other districts were able to reallocate staff or find the funds to hire new people so that more administrative time was available for conducting and writing up observations. The second was scheduling. In a few districts, administrators and even occasionally teachers noted that the central office developed a very effective schedule to make sure that all observations were done and usually done at appropriate times. Moreover, these schedules were usually monitored by the superintendent or the pilot project director to make sure that observers stayed on track to finish on time. While no one actually said during an interview that the district's schedule was "bad," some arrangements appeared unusually time-consuming. The most notable was having principals do evaluations for other schools; this required them to be out of their buildings for extended periods of time. Whether this practice was somehow required by the shortage of available observers or stemmed from poor scheduling is difficult to ascertain. Failure to develop a schedule, even just to set interim benchmarks for observations, combined with failure to carefully monitor that schedule, also contributed to some districts' problems in completing all required observations.

Data-Management Tool

The first year report on the pilot teacher evaluation program pointed out that the recording, storing, feedback, and analysis of teacher observation data had become so complex that it required computerized support to be done efficiently and accurately (Firestone et al., 2013). As one district administrator said in the first year, "You can't do this with a paper and pencil." However, computerized data-management tools create challenges as well as opportunities. Table 17 shows how many administrators had several types of problems with their systems. Only about a quarter of respondents reported rarely encountering any significant technical problems (item k). However, anywhere from 5% to almost half the respondents might encounter problems, with items like short log-in periods, problems saving information, and crashing the tool being mentioned frequently and by administrators in both cohorts and years. The substantial variation among districts in percent of administrators reporting problems of every sort suggests that either the way districts trained administrators to use the data system or the internal support they provided made a substantial difference in administrators' perceptions of the data-management tools' ease of use.

31

Table 17 – *Most Frequently Encountered Technical Problems with the Data-Management Tool as Reported by Administrators (n=283)*

		Year 1	Year 2	
		Cohort 1	Cohort 1	Cohort 2
a. Short log-in period (getting kicked out and losing data before saving)	Count	38	33	41
	Percent*	26%	21%	32%
	District Range:	0–63%	0–60%	0–67%
b. Problems saving information while working online	Count	36	35	40
	Percent	24%	23%	31%
	District Range:	11–44%	0–50%	0–67%
c. Problems with crashing of the tool	Count	28	24	43
	Percent	19%	16%	33%
	District Range:	0–69%	0–61%	0–67%
d. Personal account information issues – lengthy password changes, updates of identification information, etc.	Count	13	6	11
	Percent	9%	4%	9%
	District Range:	0–24%	0–25%	0–21%
e. Poor accessibility of previously uploaded information	Count	17	9	22
	Percent	12%	6%	17%
	District Range:	0–50%	0–28%	0–46%
f. Difficulty generating reports (about groups of teachers)	Count	N/A	7	32
	Percent	N/A	5%	25%
	District Range:	N/A	0–22%	0–62%
g. Difficulty generating reports (about individual teachers)	Count	N/A	24	28
	Percent	N/A	16%	22%
	District Range:	N/A	0–83%	0–62%
h. Difficulty in aggregating data across observers	Count	N/A	0	41
	Percent	N/A	0%	32%
	District Range:	N/A	0%	0–62%
i. Ineffective online help feature/technical support	Count	15	70	25
	Percent	10%	46%	19%
	District Range:	0–43%	11–86%	0–46%
k. I have rarely encountered any significant technical problems	Count	65	20	30
	Percent	44%	13%	23%
	District Range:	0–67%	0–40%	0–100%
Total	**Count (N)**	**147**	**154**	**129**

*Represents percent of respondents within cohort who reported that this was one of the most frequently encountered issues.

Table 18 reports what percentage of administrators found the data-management tools easy to use for different purposes. In general, few administrators reported that these tools were easy to use for any of the listed purposes. Rarely did more than a third of respondents report that the tool was easy to use for a function. The data-management tools turned out to be easiest to use for the functions that are most central to teacher evaluation; recording data about teacher practice and sharing it with the teacher observed. Things related to identifying larger patterns in observation data and preparing reports to share with broader audiences than individual teachers were more challenging. From less than a fifth to about a quarter of administrators found these tasks easy.[4]

---

[4] These findings are only for Cohort 2. The findings for Year 1 Cohort 1 are similar, but respondents were given a different set of response choices so items cannot be combined.

Table 18 – *Perceptions of Administrators Regarding the Ease of Use of Different Features of the Online Data-Management Tool (n=129)*

	Finds Easy
Use of the tool during actual teacher observations	34%
Use of the tool for communication and exchange of information	31%
Access data to provide formative feedback or to conduct post-conferences for observed teachers	36%
Access already stored data to identify patterns of practice in the school or district	17%
Access already stored data to prepare reports on teacher evaluation data for various audiences other than the individual teacher being observed	14%
Management of already uploaded information	24%

A data-management tool can be difficult to use for several reasons. For instance, site visit data suggest that developers of some data-management tools may have gone so far in protecting the confidentiality of data about individual teachers' practice that aggregating data across schools or observers could be a challenge. Districts with more sophisticated knowledge of data analysis sometimes supplemented the externally-provided data-management tools with other programs or analysis tools in order to answer their questions. Figuring out how to do analyses outside the data-management tool often took time that was especially difficult to find in the first year of implementation.

In addition, some problems may have come from lack of experience or practice. Table 19 shows how often all administrators reported using the data-management tool for various tasks. It shows that many more administrators used the tool to provide individual teacher feedback than for more analytic purposes, such as checking for inter-rater reliability or assessing the patterns of strengths and weaknesses of teachers for various purposes. While the differences between cohorts in using the data-management tool to provide feedback to teachers are modest, Cohort 1 administrators spent slightly more time examining data to find patterns and distributions, a sign that more analytical uses come with experience.

Table 19 – *Perceptions of Administrators Regarding the Frequency of Use of Various Features of the Data-Management Tool (n=265)*

	Cohort 1 (Percent reporting)			Cohort 2 (Percent reporting)		
	Weekly	Monthly	Less than Monthly	Weekly	Monthly	Less than Monthly
Provide individual teacher feedback	53%	22%	25%	53%	13%	33%
Examine observer agreement	17%	27%	55%	18%	25%	57%
Look at score distribution to find school or district issues	12%	24%	64%	7%	14%	79%
Look at score distribution to assess collective teacher capacities	10%	18%	72%	6%	15%	79%

Although the differences between cohorts are relatively small, there are some substantial differences depending on one's administrative position. Table 20 shows that supervisors provided most of the feedback to teachers. Almost four-fifths of supervisors reported providing feedback weekly. Feedback to teachers was provided by half of both principals and superintendents weekly. More project directors and superintendents than other roles examined inter-rater agreement often. Three-fifths of both these roles did so monthly or more often, as opposed to two-fifths of those in other positions. More project directors looked for patterns in the observations frequently. Almost half did so at least monthly, while the fewest supervisors (less than 20%) did so often.

Table 20 – *Percent of Administrators Using Data-Management Tools for Different Purposes, Organized By Position*

	Weekly	Monthly	Sum of Both
Superintendents and Assistant Superintendents (n=24)			
Provide feedback	50%	17%	67%
Look at agreement among scores	5%	50%	55%
Use scores to examine school or district	4%	29%	33%
Use scores to examine teachers overall	4%	33%	37%
Principals and Assistant Principals (n=149)			
Provide feedback	50%	22%	72%
Look at agreement among scores	19%	23%	42%
Use scores to examine school or district	10%	22%	32%
Use scores to examine teachers overall	8%	14%	22%
Supervisors (n=52)			
Provide feedback	71%	8%	79%
Look at agreement among scores	22%	27%	49%
Use scores to examine school or district	9%	11%	20%
Use scores to examine teachers overall	9%	11%	20%
District Directors (n=23)			
Provide feedback	30%	25%	55%
Look at agreement among scores	20%	25%	45%
Use scores to examine school or district	15%	25%	40%
Use scores to examine teachers overall	10%	30%	40%
Project Directors (n=23)			
Provide feedback	57%	9%	66%
Look at agreement among scores	20%	40%	60%
Use scores to examine school or district	17%	26%	43%
Use scores to examine teachers overall	13%	35%	48%

In sum, learning to use the data-management tools that supported the teacher practice instruments was a significant endeavor. The most frequent uses of these tools (collecting practice data and sharing it with teachers) were also the ones that most administrators soon saw as easiest. Moreover, the variation among districts in reports of the ease of use of these tools suggests that district capacity could facilitate their implementation.

Summary

This analysis has helped not only to clarify what some key barriers and facilitators were, but in some cases, showed factors that increased the likelihood they would facilitate implementation. Key findings are listed below:

- Where NJ DOE can provide clear, well-developed guidelines for districts, those guidelines are appreciated and facilitate local implementation. When such guidelines

were developed for SGOs, the implementation process was accelerated and the pilot districts responded.

- Training continues to be an important startup issue. Effective training is partly a matter of time—and observers were more likely to report getting adequate training time than teachers—and partly a matter of delivery mode. What makes a delivery mode effective may have something to do with the medium. We found significant differences in trainees' responses to live vs. video training, although administrators seemed more comfortable with video training than did teachers. Especially for teachers, however, the real issue may have been the extent to which training modeled the precepts of good teaching. Using video (or live trainers) to simply deliver information in a static format was not appreciated as much as the sort of active learning opportunities recommended by the teacher practice instruments themselves. These included opportunities to score classroom lessons—a constructive use of video—and chances for peers to share information about and experience with scoring specific teacher practice rating categories.

- Time for observations continued to be a concern in Year 2 although the magnitude of the issue declined from the first year. Site visits suggest that observation time still created challenges for administrators to do some things they saw as important parts of their jobs, like working with parents and being visible in the buildings. If observing took no more time in Year 2 than in Year 1, it may be that "the learning curve" helped Cohort 1 administrators conduct observations more quickly, and by doing so, helped overcome the increase in required numbers of observations during Year 2. Observers report that writing up observations and communicating with teachers takes more time than the actual observations, so the number of observations required may be more important than the observation time required. It may be that the time issue will be alleviated going forward because the regulations adopted in September require fewer observations than were required in Year 2 of the pilot.

- Like so many new technologies, the data-management tools that support teacher practice instruments generate notable frustration as users are learning to use them. Administrators report that the data-management tools are easiest to use for their major functions, recording teacher observations and providing feedback to teachers. They are harder to use for data aggregation and analysis. As might be expected, administrators surveyed got less practice in using the tools for data analysis. The interviews suggest that data analysis was not an activity that many observers were comfortable with. However, some of the administrators interviewed who had greater capacity in this area suggested that the data-management tools needed to be supplemented to be effective.

Conclusion

This section brings together our findings across the assessment's three sections. It first brings together conclusions about implementation, perceptions of the program, and barriers and facilitators, then offers some suggestions for future research.

Implementation

Startup during 2012–2013 went more smoothly than the year before. NJ DOE provided awardee districts with notification that they won pilot grants earlier in the summer. This notification allowed Cohort 2 districts to plan for training sooner and begin observations earlier in the year than the Cohort 1 districts did the year before. Many already had ideas about the teacher practice instrument they would select, and most of them chose to use the Danielson Framework for Teaching.

It is difficult to make firm estimates about the number of observations completed during the first two years. About half the project directors reported doubts about their capacity to complete all observations in the spring of the second year, which may reflect the increased demands on observers' time and continuing learning challenges more than the actual feasibility of completing the work. In the half of the districts where observation data are available, the evidence suggests that the requisite number of observations were completed.

NJ DOE improved the guidance it offered districts throughout the 2012–2013 academic year. The clearer guidance was used and appreciated as it became available, although it did not appear to help teachers with questions like how to design assessments that were comparable across grades and subjects. With state guidance, districts were able to actually adopt functional SGOs during the year, a step that had not been possible the year before. These SGOs relied heavily on teacher-made tests. Whether they would provide useful formative data or meet appropriate tests for accuracy remained to be seen.

Orientations

As during the first year, data from the second year suggest that more administrators than teachers saw the teacher observation process as both accurate and fair. This difference may reflect a number of factors, including the key role difference: administrators were observers while teachers were evaluated through this process. Moreover, administrators received substantially more training about how teacher observation worked.

This report highlights two issues that contribute to teachers' review of the fairness and accuracy of teacher evaluations. One is consistency among raters; does the teacher believe that different raters will give the same lesson the same score? The NJ DOE and the providers of the observation protocols have given considerable attention to increasing agreement among raters through providing training materials and encouraging districts to use teacher observation instruments embedded in systems that require dual observations (now called co-observations), certification, and the like. The site visit evidence suggests that teachers in some districts are confident that raters are consistent, but that in other districts teachers are not as confident. This variation in the site visit reports fits with the notable variation among districts in percent of teachers agreeing that observations can be fair and accurate.

Teachers also thought it was important that their observers understood the context of their classroom. In interviews, they gave examples of observers unfamiliar with their classroom who gave inappropriately low ratings because they did not understand the reasons for some actions

observed. This concern may be one reason why the requirement that all teachers be observed by out-of-building as well as in-building observers was removed as teacher evaluation moved from a pilot program to a statewide regulation.

Most teachers recognized and accepted the fundamental purposes of the teacher evaluation system. The majority recognized that this system was intended to both guide personnel decisions and provide information to help teachers improve. It is notable that most teachers did not see teacher evaluation as a threat to their tenure. Almost three-fourths of tenured teachers thought it unlikely that the teacher evaluation process would affect their continued employment, and more untenured teachers thought it would help them get tenure than harm that effort.

Still, during this pilot phase when educators were still improving the observation process, access to SGPs based on state test scores was limited, and procedures for developing student growth objectives (SGOs) based on locally-collected data were being developed, less than half the teachers surveyed found the evaluation process helpful for their teaching. It may be that evaluation data will become more important as initial implementation challenges are resolved and teachers get more experience with using the data provided by the new approach.

Barriers and Facilitators

Administrators found the state guidance provided on SGOs to be an important facilitator of progress on the teacher evaluation program during the 2012–2013 academic year. The lack of clear specifications had been a major impediment. The process of developing that guidance generated mixed reviews, with some leaders of pilot districts complaining of inconsistency and others welcoming the opportunity for input. As specific policies were clarified, districts took advantage of that clarity to more energetically develop local SGO policies. These local policies varied somewhat. It would take more time to determine which ones were most effective as tools to measure student growth for summative purposes or as means to provide effective feedback to teachers.

Training also affected implementation of the teacher evaluation program. Not only did administrators receive substantially more training than teachers—not a surprise since administrators had to act as observers, provide accurate scores, and offer useful feedback—but they seemed more satisfied with the training they received.

Whether for teachers or administrators, two factors contributed to effectiveness of training. First, those receiving the training had to get enough. Cutting training short could reduce accuracy and effectiveness among observers and raise doubts and leave unanswered questions among teachers. However, initial training could rarely provide answers to all questions. It was important to provide time for people to meet, discuss their experiences, and learn from them as implementation moved forward. Observers benefitted from the opportunity to continue refining their understanding of observation categories into the second year. Where teachers received sufficient clarification about the teacher practice instrument in the first year, training could profitably shift to learning how to improve assessed practices by the second year.

Second, after receiving initial information, all educators appeared to prefer opportunities for active learning—opportunities to score lessons and compare results, for instance—to more

passive forms when they only received lectures or other forms of information. The preference for active learning helps to explain teachers' concerns about receiving video training. Much of what they received appeared to provide information with little opportunity for active learning. When the videos they saw contributed to opportunities to share and problem-solve, those videos were appreciated more.

The data-management system that supported the teacher practice instrument was also an important facilitator of evaluation. Nothing learned in the second year contradicted our earlier conclusion that these systems facilitated the implementation of teacher practice assessment. However, like other computer or online programs, the data-management systems created a steep learning challenge in the first year. Learning how to use these systems for recording observation data and providing feedback to observed teachers usually came first. Learning to use data-management systems to aggregate, communicate, and analyze data usually took longer. These skills were rarely learned as well. Moreover, some district experts in data analysis believed that the data-management systems provided by the designers of the teacher practice instruments needed to be supplemented to meet districts' analytic needs.

Sufficient time for implementation was also an important facilitator of effective teacher evaluation program implementation. Although it was a declining problem in year 2, insufficient time could affect teacher observations by the required number of observations not being met, by less feedback being provided, and by reducing the time principals had for other school matters. However, the time problem was not universal. In fact, districts varied considerably in their capacity to absorb these new demands. This project has identified three steps that helped alleviate the time problem.[5] First, some districts had more effective schedules than others. For instance, they managed to avoid requiring principals to leave their buildings to do observations. Part of effective scheduling was monitoring the schedule's implementation to ensure that schedules did not slip. Second, districts had to provide enough staff to complete all required observations. Staffing appeared to be a more acute problem in a few districts than in others. Finally, requirements for the number of observations to be completed were reduced for the 2013–2014 academic year and going forward.

Future Evaluation Issues

Beginning in the fall of 2013, new teacher evaluation procedures will be mandated statewide. The districts using the new procedures will no longer be the early adopters who are often the most sympathetic to a new policy and have the most cosmopolitan contacts and capacity to learn new practices (Hall & Hord, 2006). It may be especially important to track several issues about teacher evaluation on this broader canvas.

One issue concerns the quality of measurement that stems from the new teacher evaluation practices. This issue may be most apparent with SGOs. Districts are organizing in a variety of ways to meet this requirement. Some districts provide more central direction than others. Some depend largely on locally-developed instruments while others use commercial student

---

[5] The NJ DOE offers some partially overlapping suggestions in its August 6, 2013 Educator Evaluation Update for School Leaders, http://www.state.nj.us/education/AchieveNJ/resources/080613AchieveNJSchoolLeadersMemo.pdf.

assessments that permit short-term assessments of student growth. Much has yet to be learned about how teachers design assessments that reliably assess student growth while providing data that can be used in the short run to adjust instruction.

Another concern has to do with the completeness and accuracy of the teacher practice data resulting from administrator observations. This survey-based assessment of their implementation has left questions that can better be answered from a review of the aggregated observation data itself. These revolve around two major issues. First, how complete are the data being accumulated? Are all the observations required by the state being collected for each teacher? Second, how accurate are those data? It will be hard to address the accuracy question until fairly complete data are available.

Third, it will be important to understand how school and district leaders organize to implement teacher evaluation. The survey data collected point to considerable variation among districts in a variety of factors from teachers' perceptions of the fairness, accuracy, and usefulness of the data collected, to such organizational resources as the capacity to use the technology required for the teacher practice instruments and the time to conduct observations. It appears that these differences are not random, but that some districts may be more able to integrate the requirements of enhanced teacher evaluation into their ongoing work than others. It will be important to verify that hypothesis and explore how much that variation reflects such factors as access to financial resources, community support, and the quality of school and district leadership.

References

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

Brown, A. B., & Clift, J. W. (2010). The unequal effect of adequate yearly progress: Evidence from school visits. *American Educational Research Journal, 47*(4), 774–798.

Danielson, C. (2011). *The framework for teaching evaluation instrument.* Princeton, NJ: The Danielson Group.

Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly,49*(5), 838–882. doi:10.1177/0013161X13485961

Evaluation Pilot Advisory Committee. (2013). *Final report.* Trenton, NJ: New Jersey Department of Education.

Firestone, W. A., Blitz, C. L., Gitomer, D. H., Gradinarova-Kirova, D., Shcherbakov, A., & Nordin, T. L. (2013). *Year 1 report: New Jersey teacher evaluation pilot program.* New Brunswick, NJ: Rutgers Graduate School of Education.

Hall, G. E., & Hord, S. M. (2006). Implementing change: Patterns, principles, and potholes (2nd ed.). Boston, MA: Pearson.

Hamilton, L. S., Stecher, B. M., & Yuan, K. (2008). *Standards-based reform in the United States: History, research, and future directions.* Santa Monica, CA: Center on Education Policy. RAND Corporation.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill and Melinda Gates Foundation.

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly, 70,*(5), 759–779.

Marzano R. J. (2011). *The Marzano Teacher Evaluation Model.* Retrieved from http://pages.solution-tree.com/rs/solutiontree/images/MarzanoTeacherEvaluationModel.pdf

Milanowski, A., & Kimball, S. M. (2003). *The framework-based teacher performance assessment systems in Cincinnati and Washoe.* (No. TC-03-07). Madison, WI: Consortium for Policy Research in Education.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Thousand Oaks, CA: Sage Publications.

New Jersey Department of Education. (2012). *Notice of Grant Opportunity: Excellent Educators for New Jersey (EE4NJ) pilot program teacher effectiveness evaluation system Cohort 2A.* Trenton, NJ: New Jersey Department of Education. Retrieved from http://education.state.nj.us/events/details.php?recid=17045

New Jersey Department of Education. (2012). *2012–2013 Teacher Evaluation Pilot Toolkit.* Trenton, NJ: New Jersey Department of Education.

New Jersey Department of Education. (2011). *Notice of Grant Opportunity: Excellent Educators for New Jersey (EE4NJ) pilot program teacher effectiveness evaluation system.* Trenton, NJ: New Jersey Department of Education.

New Jersey Educator Effectiveness Task Force. (2011). *Interim report.* Trenton, NJ: New Jersey Department of Education.

Podgursky, M., & Springer, M. G. (2007). Credentials versus performance: Review of the teacher performance pay research. *Peabody Journal of Education, 82*(4), 551–573.

Stronge, J. H. (2006). Teacher evaluation and school improvement: Improving the educational landscape. In J. H. Stronge (Ed.), *Evaluating teaching: A guide to current thinking and best practice* (1–23). Thousand Oaks, CA: Corwin Press.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Brooklyn, NY: The New Teacher Project.

White, B. R., Cowhy, J., Stevens, W. D., & Sporte, S. E. (2012). *Designing and implementing the next generation of teacher evaluation systems: Lessons learned from case studies in five Illinois districts.* Chicago, IL: University of Chicago Consortium on Chicago School Research.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57–67.

Rutgers University, Graduate School of Education

EE4NJ External Evaluation

Final Report, December 2012

Appendices

Appendix A

RU-GSE Site Visit Guide

Spring 2012

SITE VISIT GUIDE

Following Patton (2002), this is an unstructured, open-ended interview guide. Questions are intended to cover the major topics to be addressed. The interviewer is expected to adjust question wording and order to maximize rapport with the respondent and ensure that full information is provided. Following the principles of the site visit guide (Yin, 1989), this guide covers all the questions to be asked of each site since the goal of the research is to describe the situation in each school district, and individual perceptions are primarily important for clarifying district conditions. Interviewers will have discretion to ask questions that are appropriate to the respondent's position--i.e., different questions will be selected for principals and district student data coordinators--and to make sure that adequate information is collected across the district if all questions cannot be answered of every respondent.

Individual Interview Guide

TEACHER OBSERVATION SYSTEM
1. What is the district doing to ensure that it gets the requisite number of teacher observations completed for every teacher?
2. What is the district doing to ensure that it gets *accurate* teacher observation scores?
3. What makes it hard to collect teacher observation data?
4. What makes it easy to collect teacher observation data?
5. What helps or hinders your district's efforts to store and retrieve teacher observation data?
6. What have been your most important sources of knowledge about how to collect, analyze, and use teacher observation data**?**
7. What contributes to the accuracy of your teacher observation system?
8. What undermines the accuracy of your teacher observation system?
9. In what ways is the teacher observation system useful for planning supervision, professional development, changes in the curriculum or other things?
10. What could be done to improve the usefulness of the teacher observation system?
11. How is the teacher observation system facilitating or impeding collaboration among educators in this district?

STUDENT GROWTH SCORES (TESTED SUBJECTS)
1. What is the district doing to ensure that it has student growth scores for teachers in tested subjects?
2. What is the district doing to ensure that it has *accurate* growth scores for teachers in tested subjects?
3. What makes it hard or easy to get teacher growth scores?
4. What helps or hinders your district's ability to store and retrieve teacher growth scores?
5. What contributes to the accuracy of your teacher growth scores?
6. What undermines the accuracy of your teacher growth scores?
7. In what ways are your teacher growth scores useful for planning supervision, professional development, changes in the curriculum or other things?
8. What could be done to improve the usefulness of the teacher growth scores?
9. How are teacher growth scores facilitating or impeding collaboration among educators in this district?

TEST SCORES IN UNTESTED SUBJECTS
1. What is the district doing to ensure that it has measures of student growth that it can link to

teachers from untested subjects?
2. What is the district doing to ensure that it has *accurate* measures of student growth that it can link to teachers from untested subjects?
3. What helps or hinders your district's ability to store and retrieve measures of student growth in untested subjects? Measures of student growth linked to teachers?
4. What contributes to the accuracy of your growth measures in untested areas?
5. What gets undermines the accuracy of your growth measures in untested areas?
6. (Only if experience over a year) In what ways are growth measures in untested areas useful for planning supervision, professional development, changes in the curriculum or other things?
7. (Only if experience over a year) What could be done to improve the usefulness of the teacher growth scores?
8. How are teacher growth scores facilitating or impeding collaboration among educators in this district?

Teacher Focus Groups Interview Guide

QUESTIONS ABOUT OBSERVATIONS
1. Thinking about the teacher observations that you have had this year, what made them better or worse than the observations you had last year?
2. What have been your most important sources of knowledge about how to collect, analyze, and use teacher observation data?
3. In what ways is the teacher observation system useful (or not) for planning supervision, professional development, changes in the curriculum or other things?
4. How do you judge the expertise of the person who observes you?
5. How well do you know the people who have observed you this year?
6. How is the teacher observation system facilitating or impeding collaboration among educators in this district?
7. What could be done to improve the usefulness of the teacher observation system?

QUESTIONS ABOUT GROWTH SCORES
8. What progress has the district made in creating a system of growth scores for your students and classes?

STUDENT GROWTH SCORES (TESTED SUBJECTS)
9. How are teacher growth scores facilitating or impeding collaboration among educators in this district?

TEST SCORES IN UNTESTED SUBJECTS

10. How are measures of progress in untested areas facilitating or impeding collaboration among educators in this district?

Appendix B

RU-GSE Administrator Survey

Summer 2012

RU-GSE Administrator Survey, Summer 2012

Default Question Block

Q1.
This survey is part of an evaluation of Excellent Educators for New Jersey (EE4NJ), an initiative to pilot a new teacher evaluation system in several LEAs in the 2011-2012 school year. The evaluation of this pilot program is conducted by an independent contractor, Rutgers University's Graduate School of Education. Your responses to this survey will help us to learn about the successes and challenges involved in the implementation of the new teacher evaluation system and will inform future plans about improving and implementing the new teacher evaluation system statewide.

With minor exceptions, this survey contains only quick-answer, multiple-choice responses. We estimate that you should be able to complete the survey in approximately 20 minutes. Your responses to this survey will be kept strictly confidential and will be only reported at the aggregate. The results will never be reported in any way that would permit any response to be associated with a specific individual.

After completing the survey, you will be assigned an identification number that will correspond to your responses so that your name is not linked to your responses. Information gathered from the survey will be kept confidential and will be used only for the purpose of this project by the Rutgers Research Team. Because the information you provide in this study is strictly confidential, there will be essentially no risk from your participation. All study data will be kept for 7 years after the completion of the study and then will be destroyed.

The information you provide in this study will enhance our ability to understand the effectiveness of the New Jersey's teacher evaluation program, piloted through a grant funded by the New Jersey Department of Education (NJDOE). Information learned will be shared with the NJDOE and districts. Your participation in this study is completely voluntary.

If you have any questions concerning this project, please feel free to contact the Principal Investigator, Dr. William Firestone, or the Rutgers University's Institutional Review Board using the following contact information:

| Dr. William Firestone, Principal Investigator

Rutgers University Graduate School of Education
10 Seminary Place, New Brunswick, NJ
Tel: 732-932-7496 x 8231
Email: william.firestone@gse.rutgers.edu | Rutgers University Institutional Review Board for the Protection of Human Subjects

Office of Research & Sponsored Programs
3 Rutgers Plaza, New Brunswick, NJ 08901-8559
Tel: 848-932-0150
Email: humansubjects@orsp.rutgers.edu |

By selecting "Agree," you will be agreeing to the conditions of the survey. You should then click on the forward arrow to be taken to the survey.

◯ Agree

◯ Disagree

Q2. Thank you for your time and patience in completing this survey. Please read each question and the possible responses carefully, and then fill in the requested information or mark the appropriate check boxes.

Q3. **Professional Background**

Q4. School District Name:

○ Alexandria

○ Bergenfield

○ Elizabeth

○ Monroe

○ Ocean City

○ Pemberton

○ Red Bank

○ Secaucus

○ West Deptford

○ Woodstown

Q5.
What is the highest degree you have earned?

○ Bachelor's degree (B.A., B.S., etc.)

○ Master's degree (M.A., M.A.T., M.B.A., M.Ed., M.S., etc.)

○ Educational specialist or professional diploma (at least one year beyond master's level)

○ Doctorate or first professional degree (Ph.D., Ed.D., M.D., L.L.B., J.D., D.D.S.)

Q6.
What is your current position within the school system?

○ Superintendent

○ Assistant Superintendent

○ Principal

○ Assistant or Vice Principal

○ Supervisor of instruction (appropriately certified individual assigned with the responsibility for the direction and guidance of the work of teaching staff members)

○ Other district level line or staff position

○ Elementary or secondary school teacher

○ Other:

Q7. How many years have you worked in your current position **within this school system**?

○ 1-3 years

○ 4-6 years

○ 7-10 years

○ More than 10 years

Q8. How many years have you worked in each of the following positions in your career?

0	Superintendent
0	Assistant Superintendent
0	Principal
0	Assistant or Vice Principal
0	Supervisor of instruction
0	Other district level line or staff position
0	Elementary or secondary school teacher

Q9. When you taught, what grades did you teach? (Please select all that apply.)

☐ Pre-K

☐ Kindergarten

☐ Grade 1

☐ Grade 2

☐ Grade 3

☐ Grade 4

☐ Grade 5

☐ Grade 6

☐ Grade 7

☐ Grade 8

☐ Grade 9

☐ Grade 10

☐ Grade 11

☐ Grade 12

Q10.
When you taught, what subjects did you teach? (Please select all that apply.)

☐ 21st Century Life and Careers

☐ English Language Learners (ELL)/English as a Second Language (ESL)

☐ Health and Physical Education

☐ Language Arts

☐ Mathematics

☐ Science

☐ Social Studies

☐ Special Education

☐ Technology

☐ Visual and Performing Arts

☐ World Language

☐ Other:

Q11.
We are interested to learn about the experience and perspective of all who are involved in the
implementation of the new teacher evaluation system, whether they are charged with conducting the
formal observations that feed into the teacher evaluation score, are overseeing the implementation in
their capacity as school administrators, or do both.

Q12.
How would you describe your role in the implementation of the new teacher evaluation system?

○ I am a school or district administrator (principal, assistant superintendent for testing, etc.) who helps to administer the
district's system for teacher supervision and evaluation **and** who also conducts formal observations of teachers as part of
their evaluation.

○ I am a school or district administrator who helps to administer the district's system for teacher supervision and evaluation,
but who does not conduct formal observations of teachers as part of their evaluation.

○ I conduct formal observations of teachers as part of their evaluation but I do not help administer the district's system for
teacher supervision and evaluation.

○ I have no responsibility for evaluating teachers, nor do I have any responsibility or involvement in the selection or
implementation of the teacher evaluation system

Q13.
Choice & Early Implementation of New Teacher Evaluation System

Q14.
Now, we'd like to ask you a few questions about your school district's choice of the new teacher
evaluation system that you are using for the pilot teacher evaluation program.

Q15. How would you describe the amount of support your district received in the following dimensions
from the New Jersey Department of Education during implementation of the new teacher evaluation
system?

	A great deal	Some	Not much	Not at all	Don't know
a. Choosing a new teacher evaluation system	○	○	○	○	○
b. Helping to create training programs	○	○	○	○	○
c. Selecting a data management system	○	○	○	○	○
d. Helping to troubleshoot issues as they arose	○	○	○	○	○
e. General level of support	○	○	○	○	○

Q16.
How would you describe the level of support you received from the **vendor** (online tool provider)
regarding the implementation of the new teacher evaluation system?

○ Outstanding (timely and helpful)

○ Average

○ Poor

○ Nonexistent

○ Does not apply

Q17. How would you describe the level of support you received from the **school district** regarding the implementation of the new teacher evaluation system?

○ Outstanding (timely and helpful)

○ Average

○ Poor

○ Nonexistent

○ Does not apply

Q18.
To the best of your knowledge has your school and/or your school district invested new or existing resources (including human resources) into the implementation of the new teacher evaluation system in the **last three months of the 2011-2012 school year**? Resources include, but are not limited to, personnel, technology, and services from external contractors.

○ Mostly new resources

○ Mostly existing resources that were diverted to this purpose

○ About equal portion of new and existing resources

○ Don't know or unsure

Q19.
What kinds of new resources have your school and/or your school district invested in the implementation of the new teacher evaluation system in the last **three months of the 2011-2012 school year**? (Please check all that apply)

☐ Hiring new personnel to supervise, conduct and/or support the implementation of the new teacher evaluation system

☐ Purchasing new or additional designated technology (e.g., audio or video equipment, transcribers, data storage devices, etc.) to support the new teacher evaluation system.

☐ Contracting external professional services to conduct teacher evaluations.

☐ Other:

Q20.
What is being done in your school and/or school district to ensure the optimal implementation of the new teacher evaluation system? (Please check all that apply)

☐ Holding information sessions in 2011/2012 for teachers and school administrators about the new system

☐ During the 11-12 academic year training on the new system was offered to current teachers

☐ During the 11-12 academic year teacher leaders or liaisons were Identified, provided with special training on the new system and asked to serve as a first line of question answerers or problem solvers about the new system.

☐ During the 11-12 academic year all evaluators were equipped with relevant classroom observation protocols

☐ During the 11-12 academic year all evaluators were offered demonstrations of executing classroom observation protocols

☐ All evaluators received scoring and rating calibration sessions

☐ During the 11-12 academic year inter-rater reliability for observers was assessed

☐ During the 11-12 academic year verification and certification that all evaluators apply the new system validly and reliably was performed

☐ Planning for training new teachers on the observation system for the 2012-2013 school year

☐ Planning for training observers for the 2012-2013 school year

☐ Other:

Q21. Now, we'd like to ask you a few questions about your school district's choice of an online observation data management tool, used along with the teacher observation framework.

Q22.

How well prepared do you feel about using the online data management tool?

○ Fairly well prepared

○ Somewhat well prepared

○ Not well prepared

○ Does not apply

Q23.

How hard or easy to use are the different features of the online data management tool?

	Very easy	Easy	Hard	Very hard	Does not apply
a. Log into the system and input information	○	○	○	○	○
b. Use of the tool during actual teacher observations	○	○	○	○	○
c. Use of the tool for communication and exchange of information	○	○	○	○	○
d. Access already stored data to identify patterns of practice in the school or district	○	○	○	○	○
e. Access already stored data to prepare reports on teacher evaluation data for various audiences other than the individual teacher being observed	○	○	○	○	○
f. Management of already uploaded information	○	○	○	○	○

Q24.

Please rate the quality of the service offered by the provider's customer support group?

○ Poor

○ Fair

○ Good

○ Very Good

○ Excellent

○ Does not apply

Q25.

In terms of technical problems (if any) what are the most frequently encountered issues:

☐ Short log-in period (getting kicked out and loosing data before saving)

☐ Problems saving information while working online

☐ Problems with crashing/time-outing of the tool

☐ Personal account information issues – lengthy password changes, updates of identification information etc.

☐ Poor accessibility of previously uploaded information

☐ Ineffective online help feature/technical support

☐ Other:

☐ I have rarely encountered any significant technical problems

Q26.
What of the following are in your opinion legitimate concerns with regard to the online data management tool? Please rate the importance of each.

	Concern is not at all well addressed in my district	Concern is somewhat addressed in my district	Concern is well addressed in my district	I don't know how well this concern is addressed in my district	Does not apply
a. Privacy issues – transparency about issues of ownership of the personal and the observational data uploaded in the system	○	○	○	○	○
b. Security issues – built-in safeguard mechanism with regard to access of the personal and observational data uploaded in the system.	○	○	○	○	○
c. Adequate alignment of the online data management tool with the specific evaluation components of the teacher observation framework adopted in the district	○	○	○	○	○
d. Clear understanding of how personnel data should be handled in case the district is to change the current service provider	○	○	○	○	○
e. Transparent district policy with regard to training and responsibility of the data management personnel on the school district level.	○	○	○	○	○
f. Other:	○	○	○	○	○

Q27.
Experience with New Teacher Evaluation System

Q28.
How many formal teacher observations (excluding walkthroughs) during the 2011/2012 academic year did you complete using the new teacher evaluation system? If you are unsure of the exact number, please provide your best estimate.

Q29. How many walkthrough observations during the 2011/2012 academic year did you complete using the new teacher evaluation system? If you are unsure of the exact number, please provide your best estimate.

Q30.
On average, how many hours, would you say, were required to complete a single formal teacher observation (not a walkthrough), including pre- and post-observation meetings, the time to write up your observation and any other tasks involved?

Q31.
Of the formal teacher observations you completed (not walkthroughs) using the new teacher
evaluation system, what percentage would you estimate were:

	None					Some					All
	0	10	20	30	40	50	60	70	80	90	100
Completed in the presence of a second observer?											
Included collection of classroom artifacts?											
Delivered to the central office data storage system?											
Discussed with your supervisor or the person who is managing EE4NJ in your school district?											
Discussed with the teacher as feedback?											

Q32.
When you do a formal observation, which of the following records or artifacts do you keep after you
complete the observation? (Please check all that apply)

☐ A written record of the observation

☐ A file of any forms or materials given to you by the teacher (e.g., lesson plans, sample materials)

☐ A record of the ratings the teacher has received in a standardized format used by other observers in your district

☐ A report that is delivered online to a central data storage facility

☐ Other:

Q33.
When you do a formal observation, which of the following records or artifacts do you submit to the
district office? (Please check all that apply)

☐ A written record of the observation

☐ A file of any forms or materials given to you by the teacher (e.g., lesson plans, sample materials)

☐ A record of the ratings the teacher has received in a standardized format used by other observers in your district

☐ A report that is delivered online to a central data storage facility

☐ Other:

Q34.
When you do a formal observation, which of the following records or artifacts do you enter into a centrally approved data base? (Please check all that apply)

☐ A written record of the observation

☐ A file of any forms or materials given to you by the teacher (e.g., lesson plans, sample materials)

☐ A record of the ratings the teacher has received in a standardized format used by other observers in your district

☐ A report that is delivered online to a central data storage facility

☐ Other:

[   ]

Q35. Overall, how well do you think training on the data management system (i.e. iObservation or Teachscape) for your district accomplished the following:

	Very well accomplished	Accomplished	Somewhat accomplished	Not accomplished	Not at all accomplished	Does not apply
a. Help you to understand how to input data	○	○	○	○	○	○
b. Help you to understand how to provide feedback for a single teacher	○	○	○	○	○	○
c. Help you understand how to retrieve data to understand pattern of strengths and weaknesses in group of teachers	○	○	○	○	○	○

Q36. Please rate the data management system (i.e. iObservation or Teachscape) for your district on the following dimensions.

	Excellent	Good	Fair	Poor	Very Poor	Do Not Know
a. Intuitiveness	○	○	○	○	○	○
b. Ease of entering data	○	○	○	○	○	○
c. Ease of saving data	○	○	○	○	○	○
d. Ease of retrieving data	○	○	○	○	○	○
e. Stability and reliability (lack of technical issues)	○	○	○	○	○	○
f. Security of personal information	○	○	○	○	○	○

Q37.
Reflecting on your experience with teacher observation using the new teacher evaluation system, how
frequently have you encountered any, some or all of the following challenges?

	All the time	Frequently	Not frequently	Not at all	Does not apply	Don't know
a. Difficulty scheduling classroom observations.	○	○	○	○	○	○
b. Difficulty scheduling sufficient time to complete classroom observations.	○	○	○	○	○	○
c. Difficulty obtaining samples of classroom artifacts.	○	○	○	○	○	○
d. Too many distractions in the classroom.	○	○	○	○	○	○
e. Difficulty using the classroom observation protocol to the letter.	○	○	○	○	○	○
f. Difficulty storing and managing classroom observation data.	○	○	○	○	○	○
g. Difficulty coding and analyzing classroom observation data.	○	○	○	○	○	○
h. Difficulty using the online data system.	○	○	○	○	○	○

Q38. Evaluation of New Teacher Evaluation System

Q39.
Overall, how comfortable do you feel observing and providing feedback to teachers in the following
subject areas using the **new** system of teacher evaluation in your district?

	Very comfortable	Comfortable	Somewhat comfortable	Uncomfortable	Very uncomfortable	Does not apply
a. Math	○	○	○	○	○	○
b. English Language Arts (ELA)	○	○	○	○	○	○
c. Science	○	○	○	○	○	○
d. Social Studies	○	○	○	○	○	○
e. Art	○	○	○	○	○	○
f. Music	○	○	○	○	○	○
g. Physical Education (PE)	○	○	○	○	○	○
h. Evaluating teachers in general	○	○	○	○	○	○

Q40.
In comparison to your previous teacher observation system, how would you rate the current (new)
system on the following dimensions:

	The current system is much better than the previous system	The current system is better than the previous system	The current system is neither better nor worse than the previous system	The current system is worse than the previous system	The current system is much worse than the previous system	Does not apply	Don't know
a. Formalization (clear rules, steps, procedures, reporting forms)	○	○	○	○	○	○	○
b. Ease of use	○	○	○	○	○	○	○
c. Grounding in research	○	○	○	○	○	○	○
d. Intuitiveness	○	○	○	○	○	○	○
e. Usefulness for providing guidance to teachers	○	○	○	○	○	○	○

Q41.
Below is a series of statements about the new teacher evaluation system used in your school district. For each statement, please indicate whether you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree or strongly disagree with that statement. Please answer based on your personal experience and observation. Remember that your answers are confidential.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree	Does not apply	Don't know
a. I feel comfortable using my district's system for assessing teachers.	○	○	○	○	○	○	○
b. In my experience, the district's system for assessing teachers generates accurate assessments.	○	○	○	○	○	○	○
c. In my experience, the district's system for assessing teachers is fair.	○	○	○	○	○	○	○
d. In my experience, the district's system for assessing teachers generates assessments that help provide individual feedback and design professional development.	○	○	○	○	○	○	○
e. In my experience, the district's system for assessing teachers is well aligned with the district curriculum.	○	○	○	○	○	○	○
f. The district's system for assessing teachers clearly separates accomplished from unaccomplished teachers.	○	○	○	○	○	○	○
g. The district's system for assessing teachers fits well with other school/district initiatives.	○	○	○	○	○	○	○
h. The district's system for assessing teachers provides a firm basis for making teacher tenure and promotion decisions and weeding out weak teachers.	○	○	○	○	○	○	○
j. The district's system for assessing teachers helps this district meet its accountability requirements under NCLB and other external mandates.	○	○	○	○	○	○	○
k. The district's system for assessing teachers helps improve student achievement.	○	○	○	○	○	○	○
l. The district's system for assessing teachers consumes resources that could be better spent on promoting key district improvement initiatives.	○	○	○	○	○	○	○

Q42.
Please indicate how much you agree or disagree with the following statements about *your* perceptions of the new teacher observation system.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
a. I feel adequately informed about the new observation system	○	○	○	○	○
b. I feel that the new observation system takes too much time	○	○	○	○	○
c. I do not feel prepared for the new observation system	○	○	○	○	○
d. I understand the new observation system	○	○	○	○	○
e. I can give useful feedback under the new observation system	○	○	○	○	○
f. The new observation system provides a fair picture of teaching	○	○	○	○	○

Q43. Please indicate how much you agree or disagree with the following statements about your experience with the new teacher evaluation system.

	Strongly agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
a. In my school evaluation criteria and indicators are appropriate.	○	○	○	○	○
b. Existing instruments for teacher performance evaluation are clear.	○	○	○	○	○
c. Existing evaluation criteria take into account the context of teaching.	○	○	○	○	○
d. The evaluation process at my school allows teachers to explain decisions and actions.	○	○	○	○	○
e. Rating scales used to evaluate teachers' performance are appropriate.	○	○	○	○	○
f. I am able to give useful feedback to teachers.	○	○	○	○	○
g. I feel that in my school teachers' work and achievements are recognized.	○	○	○	○	○
h. I feel that I have the required knowledge and competencies to appraise teachers.	○	○	○	○	○
j. I feel that I have received adequate training to perform their job.	○	○	○	○	○
k. In general, I think that the feedback that I give focuses upon suggestions for improvement.	○	○	○	○	○

Q44. In general, what kind of an effect do you think the new teacher evaluation system has had on:

	Positive	Negative	No Effect	Don't Know
a. Professional Development	○	○	○	○
b. Collaboration	○	○	○	○
c. Your School	○	○	○	○

Q45. Since the implementation of the new teacher observation system, please estimate how your workload has changed:

	Much more time spent	More time spent	About the same time spent	Less time spent	Much less time spent	Does not apply
Doing observations	○	○	○	○	○	○
Entering data	○	○	○	○	○	○
Other administrative tasks or job responsibilities	○	○	○	○	○	○

Q46. If you have any other comments or thoughts you would like to add which have not been captured by the previous questions, please write them below:

Q47.
Closing and Additional Informed Consent

Q48.
That completes our survey. Thank you very much for your time and cooperation. If you have any questions, you may contact Dr. William Firestone at 732-932-7496, X8231. If you have any questions about your rights as a research participant, you may contact the administrator of the Rutgers Institutional Review Board at: 732-932-0150, ext. 2104.

Please do not forget to click on the forward arrow in the right bottom corner in order to successfully SUBMIT THE SURVEY.
Thank you!

Appendix C

RU-GSE Teacher Survey, Spring 2012

Default Question Block

.
This survey is part of an evaluation of Excellent Educators for New Jersey (EE4NJ), an initiative to pilot a new teacher evaluation system in several school districts during the 2011-2012 school year. The evaluation of this pilot program is conducted by an independent contractor, Rutgers University's Graduate School of Education. Your responses to this survey will help us to learn about the successes and challenges involved in the implementation of the new teacher evaluation system and will inform future plans about improving and implementing the new teacher evaluation system statewide.

With minor exceptions, this survey contains only quick-answer, multiple-choice responses. We estimate that you should be able to complete the survey in approximately 20 minutes. Your responses to this survey will be kept strictly confidential and will be only reported at the aggregate. The results will never be reported in any way that would permit any response to be associated with a specific individual.

Your participation in this survey is completely confidential - we will not ask you to provide your name or other identifiable information beyond your professional experience and school district affiliation. Information gathered from the survey will be kept confidential and will be used only for the purpose of this project by the Rutgers Research Team. Since the information you provide in this study is strictly confidential, there will be essentially no risk from your participation. All study data will be kept for 7 years after the completion of the study and then will be destroyed.

The information you provide in this study will enhance our ability to understand the effectiveness of the New Jersey's teacher evaluation program, piloted through a grant funded by the New Jersey Department of Education (NJDOE). Information learned will be shared with the NJDOE and districts. Your participation in this study is completely voluntary.

If you have any questions concerning this project, please feel free to contact the Principal Investigator, Dr. William Firestone, or the Rutgers University's Institutional Review Board using the following contact information:

| Dr. William Firestone, Principal Investigator

Rutgers University Graduate School of Education
10 Seminary Place, New Brunswick, NJ
Tel: 732-932-7496 x 8231
Email: william.firestone@gse.rutgers.edu | Rutgers University Institutional Review Board for the Protection of Human Subjects

Office of Research & Sponsored Programs
3 Rutgers Plaza, New Brunswick, NJ 08901-8559
Tel: 848-932-0150
Email: humansubjects@orsp.rutgers.edu |

Thank you for your time and patience in completing this survey. Please read each question and the possible responses carefully, and then fill in the requested information or mark the appropriate check boxes.

 By selecting "Agree," you will be agreeing to the conditions of the survey. You should then click on the forward arrow to be taken to the survey.

⚪ Agree

⚪ Disagree

. **Professional Background**

Q1. Are you a teacher currently teaching students?

⚪ Yes

⚪ No

Q2. School District Name:

◯ Alexandria

◯ Bergenfield

◯ Elizabeth

◯ Monroe

◯ Ocean City

◯ Pemberton

◯ Red Bank

◯ Secaucus

◯ West Deptford

◯ Woodstown

Q3.
What is the highest degree you have earned?

◯ Bachelor's degree (B.A., B.S., etc.)

◯ Master's degree (M.A., M.A.T., M.B.A., M.Ed., M.S., etc.)

◯ Educational specialist or professional diploma (at least one year beyond master's level)

◯ Doctorate or first professional degree (Ph.D., Ed.D., M.D., L.L.B., J.D., D.D.S.)

Q4. Which grades do you teach? Select all that apply.

☐ Pre-K

☐ Kindergarten

☐ Grade 1

☐ Grade 2

☐ Grade 3

☐ Grade 4

☐ Grade 5

☐ Grade 6

☐ Grade 7

☐ Grade 8

☐ Grade 9

☐ Grade 10

☐ Grade 11

☐ Grade 12

Q5.
Which statement best describes the way YOUR classes at your current school are organized?

○ You instruct several classes of different students most or all of the day in one or more subjects (sometimes called Departmentalized Instruction).

○ You are an elementary school teacher who teaches only one subject to different classes of students (sometimes called an Elementary Subject Specialist).

○ You instruct the same group of students all or most of the day in multiple subjects (sometimes called a Self-Contained Class).

○ You are one of two or more teachers, in the same class, at the same time, and are jointly responsible for teaching the same group of students all or most of the day (sometimes called Team Teaching).

○ You instruct a small number of selected students released from or in their regular classes in specific skills or to address specific needs (sometimes called a "Pull-Out" Class or "Push-In" Instruction).

Q6.
What is/are the subjects you currently teach? Select all that apply.

☐ 21st Century Life and Careers

☐ English Language Learners (ELL)/English as a Second Language (ESL)

☐ Health and Physical Education

☐ Language Arts

☐ Mathematics

☐ Science

☐ Social Studies

☐ Special Education

☐ Technology

☐ Visual and Performing Arts

☐ World Language

☐ Other:

Q7. How many years will you have been teaching at the end of the current school year?

○ 1-3

○ 4-6

○ 7-10

○ 11 or more

.
Current Teacher Evaluation in Comparison to Previous System

Q8. Please indicate how much you agree or disagree with the following statements about teacher evaluation systems in general.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Teacher evaluation is essential to raise the standards of teaching and learning.	○	○	○	○	○
Teacher evaluation should primarily focus on the identification of my professional development needs.	○	○	○	○	○
Teacher evaluation aims at meeting the minimum standards.	○	○	○	○	○
Teacher evaluation aims at providing useful information for teachers to improve their performance.	○	○	○	○	○
Teacher evaluation should be based upon a list of professional competencies or behaviors.	○	○	○	○	○
As a professional, I am entitled to have my performance appraised.	○	○	○	○	○
Teacher evaluation should aim primarily at making managerial decisions.	○	○	○	○	○
Teacher evaluation aims to enhance teachers' reflection on their practice.	○	○	○	○	○
Teacher evaluation should be used both for professional development and accountability purposes.	○	○	○	○	○

Q9.
In comparison to your previous teacher observation system, how would you rate the current (new) teacher observation system on the following dimensions:

	The current system is much better than the previous system	The current system is better than the previous system	The current system is neither better nor worse than the previous system	The current system is worse than the previous system	The current system is much worse than the previous system	Does not apply	Don't know
Formalization (clear rules, steps, procedures, reporting forms)	○	○	○	○	○	○	○
Ease of use	○	○	○	○	○	○	○
Grounding in research	○	○	○	○	○	○	○
Intuitiveness	○	○	○	○	○	○	○
Usefulness for providing guidance to teachers	○	○	○	○	○	○	○

Q10.
Were you responsible for informing other teachers about the new teacher evaluation system?

○ Yes

○ No

.
Training on the New Teacher Evaluation System

.
The next set of questions is about the training you received on the **new** teacher evaluation system in your district as part of the EE4NJ pilot program. There are no right or wrong answers. We are simply interested to learn about your experience with the new teacher evaluation system.

Q11.
How many hours of training or education have you personally received on the new teacher observation system:

○ 0

○ 1-2

○ 3-4

○ 5-8

○ 9-16

○ 17-24

○ 25-32

○ 33-40

○ More than 40

Q12.
Overall, how well, would you say, the training accomplished each of the following:

	Very well accomplished	Accomplished	Somewhat accomplished	Not accomplished	Not at all accomplished	Does not apply
Help you understand your district's system of assessing teachers	○	○	○	○	○	○
Help you understand the main components of the teacher evaluation: teacher practice and direct measures of student achievement	○	○	○	○	○	○
Help you understand the process of linking student growth scores to teacher observations in tested subjects	○	○	○	○	○	○
Help you understand the process of linking student growth scores to teacher observations in non-tested subjects	○	○	○	○	○	○
Help you understand the information needed for you to be accurately assessed	○	○	○	○	○	○
Help you understand the criteria for assessment of teachers' planning process	○	○	○	○	○	○
Help you to understand the criteria for assessment of teachers' instructional practices	○	○	○	○	○	○
Help you to understand the feedback after an observation	○	○	○	○	○	○
Help you to understand what underlies judgments of teacher quality	○	○	○	○	○	○
Help you to understand potential biases in the way teachers are evaluated	○	○	○	○	○	○

Q13.
Have you been evaluated by the new teacher observation system?

○ Yes, I have been evaluated as part of the new evaluation system at least once.

○ No, but I will be evaluated in the future.

○ No, and I will not be evaluated.

.
Evaluation of New Teacher Evaluation System

.

This question asks for your personal evaluation of the new teacher evaluation system based on your experience.

Q14.
Below is a series of statements about the **new teacher evaluation system** used in your school district. For each statement, please indicate whether you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree or strongly disagree with that statement. Please answer based on your personal experience and observation. Remember that your answers are confidential.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree	Does not apply	Don't know
I feel comfortable being assessed by the district's new evaluation system.	○	○	○	○	○	○	○
The district's system for assessing teachers generates accurate assessments.	○	○	○	○	○	○	○
The district's system for assessing teachers is fair.	○	○	○	○	○	○	○
The district's system for assessing teachers generates assessments that provide constructive individual feedback and promote professional development.	○	○	○	○	○	○	○
The district's system for assessing teachers is well aligned with the district's curriculum.	○	○	○	○	○	○	○
The district's system for assessing teachers clearly separates accomplished from unaccomplished teachers.	○	○	○	○	○	○	○
The district's system for assessing teachers fits well with other school/district initiatives.	○	○	○	○	○	○	○
The district's system for assessing teachers provides a firm basis for making teacher tenure and promotion decisions and weeding out weak teachers.	○	○	○	○	○	○	○
The district's system for assessing teachers helps this district meet its accountability requirements under NCLB and other external mandates.	○	○	○	○	○	○	○
The district's system for assessing teachers helps improve student achievement.	○	○	○	○	○	○	○
The district's system for assessing teachers consumes resources that could be better spent elsewhere.	○	○	○	○	○	○	○
The district's system for assessing teachers is relevant for my subject area and teaching methodology	○	○	○	○	○	○	○

Q15.
Please indicate how much you agree or disagree with the following statements about *your* perceptions of the new teacher observation system.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I feel adequately informed about the new observation system	○	○	○	○	○
I feel that the new observation system takes too much time	○	○	○	○	○
I do not feel prepared for the new observation system	○	○	○	○	○
I understand the new observation system	○	○	○	○	○
I receive useful feedback from observers under the new observation system	○	○	○	○	○
The new observation system provides a fair picture of my teaching	○	○	○	○	○

Q16. Please indicate how much you agree or disagree with the following statements about your experience with the new teacher evaluation system.

	Strongly agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
In my school evaluation criteria and indicators are appropriate.	○	○	○	○	○
Existing instruments for teacher performance evaluation are clear.	○	○	○	○	○
Existing evaluation criteria take into account the context of teaching.	○	○	○	○	○
The evaluation process at my school allows teachers to explain decisions and actions.	○	○	○	○	○
Rating scales used to evaluate my performance are appropriate.	○	○	○	○	○
I am given useful feedback by the evaluator.	○	○	○	○	○
I feel that in my school teachers' work and achievements are recognized.	○	○	○	○	○
I feel that the evaluators in my school have the required knowledge and competencies to appraise teachers.	○	○	○	○	○
I feel that the evaluators in my school have received adequate training to perform their job.	○	○	○	○	○
In general, I think that the feedback that I am given focuses upon suggestions for improvement.	○	○	○	○	○

Q17. Please indicate how much you agree or disagree with the following statements about your perceptions of the effects of the new teacher evaluation system.

	Strongly agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
The evaluation system encourages me to reflect on my teaching.	○	○	○	○	○
The evaluation system has made me more aware of my strengths and weaknesses as a teacher.	○	○	○	○	○
The evaluation system has led to an intensification of my work.	○	○	○	○	○
The evaluation system has increased the bureaucratic work at school.	○	○	○	○	○
The evaluation system has led to tensions among staff.	○	○	○	○	○

Q18.
I would prefer to be evaluated by:

○ A superior (principal, etc.) with whom I have a developed professional relationship

○ A superior with whom I am not very familiar

○ A teacher with whom I have a developed professional relationship

○ A teacher with whom I am not very familiar

○ Someone whom I have never met

Q19. Why do you prefer to be evaluated by that person?

Q20. Which of the below BEST describes the relationship you have with the people who observe you

○ I know them and they know me/my classroom well

○ I know them and they know me/my classroom a little

○ I know who they are but have no relationship with them

○ I do not know them

Q21. Please indicate how much you agree or disagree with the following statements about your perceptions of how you would prefer to be evaluated in the new teacher evaluation system.

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
I will excel under the new evaluation system	○	○	○	○	○
I am confident that I will be accurately evaluated in the new system	○	○	○	○	○
I feel comfortable being observed and evaluated by the current person responsible for it	○	○	○	○	○
I am confident I would score well on an evaluation done by my principal	○	○	○	○	○
I am confident I would score well on an evaluation done by an impartial observer	○	○	○	○	○
I am more likely to be accurately assessed by someone who knows my classroom and teaching well	○	○	○	○	○
Impartial observers will not understand the context of my classroom	○	○	○	○	○
An impartial observer may give a more accurate evaluation of my teaching than someone who knows me	○	○	○	○	○
I would rather be evaluated by a **direct superior** (i.e. a principal) than a peer or master teacher in my content area for purposes of **accountability**	○	○	○	○	○
I would rather be evaluated by a **direct superior** (i.e. a principal) than a peer or master teacher in my content area for purposes of **professional development**	○	○	○	○	○
I would rather be evaluated by an **impartial observer** than someone who knows me for purposes of **accountability**	○	○	○	○	○
I would rather be evaluated by an **impartial observer** than someone who knows me for purposes of **professional development**	○	○	○	○	○

Q22. Please indicate how much you agree or disagree with the following statements about your perceptions of how content knowledge affects evaluation.

	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree	Not Applicable
Evaluating good teaching in my subject area is different from evaluating good teaching in other subject areas.	○	○	○	○	○	○
A strong understanding of the pedagogy specific to my subject matter (i.e. the pedagogy of science or special education) on the part of the observer is essential for an accurate observation of my teaching	○	○	○	○	○	○
The person who evaluates me has a robust knowledge of the content I teach	○	○	○	○	○	○
The person who evaluates me has a robust understanding of what good teaching looks like in my subject area	○	○	○	○	○	○
I would prefer to be evaluated by someone who understands my content area deeply	○	○	○	○	○	○
I would be more accurately evaluated by someone who understands my content area deeply	○	○	○	○	○	○
The people who evaluate me do not understand the intricacies of teaching my subject	○	○	○	○	○	○
The new evaluation system accounts for the importance of content knowledge and content-specific pedagogy in evaluation	○	○	○	○	○	○
It is fair to be evaluated on my teaching by someone who is an expert on effective pedagogy even if they are not familiar with my subject area.	○	○	○	○	○	○
Effective teaching is generally the same across all content areas	○	○	○	○	○	○

Q23. In general, what kind of an effect do you think the new teacher evaluation system has had:

	Positive	Negative	No effect	Don't know
On your professional development	○	○	○	○
On collaboration with others	○	○	○	○
On your school	○	○	○	○

Q24. If you have any other comments or thoughts you would like to add which have not been captured by the previous questions, please write them below:

.
End of survey.

.
That completes our survey. Thank you very much for your time and cooperation. If you have any questions, you may contact Dr. William Firestone at 732-932-7496, X8231. If you have any questions about your rights as a research participant, you may contact the administrator of the Rutgers Institutional Review Board at: 732-932-0150, ext. 2104.

Please do not forget to click on the forward arrow in the right bottom corner in order to successfully SUBMIT THE SURVEY.
Thank you!