An Evaluation of the Pilot Implementation of Connecticut's System for Educator Evaluation and Development

Submitted by

UConn Center for Education Policy Analysis
Neag School of Education

Morgaen L. Donaldson
Casey D. Cobb
Kimberly LeChasseur
Rachael Gabriel
Richard Gonzales
Sarah Woulfin
Aliza Makuch

January 1, 2014

Report on the Pilot Implementation of Connecticut's System for Educator Evaluation and Development

Executive Summary

Purpose

In June 2012, the Connecticut General Assembly passed legislation mandating that that State Board of Education, in consultation with the Performance Evaluation Advisory Council, develop guidelines for a model program to evaluate and support public school teachers and administrators. This legislation also directed the University of Connecticut's Neag School of Education to study the pilot implementation of the state model and report findings to the State Board of Education and Education Committee to inform modifications to the model. Developed by the State Department of Education, the state model was entitled the System for Educator Evaluation and Development (SEED), and was implemented in 10 pilot school districts/consortia, encompassing 14 districts in total. The purpose of our study was to examine the implementation and short-term outcomes of Connecticut's new educator evaluation model in these districts. Our broader aim was to provide feedback on SEED's pilot implementation so that the model could be improved and better support improvements to teachers' practice and students' learning.

Based on data collected in all 14 pilot districts, we find that SEED was implemented in the pilot sites with a relatively high degree of fidelity to the model. Moreover, most educators in the pilot sites supported the model in the abstract and found several aspects of SEED (e.g., post-observation conferences with evaluators) quite valuable. There is some evidence of changed practices on the part of teachers and administrators due to SEED. While SEED's components were generally implemented with fidelity, opportunities for professional growth through SEED were not fully realized. To a great extent, these challenges were due to the lack of clear and consistent communication to educators about SEED and inadequate opportunities for educators to construct a robust understanding of this new approach to evaluation. Educators attempted to understand the complexities of the technical and practical aspects of SEED. However, the tight implementation timeline in some cases encouraged compliance behaviors rather than full engagement with the model.

This emphasis on compliance is not uncommon during the initial stages of a major reform initiative as participants and implementers attempt to adapt to a new set of rules and routines. We expect that districts will begin to experience greater benefits of SEED as educators become accustomed to its components and adjust SEED to their own contexts. Our data suggest that, if sufficiently supported and properly implemented, SEED has the potential to improve teachers' practice and students' learning. Therefore,

we recommend that the SEED model be maintained but that the infrastructure to clarify and support its implementation be strengthened. Accordingly, we conclude by providing eight recommendations that we believe will enhance SEED's positive impact on districts, schools, teachers, and students.

Methods

Fourteen (14) public school districts (eight stand-alone districts and two consortia) were chosen by the State Department of Education (SDE) to serve as pilot sites. Between September 2012, and October 2013, Neag researchers from UConn's Center for Education Policy Analysis collected data in each of the pilot districts. We collected three rounds of interviews with between 200 and 500 individuals in each round; two rounds of surveys with teachers with over 500 respondents in each round; and one round of surveys with school administrators (*n*=22).

Within each district, researchers interviewed superintendents, assistant superintendents (where applicable), and a sample of principals, teachers, and other educators who are evaluated under the SEED framework. We also interviewed presidents of union locals, state-level union leaders, and Regional Educational Service Center (RESC) representatives who provided the training on the SEED model to educators in the pilot districts.

The 14 pilot districts were identified by the SDE from a larger group of districts that volunteered to take part in the pilot program. In each of the 14 districts, we selected a sample of schools and educators within those schools using a two-stage, stratified sampling technique. We selected a minimum of 20% of the schools at each level (high school, middle school, elementary school) based on characteristics that might influence policy implementation such as the student population served, leadership turnover, and initiatives or programs of study (e.g., dual language program, alternative education program). Thirty-seven schools were purposely sampled to reflect a wide range of school settings and grade levels. In the consortia, we collected data in all schools. In each school, we interviewed the principal and, in some cases, an assistant principal.

Educators (teachers and others who are evaluated with the same framework) within these same schools were randomly selected to obtain an interview sample representing teachers of a range of grade levels and subjects and non-teaching personnel with a range of assignments. We deliberately stratified our samples of educators by three types: core academic teachers, related arts educators, and student support personnel. We used the ratios across these categories present in each school to determine the number of interview participants in each category for that school. Random samples of educators were then drawn by researchers.

Findings

We gathered data related to SEED's implementation, educators' experiences with SEED, and initial outcomes of SEED.

Implementation

We found that all pilot districts implemented almost all aspects of the SEED model. Districts focused their efforts on two components of the SEED model in particular: the Student Learning Objectives (SLOs) component, worth 45% of a teacher's rating, and Instructional Practice, constituting 40% of a teacher's rating. Participants reported that they had received more observations and feedback than in prior years.

Specifically, we found:

- The number of baseline observations required by SEED (3 formal, 3 informal) proved challenging, but most districts were able to complete 2 formal and 2 informal observations:
    - 69% of teachers surveyed reported having two or more informal observations; 64% reported two or more formal observations
- Educators reported increased time on evaluation activities compared to prior (pre-SEED) years:
    - 50% of teachers reported that they had been observed more than in previous years
    - 74% of teachers reported spending more time on goal setting than under previous evaluation systems

We found that the following components of SEED were implemented with less fidelity to the model than the SLO or instructional practice components: parent/peer feedback, student feedback, and the development of professional learning opportunities connected to evaluation results. Moreover, SEED for school administrators was implemented with less fidelity and on a much more compressed timeline than was SEED for teachers. This seemed largely due to districts' emphasis on implementing SLOs and observation for teachers.

Experiences

We found that most educators in the pilot districts agreed with the SEED model in theory. Very few educators rejected SEED out of hand. Participants, especially teachers, raised concerns about the way in which the SEED pilot was implemented. Teachers were particularly concerned about lack of training and information on how to develop SLOs. Specialists were particularly vocal about the ways in which SEED did not address their positions. Principals raised objections to the number of formal and informal observations required by the model. District leaders raised few objections other than concerns about the rushed implementation of the model in the pilot year. On the whole, participants feel that SEED should be better supported through additional guidance on SLOs and increased professional development. This may help shift the perception of many educators in the pilot districts that SEED was a system emphasizing their evaluation rather than their development.

In addition, we found:

- 58% of teachers and 59% of administrators surveyed felt that their summative ratings under SEED were accurate
- 57% of teachers surveyed felt that their post-observation conferences under SEED were "valuable"
- Of those teachers who reported that they had spent more time in post-observations conferences under SEED than under prior systems, 69% felt that these meetings were valuable
- 55% of teachers surveyed reported that being observed under SEED was somewhat or very valuable
- 94% of administrators reported that observing teachers under SEED was somewhat or very valuable to them
- 68% of teachers surveyed found analyzing student data valuable

Educators raised concerns about whether their evaluators had sufficient time and resources to carry out SEED as the 2012-13 model prescribed. Other survey and interview data suggest that evaluators found time to do the evaluations, but that scaling back some requirements may be warranted. Specifically:

- While 51% of teachers felt their evaluators had the knowledge to evaluate them accurately, only 17% of teachers felt that their supervisor had the time and resources to implement SEED.

Outcomes

We also examined short-term outcomes of SEED. Given that 2012-13 was a pilot year and that many of the resources to support SEED were being developed as the model was implemented for the first time, we would not expect to see large positive outcomes. At this stage of implementation, it is reasonable to expect modest changes in practice.

Consistent with this expectation, teachers expressed mixed views on whether SLOs changed their practice, but were more positive about the influence of observations on their teaching. Administrators reported modest changes as a result of administrator SEED.

- 44% of teachers surveyed agreed that feedback from observations prompted them to change their practice
- 55% of administrators surveyed indicated that setting SLOs led teachers to make changes in their teaching practice
- 42% of teachers surveyed felt that with sufficient resources, such as time and staffing, SEED could improve teacher practice at their school
- 74% of administrators surveyed felt that with sufficient resources, such as time and staffing, SEED could improve teacher practice at their school

Variations

We found variations by school level, teachers' tenure status, and district. Elementary teachers and non-tenured teachers reported that they had received a higher number of observations than their secondary school and tenured counterparts. They also reported SEED to be more valuable and have greater potential benefits than did their secondary school and tenured counterparts. Districts with lower student performance and multiple initiatives underway reported less robust implementation and less favorable attitudes towards SEED's current or potential impact on practice.

Recommendations

Based on these findings, we recommend that the State Department of Education carry out the following:

1. **Provide additional opportunities for all educators to learn about SEED.** There is a need for additional professional learning opportunities for all educators with respect to SEED. Professional learning needs fall into two categories: one is better understanding and implementing the technical aspects of SEED (setting goals, conducting observations, and providing post-observation feedback) and the other is improved development of educators (by providing individualized and targeted professional development). We strongly recommend that both administrators *and* teachers receive this training, rather than relying on a train-the-trainer model as was the case in the pilot year of SEED.

2. **Build the skills of evaluators, in particular.** Teacher survey and interview data indicated substantial variability in the perceived skill level of evaluators. This variance occurred within and across districts, and even within schools. Such a finding points to the need for administrators to develop and refine their supervisory skills. We suggest there be processes for identifying evaluators in need of improvement and then offering specialized training to these individuals. The SDE, along with professional organizations and regional consortia, should provide professional development to evaluators in using the CCT Rubric for Effective Teaching, conducting formal and informal observations, and providing verbal and written feedback. These professional learning opportunities are critical to the success of any teacher evaluation reform.

3. **Increase the use of complementary observers.** Educators reported that SEED places significant time demands on school administrators. We recommend that districts consider including complementary observers within their teacher evaluation systems. Including complementary evaluators not only reduces the time demands on principals and assistant principals, but also enhances the professional role of teachers by providing additional leadership roles for teachers. Under the moniker of peer assistance and review, such systems have been adopted by several high-profile

districts in other states (e.g. Cincinnati, Montgomery County, MD) and have shown promising outcomes in terms of teacher performance and student learning. The SDE can help by sharing effective models with those in the field and providing the required training for complementary observers.

4. **Provide additional guidance on Student Learning Objectives and Indicators of Academic Growth and Development.** As SEED has begun in earnest, we recommend that the SDE continue to provide clear guidance on the identification of valid IAGD targets. The SMART goals heuristic is helpful although does not dictate what performance level or student growth target is both "attainable" and sufficiently challenging. Some teachers or administrators are selecting far too challenging targets while others are choosing far too easy. The setting of IAGD growth targets is in most cases inherently arbitrary. Should 100% of students score a 70% on an exam or should 70% of students score at 100%? If half the students fall below a certain performance level at the beginning of the year, what percentage should reasonably be expected to meet it by the end of the year? Further, what is meant by growth differs based on the measures available (e.g., contrast pre and post measures from a vertically-scaled assessment with static measures of performance on a locally developed test). There are checks and balances built into the system, such as the mid-year check-ins, that are intended to provide at least one opportunity to adjust growth goals. However, the selection of IAGD targets is an area that deserves close attention as the SEED model evolves.

5. **Clarify and contextualize SEED to district and school personnel.** Teachers in pilot districts expressed trepidation over the number and magnitude of new reform initiatives they faced, and viewed teacher evaluation as adding "one more thing" to their already busy agenda. It would be helpful if educational leaders and policy makers demonstrated how educator evaluation interfaces with the implementation of other major initiatives such as the Common Core State Standards and Smarter Balanced Assessments. Moreover, make clear to districts any flexibility they have with regard to implementing SEED or the Core Requirements. Administrators are interpreting SEED with more rigidity than the model actually requires.

6. **Disseminate promising practices.** We recommend a coordinated effort by SDE and/or regional district consortia to identify and disseminate "promising practices" relative to SEED and educator evaluation. Connecticut educators represent an extraordinarily rich source of ingenuity and talent and should be actively engaged in the continuous improvement of evaluation as they adapt the model to their own district contexts and explore new processes. In some settings we found evidence of schools using SEED to promote deep conversations about teaching and learning. For example, one school piloted the use of video cameras to facilitate the evaluation of teachers. Teachers reported that this approach reduced their anxiety to "perform" in front a live observer and facilitated conversations with their evaluator by helping

them understand the evaluator's feedback.

7. **Provide additional assistance to low-performing districts to support SEED's implementation.** We believe that the state should provide additional assistance to low-performing districts to support the implementation of SEED. This recommendation is supported by our finding that lower-performing districts with multiple initiatives underway encountered particular challenges in implementing SEED. Providing additional support and assistance to these districts may help them integrate SEED with other initiatives and increase the chance that the reform spurs positive change in their schools and classrooms.

8. **Continue to track SEED's implementation and effects.** We recommend that the state continue to gather data from educators at regular intervals to inform the continuous improvement of the model. In addition, we recommend tracking the model's utility in rating the annual performance of educators and administrators. Specifically, we recommend that the relationship between SEED and student achievement be examined.

Introduction

In June 2012, the Connecticut General Assembly passed legislation mandating that the System for Educator Evaluation and Development (SEED) be piloted in the 2012-13 school year. The Neag School of Education was directed to study the enactment of SEED or the Core Requirements in the pilot school districts and report findings to the State Board of Education and Education Committee to inform future implementation of SEED. The purpose of our study was to examine the implementation and short-term outcomes of Connecticut's new educator evaluation model in the 10 pilot school districts/consortia (14 districts in total). Our broader aim was to provide feedback on SEED's pilot implementation so that the model could be improved and better support improvements to teachers' practice and students' learning.

Between fall, 2012, and fall, 2013, we collected quantitative and qualitative data in all 14 pilot districts. Data addressed SEED's implementation; educators' experiences with SEED; and key outcomes of SEED's pilot year.

Based on these data, we find that SEED was implemented in the pilot sites with a relatively high degree of fidelity to the model. Moreover, most educators in the pilot sites supported the model in the abstract and found several aspects of SEED (e.g., post-observation conferences with evaluators) quite valuable. There is some evidence of changed practices on the part of teachers and administrators due to SEED. While SEED's components were generally implemented with fidelity, opportunities for professional growth through SEED were not fully realized. To a great extent, these challenges were due to the lack of clear and consistent communication to educators about SEED and inadequate opportunities for educators to construct a robust understanding of this new approach to evaluation. Educators attempted to understand the complexities of the technical and practical aspects of SEED. However, the tight implementation timeline in some cases encouraged compliance behaviors rather than full engagement with the model.

This emphasis on compliance is not uncommon during the initial stages of a major reform initiative as participants and implementers adapt to a new set of rules and routines. We expect that districts will begin to experience greater benefits of SEED as educators become accustomed to its components and adapt SEED to their own contexts. Our data suggest that, if sufficiently supported, SEED has good potential to improve teachers' practice and students' learning.

The report is organized as follows. Section II presents the methods employed for this study. Section III presents the findings and is organized into three sub-sections: one on implementation, one on educators' experiences with SEED, and the final section devoted to outcomes of SEED for its pilot year. We conclude with section IV, which presents recommendations based on our analyses of the data. This report represents

the final deliverable as requested by Public Act No. 12-116. During our year of data collection and analysis, oral presentations were made to the Performance Evaluation Advisory Council on February 4, 2013 and July 10, 2013. These interim reports were designed to provide state policy makers with timely feedback on the first two phases of the evaluation cycle.

Methods

Our study followed a mixed methods research design. Between September 2012 and October 2013, we collected qualitative and quantitative data describing the implementation of SEED. We examined three sets of research questions related to policy implementation, educator experiences, and policy outcomes:

SEED Policy Implementation

1. How was SEED enacted in each pilot site? Was there fidelity to the State Model?
2. Did SEED create opportunities for professional growth for teachers and administrators? To what extent were these differentiated across individuals or subgroups?
3. What variations occurred? What explains these variations?

Educator Experiences with SEED

4. How did educators in pilot sites experience SEED?
5. What variations occurred? What explains these variations?

SEED Policy Outcomes

6. To what extent did educators report changing their practices as a result of the SEED?
7. To what extent was there variation in evaluation ratings within schools and districts at the conclusion of the SEED pilot year?
8. What variations occurred? What explains these variations?

Sample

District Sample

Fourteen public school districts (eight stand-alone districts and two multi-district consortia) were chosen by the State Department of Education (SDE) to serve as pilot sites. Participating districts represent a range of settings from across the state (see Tables 1 and 2 for district characteristics). CSDE selected the 14 pilot districts from a larger group of districts that volunteered to take part in the pilot program. These 14 districts employed 3,406.5 full time equivalent (FTE) general education teachers in 2010-2011, which was nine percent of the FTE general education positions in the state. These districts were responsible for educating 52,401 students that same year, which is also nine percent of the state's enrollment.

Table 1. Student and Teacher Demographics for Pilot Districts

District	# of Schools	# of Students	% FRPL	% Special Ed	% ELL	# FTE Gen Ed Teachers	Mean Gen Ed Salary
Bethany	1	491	4.5%	12.5%	< 1.0%	37.9	$54,271
Branford	5	3,363	18.8%	13.4%	3.8%	244.3	$67,956
Bridgeport	32	20,087	98.8%	12.4%	13.0%	1,133.3	$64,787
CEFS							
Columbia	1	539	10.9%	7.9%	< 1.0%	41.8	$66,383
Eastford	1	178	9.6%	12.8%	-	13.9	$58,244
Franklin	1	222	13.1%	10.9%	-	15.5	$62,929
Sterling	1	482	32.0%	10.5%	-	32.0	$54,472
CREC	12	4,241	43.1%	10.5%	2.2%	323.5	$65,349
Litchfield/ Region 6							
Litchfield	3	1,169	9.4%	10.0%	< 1.0%	85.9	$67,578
Region 6	4	1,042	12.1%	11.0%	1.2%	78.8	$65,110
Norwalk	19	11,050	43.2%	10.1%	12.3%	711.2	$80,660
Waterford	5	2,800	12.4%	11.2%	1.2%	192.6	$74,198
Windham	6	3,124	75.5%	17.4%	26.7%	221.3	$61,825
Windsor	7	3,613	28.1%	12.5%	2.7%	274.5	$61,827
State	*1.017*	*570,494*	*34.4%*	*11.6%*	*5.5%*	*35,977.4*	*$67,878*

Data from the State Department of Education, 2010-2011

Table 2. Student Achievement in Pilot Districts

District	% Students At Goal in Grade 3 CMT - Reading	% Students At Goal in Grade 3 CMT - Writing	% Students At Goal in Grade 3 CMT - Math
Bethany	74.6%	66.2%	66.2%
Branford	66.7%	62.6%	74.6%
Bridgeport	22.5%	29.7%	27.5%
CEFS			
Columbia	64.9%	65.5%	72.4%
Eastford	-	-	-
Franklin	-	-	-
Sterling	52.9%	60.8%	62.7%
CREC	70.9%	65.1%	62.0%
Litchfield/ Region 6			
Litchfield	64.6%	67.4%	68.7%
Region 6	65.4%	50.0%	61.7%
Norwalk	51.9%	56.4%	62.5%
Waterford	70.9%	77.9%	70.0%
Windham	36.6%	34.0%	35.9%
Windsor	46.8%	43.7%	52.8%
State	*58.4%*	*61.1%*	*63.3%*

Data from the State Department of Education, 2010-2011

School Sample

In each of the pilot districts, we purposively selected a sample of schools in which to conduct focus groups and interviews. We selected a minimum of 20% of the schools at each level (high school, middle school, elementary school) based on characteristics that might influence policy implementation, such as the student population served, leadership turnover, and initiatives or programs of study (e.g., dual language program, alternative education program). In small districts with only one school at each level, we included all of that district's schools in the sample. In the consortia sites, we included schools from each district. In total, 41 schools participated in the focus groups and interviews for this study, representing 42% of the schools in the pilot districts.

All schools in the pilot districts were invited to participate in the two surveys; further detail is provided below.

Interview and Focus Group Sample

Superintendents from each pilot district and, in some cases, other central office personnel, participated in interviews.

We selected educators within each sample school to participate in interviews and focus groups. We included all principals and, in some schools, also an assistant principal or Dean of Students, from each school in our sample. In total, 37 school leaders participated in the study (see Tables 3-5 for interview and focus group participants by district/consortium for each round of data collection).

In addition to school leaders, we sampled educators who are evaluated under the teacher portion of SEED. For this group, we selected participants to represent a range of grade levels, subjects, and roles within the school. We deliberately included other personnel, such as school psychologists and speech pathologists, who are evaluated under SEED. We randomly selected teachers and specialists to represent 20% of each school's faculty. We weighted these random samples to reflect the proportions of core academic teachers (i.e. teachers of mathematics, science, English Language Arts, and social sciences), related arts teachers (e.g., arts, physical education, world languages), and specialists (e.g., special education teacher, school counselor) present in the school. Samples were drawn by researchers; school or district leaders were not involved in sample selection (see Tables 3-5 for interview and focus group samples by phase of data collection).

Table 3. Interview and Focus Group Sample in Phase 1 by Pilot District/Consortium

	Number of Interview/Focus Group Participants in Phase 1					
District/Consortium	Schools*	District Leaders	School Leaders	Teachers	Specialists	*Total*
Bethany	1	0	2	10	2	*15*
Branford	2	2	0	11	3	*18*
Bridgeport	7	1	7	73	9	*97*
CEFS[1]	4	4	5	41	3	*57*
CREC[2]	6	1	6	31	2	*46*
Litchfield/Region 6	4	2	5	13	8	*32*
Norwalk	3	1	3	36	5	*48*
Waterford	2	2	3	27	3	*37*
Windham	4	2	4	26	23	*59*
Windsor	2	2	2	28	9	*43*
Total	*35*	*17*	*37*	*296*	*67*	*452*

*Schools in target sample; [1] Columbia-Eastford-Franklin-Sterling; [2] Capitol Region Education Council

Table 4. Interview and Focus Group Sample in Phase 2 by Pilot District/Consortium

	Number of Interview/Focus Group Participants in Phase 2					
District/Consortium	Schools*	District Leaders	School Leaders	Teachers	Specialists	*Total*
Bethany	1	0	2	6	0	*8*
Branford	3	0	0	10	2	*12*
Bridgeport	3	2	1	25	8	*36*
CEFS[1]	4	2	3	34	0	*39*
CREC[2]	5	1	4	29	2	*36*
Litchfield/Region 6	5	0	3	19	2	*24*
Norwalk	1	2	0	1	0	*3*
Waterford	0	2	0	0	0	*2*
Windham	4	2	4	31	11	*48*
Windsor	1	0	1	0	0	*1*
Total	*27*	*11*	*18*	*155*	*25*	*209*

*Schools in target sample; [1] Columbia-Eastford-Franklin-Sterling; [2] Capitol Region Education Council

Table 5. Interview and Focus Group Sample in Phase 3 by Pilot District/Consortium

	Number of Interview/Focus Group Participants in Phase 3					
District/Consortium	Schools*	District Leaders	School Leaders	Teachers	Specialists	*Total*
Bethany	1		2	4	2	*8*
Branford	3	2		10	1	*13*
Bridgeport	6		1	42	15	*58*
CEFS[1]	4	2	1	29		*32*
CREC[2]	6	1		34	1	*36*
Litchfield/Region 6	3	2	3	19	6	*30*
Norwalk	5	2	3	56	2	*63*
Waterford	3	1	2	29	3	*35*
Windham	4	2	5	24	15	*46*
Windsor	2	2	7	22	5	*36*
Total	*37*	*14*	*24*	*269*	*50*	*357*

*Schools in target sample; [1] Columbia-Eastford-Franklin-Sterling; [2] Capitol Region Education Council

We also interviewed union presidents in all but one pilot district and state-level union representatives from CEA and AFT-Connecticut (n=16). Lastly, we interviewed trainers from the Regional Educational Support Councils (RESCs) involved with SEED trainings and roll-out. These participants provided additional information about the context of SEED design and implementation and provided a source for triangulation of data provided by district and school leaders and teachers.

Survey Samples

We invited all educators within the sample schools to participate in surveys during winter/spring, 2012, and again in fall, 2013. Between 17% and 81% of teachers within each sample school responded to the first survey, yielding a total number of 684 participants; between 13% and 53% of teachers within the sample schools responded to the second survey, yielding 533 participants[1] (see Tables 6-7 for survey samples and response rates and Table 8 for characteristics of teachers in survey samples).

[1] It was important to provide anonymity to educators so they could feel comfortable being honest in their survey responses. Therefore we did not ask participants to identify themselves on the surveys and cannot determine what percentage of educators responded to both surveys.

Table 6. Survey Sample and Response Rates for First Survey

District/Consortium	Schools*	Sample		School Population**		District Population***	
	n	n	%	n	%	n	%
Bethany	1	16	2%	67	24%	67	24%
CEFS	3	69	10%	120	58%	130	53%
CREC	6	142	21%	175	81%	516	28%
Litchfield/Region 6	5	80	12%	177	45%	209	38%
Norwalk	3	58	9%	342	17%	892	7%
Waterford	3	84	12%	160	53%	244	34%
Windham	3	138	20%	217	64%	302	46%
Windsor	1	97	14%	288	34%	387	25%
Total	*25*	*684*	*100%*	*1,546*	*44%*	*2,747*	*25%*

*Bridgeport did not participate in the survey dissemination

**Percentage of target sample

***Percentage of entire district/consortia; not adjusted for reduced sample of schools

Table 7. Survey Sample and Response Rates for Second Survey

District/Consortium	Schools*	Sample		School Population**		District Population***	
	n	*n*	*%*	*n*	*%*	*n*	*%*
Branford	3	68	13%	215	32%	309	22%
Bridgeport	4	91	17%	373	24%	1,607	6%
CEFS	2	16	3%	66	24%	71	23%
CREC	5	92	17%	175	53%	516	18%
Litchfield/Region 6	6	34	6%	177	19%	209	16%
Norwalk	6	47	9%	342	14%	892	5%
Waterford	5	59	11%	160	37%	244	24%
Windham	6	90	17%	217	41%	302	30%
Windsor	1	36	7%	288	13%	387	9%
Total	*38*	*533*	*100%*	*2,067*	*26%*	*4,596*	*12%*

*Bethany, Franklin, and Sterling did not participate in the second survey

**Percentage of target sample

***Percentage of entire district/consortia; not adjusted for reduced sample of schools

Of the 683 teachers who responded to the first survey, 25% (n=173) were not tenured, 56% (n=383) were tenured, and 19% (n=127) declined to indicate tenure status. Twenty-five percent (n=167) taught at the elementary school level, 15% (n=101) taught at the middle school level, and 39% (n=263) taught at the high school level; 13% (n=92) declined to indicate grade level. There were similar patterns of representation in the sample for the second survey. Of the 533 teachers who responded to the second survey, 15% (n=82) were not tenured, 67% (n=356) were tenured, and 18% (n=95) declined to indicate their tenure status. A third (33%, n=178) of respondents teach at the

elementary school level, 19% (n=101) teacher at the middle school level, and 27% (n=143) teach at the high school level; 21% (n=111) declined to indicate their grade level.

Table 8. Characteristics of Teachers in Survey Samples

	Mid-Year Survey		End of Year Survey	
	n	*%*	*n*	*%*
Experience Teaching				
1-4 years	202	30%	46	9%
5-10 years	164	24%	100	19%
11-15 years	83	12%	94	18%
16-20 years	41	6%	80	15%
> 20 years	55	8%	125	24%
Declined to Answer	138	20%	88	17%
Highest Degree				
Bachelor's	5	1%	20	4%
Masters of Education	317	46%	256	48%
Other Masters	198	29%	159	30%
Doctorate	20	3%	11	3%
Declined to Answer	143	21%	93	17%
Race/Ethnicity				
Black/African	17	3%	16	3%
Asian/Pacific Islander	12	2%	8	2%
Hispanic/Latino	18	3%	25	5%
Native American	3	< 1%	2	< 1%
White	457	67%	369	69%
Declined to Answer	176	24%	118	22%
Gender				
Female	431	63%	336	63%
Male	49	7%	76	14%
Declined to Answer	203	30%	121	23%

We also administered one survey to school administrators in fall, 2013. Twenty-two administrators from eight districts and 14 schools completed the survey. The majority of respondents were principals (55%), with 23% identifying as assistant principals and the remaining 23% declining to indicate. Fifty percent of respondents reported having fewer than 5 years of experience as an administrator.

Table 9. Respondent Demographics for Administrator Survey (n=22)

	n	*%*
Role		
Principal	12	55%
Assistant Principal	5	23%
Declined to Indicate	5	23%
Years of Experience		
1 – 4 Years	11	50%
5 – 10 Years	5	23%
11 – 15 Years	2	9%
16 – 20 Years	0	0%
More than 20 Years	0	0%
Declined to Indicate	2	10%
Years at Current School		
1 – 4 Years	6	27%
5 – 10 Years	4	18%
11 – 15 Years	5	23%
16 – 20 Years	1	5%
More than 20 Years	2	9%
Declined to Indicate	2	9%

Data Collection

We collected data in four phases.

Phase 1: Between October and December 2012, we interviewed superintendents (and other district leaders, in some cases), principals, district union presidents and state-level union representatives. We also conducted focus groups with teachers and student support personnel in sample schools.

Phase 2: Between January and mid-April 2013, we conducted a second round of interviews and focus groups with principals, teachers, and student support personnel. We also administered surveys to teachers and student support personnel in the sample schools.

Phase 3: In May and June 2013, we conducted a final round of interviews and focus groups with superintendents, principals, teachers, and student support personnel. Follow-up interviews with union personnel were also conducted.

Phase 4: In September 2013, we administered a second survey to teachers and student support personnel.

Data Analysis

Data were analyzed using standard quantitative and qualitative analytic techniques. For survey analysis, we computed descriptive statistics and performed t-tests to discern differences between groups. For interview and focus group data, we completed thematic summaries for many interviews and completed cross-case analyses of the data.

Limitations

As with all research, this study has limitations. The main weakness of this study's design is its reliance on self-reported data obtained through interviews and surveys. We attempted to address this weakness in the design through triangulation of data sources and findings. Specifically, we gathered data on similar topics from multiple stakeholders (e.g. teachers and principals), through varied data collection instruments (i.e. interview and survey), and at multiple points during SEED's implementation. Through these varied forms of data, we attempted to reduce the bias that can arise through self-report.

Note on Terminology

Throughout this report we use the general term "teachers" to refer to those individuals evaluated under SEED's teacher portion. Thus, "specialists" such as school psychologists are included under the umbrella term of "teacher." When the data indicated differences between classroom teachers and specialists, we note that by explicitly referring to the experiences of "specialists."

Findings on the Implementation of SEED

Consistent with our charge, we examined the implementation of SEED within the 14 pilot districts. In this section, we discuss our findings related to the implementation of SEED for teachers and school administrators. We present data that responds to the following questions:

1. How was SEED enacted in each pilot site? Was there fidelity to the State Model?
2. Did SEED create opportunities for professional growth for teachers and administrators? To what extent were these differentiated across individuals or subgroups?
3. What variations occurred? What explains these variations?

In general, we found that schools and districts in the pilot program implemented all of the components of the SEED model. However, in many cases districts did not fully implement each component of the model or implemented components in ways that did not capitalize on SEED's potential to improve teachers' or leaders' practices. This finding is consistent with research on the first year of other new initiatives in education or other sectors.[2] In addition, many schools focused on basic compliance with the SEED policy rather than using SEED to drive improvements in educators' practices. This focus on compliance is to be expected given the pilot nature of the implementation in 2012-13. Districts' timeline for implementing SEED was fast and the development of resources to support SEED sometimes lagged implementation in pilot districts. We further found that schools with the greatest challenges (i.e., very low student achievement and multiple reform initiatives) seemed to have the most difficulties in using the model to drive improvements in practice.

In the sections below, we first discuss the fidelity of implementation of various components of SEED, including observations, post-observation conferences, professional growth opportunities, and the administrator-evaluation portion of SEED. We then turn to variations in implementation.

Fidelity to SEED

Assessing fidelity of implementation is important because it can help surface areas of the model that are particularly challenging, unclear, or working well. Fidelity data are also helpful as a backdrop when judging the short-term outcomes of the model. Our

[2] See, for example, the Concerns-Based Adoption Model, which posits a predictable series of behavior among individuals faced with significant organizational change (Hord, Rutherford, Huling-Austin, & Hall, 1987). A cogent summary can be found on the National Academy of Sciences website at http://www.nas.edu/rise/backg4a.htm.

purpose was not to “catch” districts not adhering to the model but rather to assess the efficacy of the model among pilots and to inform our own work as evaluators.

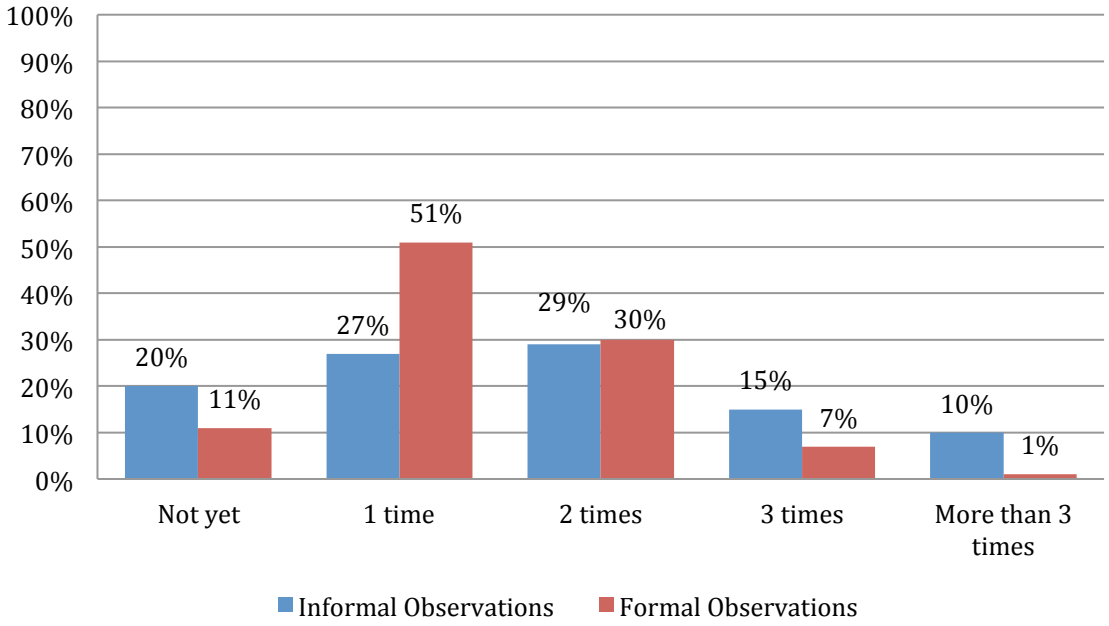
SEED requires that teachers be assessed based on the degree to which they attain their Student Learning Objectives (SLOs); their performance on three formal and three informal observations; parent feedback; and whole-school measures of student learning or student feedback. We assessed the extent to which SEED was implemented with fidelity by gathering data on the degree to which the following components of the model were enacted: Student Learning Objective (goal) setting, classroom observations, and conferences between educators and their evaluators, including the mid-year and summative conference. We further ask whether professional growth opportunities were made available to teachers in connection with SEED and whether they were assessed based on parent feedback and whole-school measures of student learning or student feedback. At the conclusion of this section, we discuss the fidelity of implementation of SEED’s school administrator portion.

Observations

A major component of SEED is classroom observations, guided by a rubric specifying different gradations of quality in instructional practice. SEED prescribed three formal and three informal observations for all teachers in the pilot districts. We found that principals encountered substantial challenges to completing the prescribed number of observations. Overall, we found that while a sizeable portion of the teachers reported that they had received the prescribed three observations in each category, the majority reported they had received fewer.

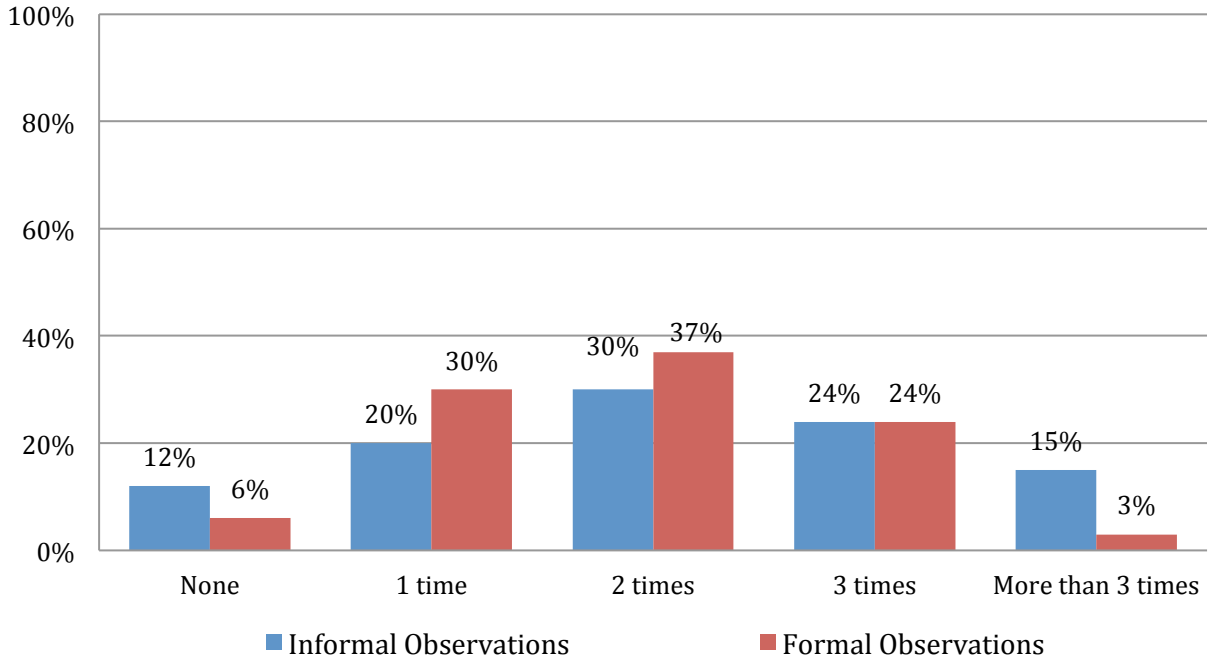
In surveys administered between February and April, 2013, 25% of teachers reported that they had been informally observed at least three times and 8% reported that they had been formally observed three or more times (see Figure 1).

Figure 1. Number of Informal and Formal Observations (Spring 2013 Survey, *n*=684)



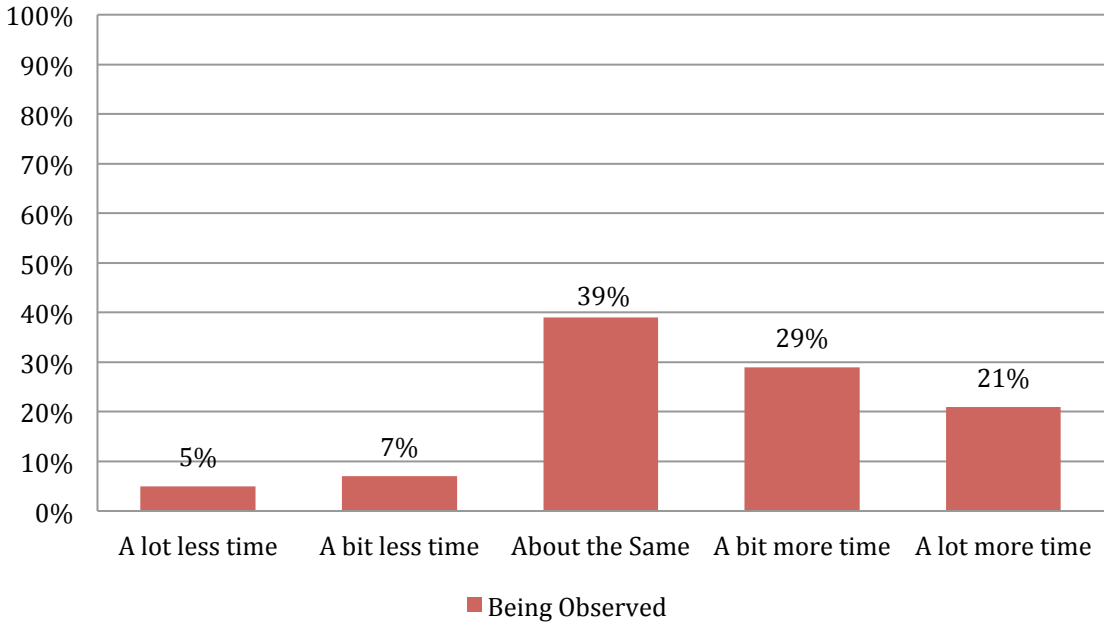
As would be expected, surveys administered in September 2013 demonstrated a higher number of observations completed. Thirty-nine percent of teachers reported that they had been observed informally at least three times with 27% reporting that they had been observed formally this often (Figure 2). The fact that 69% of teachers reported that they had received two or more informal observations and 64% reported that they had received two or more formal observations suggests that the requirement of two observations in each category (i.e., four observations in total) appears more feasible in a baseline year than the three of each observation mandated by the 2012-13 version of SEED.

Figure 2. Number of Informal and Formal Observations (Fall 2013 survey, *n*=533)



In interviews conducted in winter and spring, 2013, principals reported that they were doing more observations than in previous years. A considerable number of teachers reported that they received more observations under SEED than under their district's prior evaluation system. When surveyed, 21% of teachers reported that they spent a lot more time being observed this year (under SEED), compared to the prior year (under the prior evaluation system); another 29% reported that they spent a bit more time being observed in 2012-13 (fall survey, 2013). Thus, 50% of teachers surveyed reported being observed more under SEED than under their district's previous system (Figure 3).

Figure 3. Time Spent Being Observed Compared to Pre-SEED (Fall survey 2013, *n*=533)



School Administrators' Views

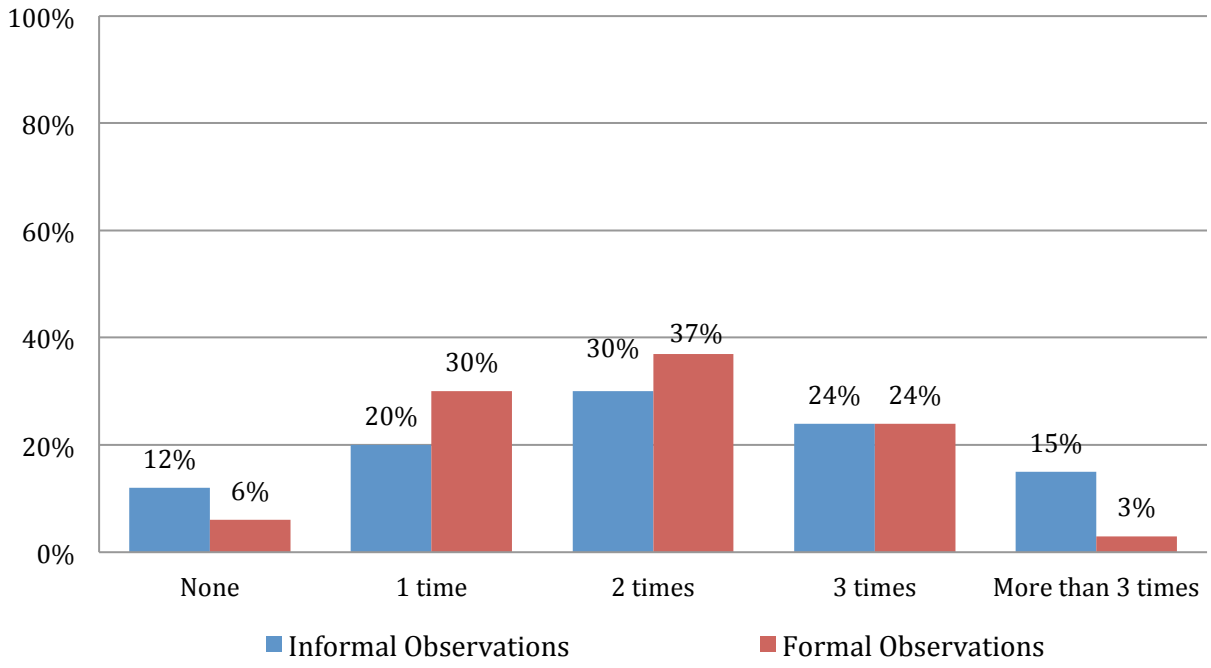
In fall 2013, we surveyed school administrators regarding the implementation of SEED for teachers. The sample size was relatively low (*n*=22), so we caution readers against putting too much faith in the exact estimates presented here. However, all statistics reported here are consistent with our qualitative findings, which increases our confidence in the validity of the survey findings despite the small sample size.

Administrators reported that evaluation loads ranged widely, from one elementary school assistant principal responsible for evaluating six teachers to another evaluator with a load of 56 teachers. The average evaluation load reported by surveyed administrators was 25 teachers, which was situated within a 95% confidence interval of between 19 to 31 teachers. A small subset of the administrators we surveyed (16%) were solely responsible for all teacher evaluations within their schools; approximately half (53%) of the administrators reported two or three evaluators in their schools and a third (32%) reported four or five evaluators within their schools. As discussed below, at least three schools used complementary observers to reduce administrators' evaluation loads.

Turning to number of observations reported, 27% of administrators indicated that they completed three or more formal observations per teacher and 39% indicated that they completed three or more informal observations per teacher (see Figure 4). These findings align with teacher reports. Approximately a third of administrators reported approaching the required number of observations expected in SEED, with 37% reporting

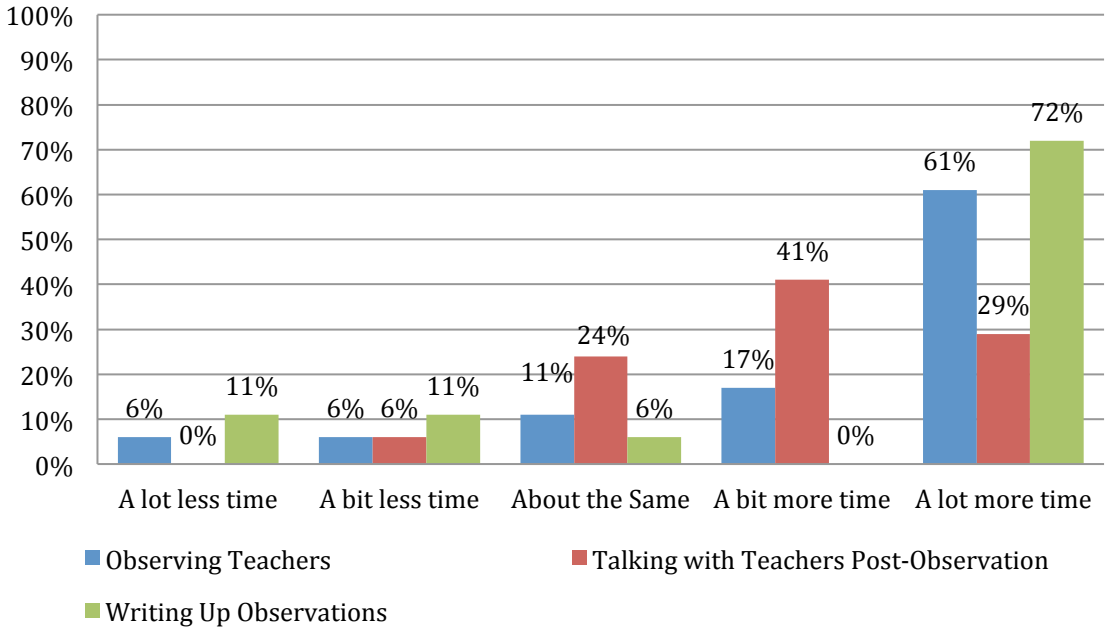
two formal observations and 30% reporting two informal observations; again, these findings align with teacher reports.

Figure 4. Distribution of Observations Reported by Administrators (*n*=22)



Consistent with these statistics, administrators reported spending considerably more time on observations than under previous evaluation systems. Seventy-eight percent of administrators surveyed reported that they had spent more time observing teachers under SEED than under their previous evaluation systems (Figure 5). Moreover, 72% of respondents indicated that they spent a lot more time writing up observations than they had done previously.

Figure 5. Administrator Time on Observation Tasks (*n*=22)



In interviews, almost all principals reported struggling to complete the required number of observations. Principals in many pilot sites stated that they devoted substantial time on a near-daily basis attending to observations and the corresponding pre- and post-observation conferences, and documenting the process. In response to principal feedback, three districts officially reduced the number of observations that they required principals to perform. Principals with larger evaluation loads were especially challenged in their efforts to complete observations. This was compounded in schools where principals were coordinating multiple initiatives in addition to SEED.

Consistent with these reports, at the end of the school year, teachers and leaders reported an increased frequency of observations as leaders attempted to complete the prescribed number of observations. Some teachers reported receiving two to three observations in the last 3-4 weeks of school. Asked in late May if he had completed the required number of observations, one principal said, “not as many as I’d like. It’s a one-man show here.” Another elementary principal responded: “for the most part.”

Some districts brought in additional observers to help principals complete them. Administrators in several schools, located in Bridgeport, CREC, and Windham, reported using complementary evaluators in addition to evaluators in traditional school leadership roles. The sites in CREC and Windham used master teachers as peer observers and the Bridgeport site used central office administrators, coaches, and lead teachers as complementary evaluators.

Windham also drew on district-level administrators, particularly in schools with few administrators. One principal said this had ancillary benefits: it “saved me time and

helped me calibrate" and argued, "I like the idea of the complementary evaluators--it works."

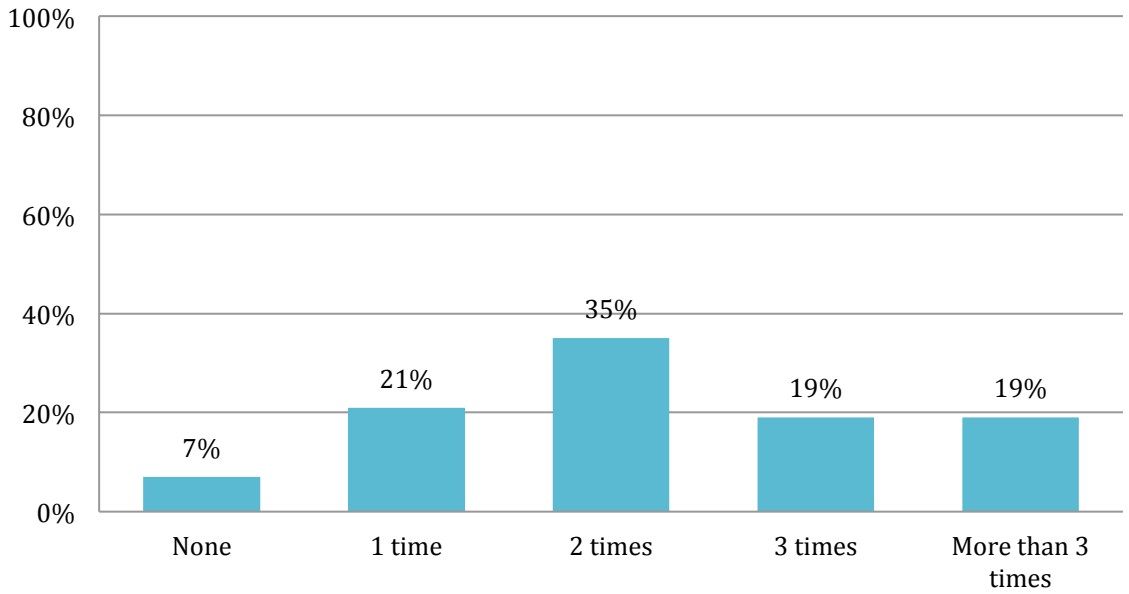
Several elementary principals reported that they had completed the prescribed number of observations but that this had greatly increased their workload. They reported that they were working 7-day weeks to complete their observations, whereas in previous years they could complete their work in 5-6 days per week.

On the whole, these data suggest that the prescribed three formal and three informal observations are hard to enact in a baseline year where every educator is evaluated. Two of each type of observation (four total) seems more attainable. Requiring a minimum of two types of each observation would encourage administrators to conduct regular observations while also protecting time so that they can maximize the value of each observation. The Core Requirements also allow for differentiation on the number of observations based on teacher experience, prior ratings, and needs and goals. SEED also affords administrators some discretion on this front.

Feedback and Conversations related to Observations

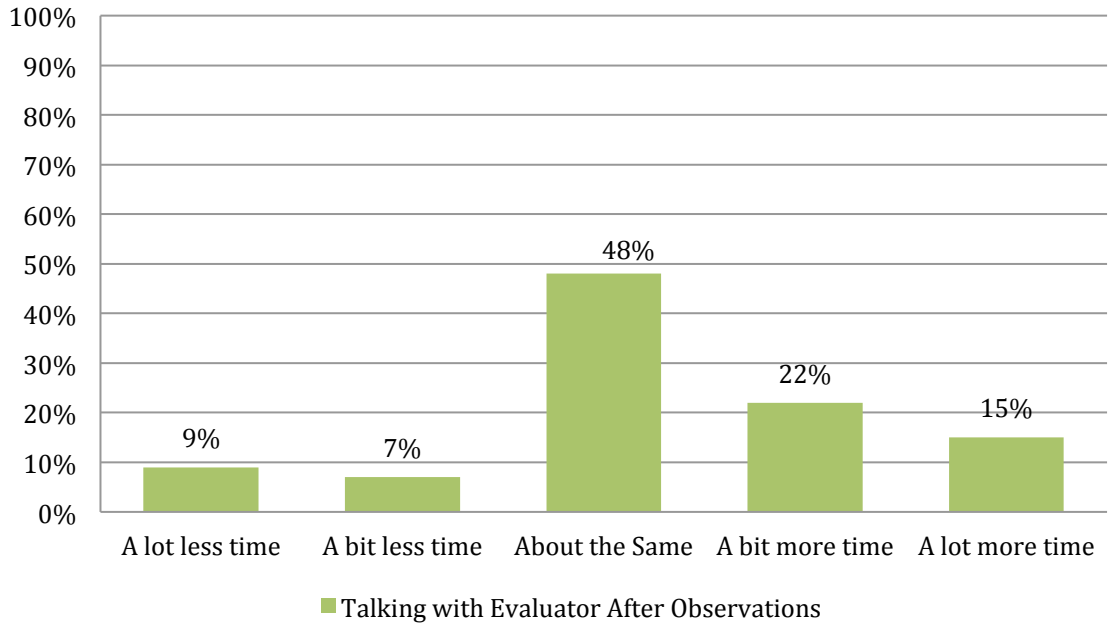
In our view, the most powerful aspect of the SEED model is its potential to improve practice through evidence-based conversations between leaders and teachers about teachers' practice. Seventy-three percent of teachers surveyed reported that they had had two or more post-observation feedback sessions (Figure 6). Seven percent said that they had no such sessions.

Figure 6. Frequency of Post-Observation Conferences (fall survey 2013, *n*=533)



More than a third (37%) of teachers surveyed in fall 2013 reported that they spent more time in post-observation conferences than they had in previous evaluation systems (see Figure 7). The plurality (48%) reported that the time spent in this activity was similar to that in the past.

Figure 7: Time Spent In Post-Observation Conferences (Fall Survey 2013, *n*=533)



As noted above, 70% of administrators reported spending more time in post-observation conferences with teachers than in years prior to SEED's implementation. The discrepancy between teachers and administrators regarding whether they had spent more time in post-observation conferences likely reflects the different experiences of these different educators. In prior years, tenured teachers were not evaluated annually. Under SEED, administrators needed to observe and conference with all teachers, which would generally result in an increase in overall time spent in conferences compared to prior years. Teachers may have interpreted this question to mean *compared to the years in which they underwent observation*; the time spent in post-observation conferences, at an individual level, might thus be higher for a lower proportion of teachers than administrators.

Although conferences were generally occurring and a sizeable minority of teachers and administrators spent more time in conferences than under previous evaluation systems, interviews with principals and teachers suggested that such meetings could be more substantive and meaningful. This comes as little surprise, given that this was a pilot year in which administrators were getting used to the system and focused on executing its parts. For example, principals reported that they had prioritized conducting observations over completing post-observation conferences. They said that sometimes they were unable to hold these meetings or, if they met with teachers, to make these

meetings as robust as they would have wanted. Moreover, teachers were asked if they were receiving and using feedback in a variety of ways.

- Overall, almost half (43%) of the teachers surveyed agreed or strongly agreed that they were receiving and using feedback under SEED (spring survey, 2013)
- Significantly fewer tenured teachers reported receiving and using feedback than non-tenured teachers, with $t(359) = 6.16$, $p < .001$. Only 35% of tenured teachers agreed or strongly agreed that they were receiving and using feedback via SEED, compared with 60% of non-tenured teachers (spring survey, 2013)

The difference between tenured and non-tenured teachers' responses was often an artifact of school leaders prioritizing observations and post-observation conferences with non-tenured teachers. Leaders reported that they felt that non-tenured teachers needed more assistance and feedback.

Our interviews with teachers and leaders suggested that debriefing conversations were often late and relatively brief (about 15-25 minutes long). Some debrief "conversations" were conducted through *My Learning Plan* or via email. In one school, no debrief conversations were reported. In others, the quality of debriefs varied across multiple evaluators within the same school.

Principals reported that they had made adaptations to SEED regarding the post-observation conference. Said one high school principal: "it's easy to get in [to classes]. The hard piece is to schedule the pre and post conference." The principal further explained that his strategy for post-observation conferences had become to "mostly catch teachers," "bump into them." Some principals stopped doing conferences with veterans whose observations indicated they were doing well and chose to spend their time instead with struggling veterans and more junior teachers with more areas to improve. Some principals held summative meetings in a group with teachers. One principal observed that it has been "a scheduling nightmare to get it all done." Another principal wished that specific (and single) domains within the rubric could be targeted during observations with early career teachers, thus affording the opportunity to achieve greater depth in one area of improvement rather than spread feedback over a wide array of domains, potentially diluting its impact. It is worth noting that several principals who were new to their building reported using the pre- and post-conferences as a way to get to know their faculty.

Mid-year Check-ins

We also gathered data specifically focused on mid-year check-ins. All districts enacted mid-year check-ins in the context of SEED. In the fall 2013 survey, the vast majority--86%--of teachers indicated that they had had a mid-year conference. However, 14% of teachers reported that they had not had these meetings, and interview data revealed that some mid-year check-ins happened quite late, in March or April.

Regarding the substance of mid-year check-ins, teachers reported that mid-year check-ins were generally brief discussions that did not delve deeply into teachers' instructional practice. According to teachers, most mid-year check-ins lasted 15-20 minutes. Many teachers were not aware that they could change their SLOs if revisions were justified and agreed to by their evaluator. Teachers in only four schools out of the entire sample reported that they had altered their goals in these meetings.

All principals interviewed or surveyed reported that they had held mid-year check-ins. When asked what constrained these opportunities, principals noted the high number of observations, pre-conferences, and post-conferences required by SEED. Principals reported that they had conducted some mid-year check-ins over e-mail because of the sheer number of meetings they needed to hold.

Summative Conferences

In our final round of data collection we asked teachers and school leaders about summative conferences. In the fall 2013 survey, 89% of teachers reported that they had had a summative conference with their evaluator for the previous school year, while 11% indicated that they had not had this meeting. A small percentage--4% of teachers (*n*=21)--indicated that they had had neither a mid-year conference nor a summative conference. These teachers many have been evaluated under their district's former evaluation system. Some principals reported that they were using their district's pre-existing evaluation systems for teachers in need of assistance. They and their supervisors feared that because SEED was in pilot status, any results based on the pilot year of SEED would not withstand legal scrutiny.

Moreover, in fall 2013 surveys, 90% of administrators indicated that they conducted summative conferences with all teachers assigned to them, but 10% indicated that they did not. About two thirds (63%) of administrators began holding summative conferences in May; 74% of administrators held summative conferences in June. A subset of administrators (16%) did not hold summative conferences before the 2012-13 school year ended; 11% held them over the summer and another 5% held them in September after the new school year began.

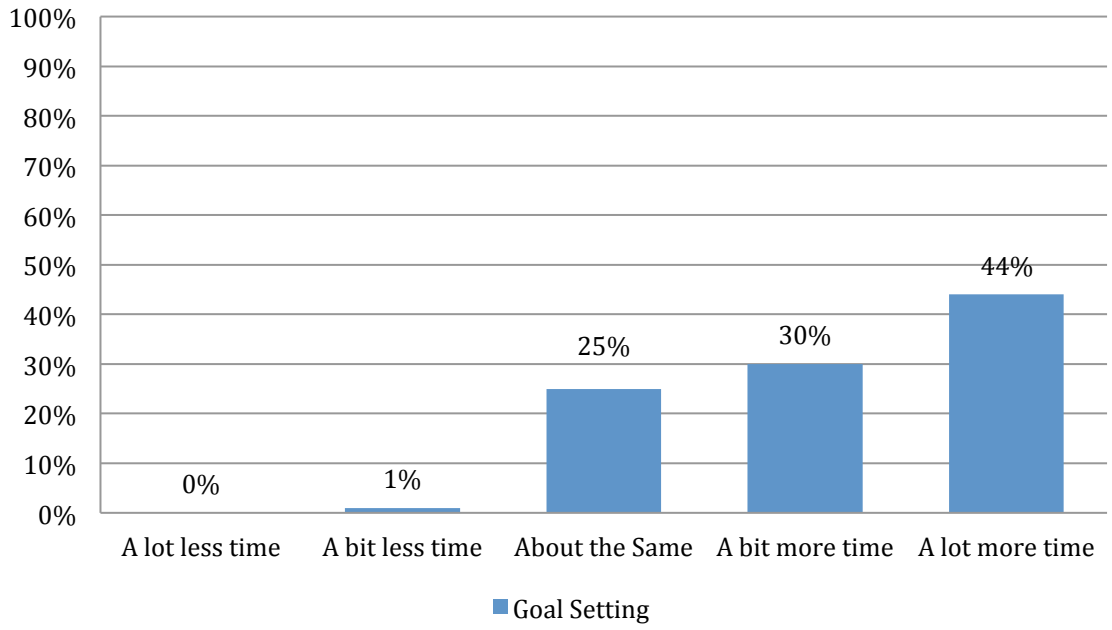
Goal Setting

Through all three rounds of our data collection, participants reported that goal setting [establishing Student Learning Objectives (SLOs) and Indicators of Academic Growth and Development (IAGDs)] consumed a substantial portion of their time. These goal-setting activities caused considerable stress for educators in the fall, with many reporting that they received insufficient guidance in how to write appropriate SLOs and IAGDs. By spring 2013, teachers reported less anxiety about this component of SEED.

Most teachers reported spending more time setting goals for themselves as part of the SEED process than they did under previous evaluation systems. Almost half (44%) of teachers surveyed in fall 2013, reported that they spent "a lot more time" on goal

setting than in previous years (see Figure 8). In all, 74% of teachers reported that they spent more time on these activities than in the past and essentially no teachers reported spending less time.

Figure 8. Time Spent on Goal Setting (Fall 2013 survey, *n*=533)



According to SEED, goals must be "rigorous" and "attainable." The majority of teachers (59%) reported that their SLOs were both rigorous and attainable. Moreover, less than half of the 44% of teachers who reported that they spent "a lot" more time on goal setting in 2012-13 reported that time spent on goal setting was very valuable. These findings may be due to the fact that teachers and specialists reported that they received minimal training or information about SLOs as they developed them and thus spent much of the fall trying to figure out the process. As a result, many teachers and specialists reported a lack of clarity or confusion about SLOs. Key confusions included the following:

- What makes for a strong SLO? What does a strong SLO look like across different subject areas and grades? What is "ambitious but attainable"?
- What are standardized and non-standardized measures in different subject areas and grades?
- What percentage of students should Student Learning Objectives/Indicators of Academic Growth and Development cover?
- What constitutes a good professional practice goal?
- In middle and high schools, should teachers have SLOs and IAGDs for each class they teach, overarching goals for all classes, or target a subset of their classes?
- When is it appropriate to change SLOs and IAGDs before the end of the school year?

In the first round of data collection, teachers and specialists reported having great difficulty understanding what constituted an acceptable SLO. In some cases, teachers or specialists worked together to craft an SLO. In a number of cases, they employed different strategies to complete this task but did not engage in the goal-setting process as the model envisions. For example, the lack of clarity on SLOs frustrated teachers such that one reported saying to her administrator: "just tell me what to write." This sentiment of compliance was echoed across multiple focus groups in the first round of data collection. Teachers at another school shared that they copied the state model SLO and changed the numbers to fit their classroom situation. Several groups of teachers reported finding colleagues who understood SLOs and asking these colleagues to write their SLOs. Finally, several groups of teachers reported that a trainer "basically wrote all our SLOs."

Many educators reported struggling in particular to find good assessments on which to base IAGDs. This was complicated by the move towards implementing Common Core State Standards and accompanying assessments in 2012-13. For example, a good number of teachers reported not having sufficient formative assessments in place and, as a result, some teachers did not give a pre-test. Other schools started using new assessments mid-year and teachers developed SLOs and IAGDs based on assessments that they had not yet seen. We anticipate these problems will decline as districts develop assessments and the state provides much more guidance on goal setting. However, it is clear that SEED requires relatively sophisticated assessment literacy and it is unclear whether educators possess this knowledge or that districts have plans or the capacity to help educators develop it.

During the last round of data collection, participants reported that Student Learning Objectives remained somewhat ambiguous even at the end of the pilot year. An interim principal who had not been present in the school when SLOs were set in the fall observed that teachers "were confused, goals were arbitrary, some were too ambitious." He added that teachers were used to aiming high with few consequences if they came up short. Of the overly ambitious goals, he said, "We tried to change some of these mid-year."

Another elementary principal suggested that there were systematic differences in rigor of SLOs within her school. She observed that teachers of the tested grades were less likely to achieve their goals than teachers of K-2 in her school. In her opinion, this was because the stakes were higher in the upper grades and goals depended on students having mastered skills prior to their grade. In the upper grades, "there is a longer history of learning or not learning," she stated. Another factor was the evaluator. She had overseen SLO creation for grades 3-5 but her assistant had managed K-2. "In grades 3-5 some teachers put 100% for their goals and I let them do that whereas in K-2 the assistant principal said she advised against that." With the extremely high goal of 100% of students reaching a benchmark, many teachers in grades 3-5 failed to achieve their goals.

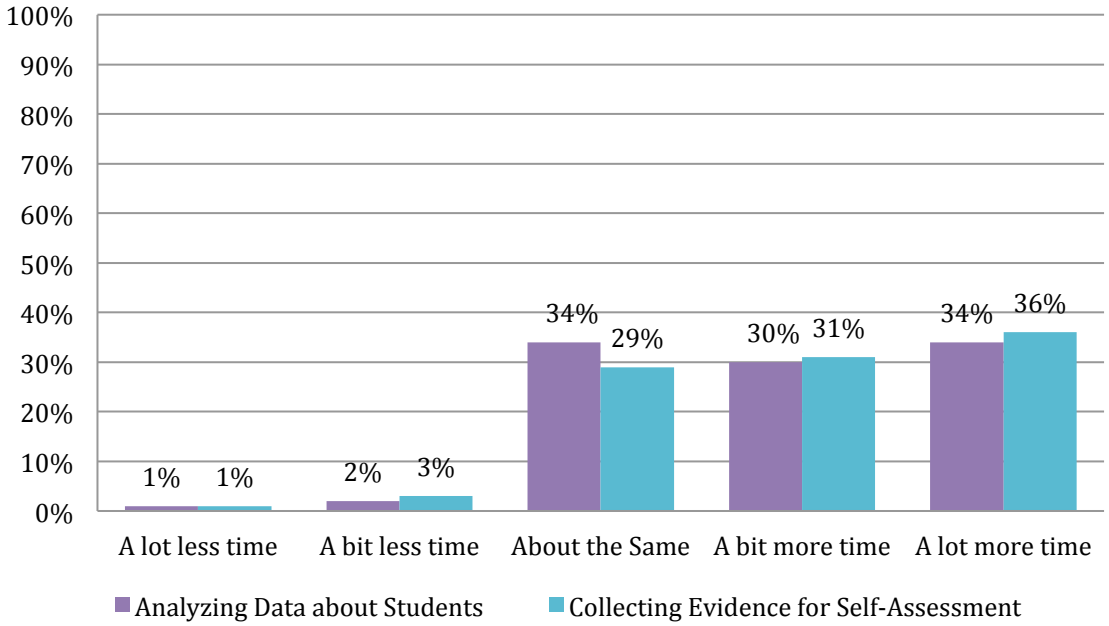
In some instances, SLOs were implemented in ways that were directly contrary to the SEED model. For example, a very small minority of teachers was forced to adopt SLOs/IAGDs established by principals. Other teachers were told they could not change their goals at the mid-year conference. Other teachers reported that their IAGDs did not measure growth, since there was no pre-test. Lastly, some teachers reported that they were using school wide measures for their SLOs/IAGDs. Teachers and principals were uncertain about when this was appropriate and when it was not (e.g., an art teacher measured on ELA CAPT; a music teacher with a CMT writing goal; a math teacher with a school wide ELA goal).

We believe the confusion with respect to goal setting were in part a result of the rushed implementation of the pilot and the lack of proper guidance that ensued. In addition, and not surprisingly, there appears to be a relationship between skill level of administrators and quality of the process.

Data Use

SEED has the potential to increase educators' use of student data. Educators are asked to set and monitor goals based on student data. Given that a substantial portion of their evaluation is based on whether or not their students achieve the goals they set, educators are likely to pay close attention to their students' performance throughout the year. Consistent with this assumption, we found that teachers reported a substantial increase in their analysis of data about their students. Specifically, 64% of teachers surveyed in fall 2013, reported that they spent more time analyzing student data in 2012-13 than in previous years. Two-thirds of teachers surveyed (67%) reported spending more time gathering data to assess their own progress than in previous years (Figure 9). It appears that SEED is spurring teachers to spend more time gathering and examining data. SEED has increased teachers' consideration of and use of student data, key precursors to instructional change. This increased data use is clearly a positive effect of SEED and one worth building upon in future years of the system. In particular, it will be important to assess and support the quality of this additional time invested in gathering and analyzing student data.

Figure 9: Teachers' Data Use Within SEED (Fall 2013 Survey, *n*=533)



Professional Learning Opportunities

According to the SEED model, teachers' summative conferences should culminate with a rating and recommendations for professional learning to address weaknesses identified through the evaluation process. Very few participants reported the presence of professional learning opportunities linked to evaluation outcomes. Almost no teachers reported that they had received specific recommendations of professional growth opportunities in such debrief situations. Almost all teachers reported that leaders did not make suggestions of professional development that they should pursue given their progress as indicated by classroom observations and student data. This was true even in schools that implemented SEED relatively robustly.

Although this is a key part of the theory of action undergirding SEED, pilot districts were not able to develop individualized professional development for implementation during the pilot year. In some cases, this was due to the fact that professional development was orchestrated at the district level. District leaders reported that there were few resources to devote to this component of SEED. In other cases, it appeared that the focus on other aspects of SEED (observations, designing the SLOs) overshadowed the development of professional learning opportunities that could be targeted to individual teachers' needs. Districts focused on carrying out the other aspects of SEED, namely the goal setting and monitoring and the classroom observation element, and the professional learning component generally received very little attention.

We see this lack of attention to the last portion of SEED—professional learning opportunities--as predictable given that this was the pilot year of the program and

districts and schools put a great deal of effort into implementing all aspects on the prescribed timeline. We view the professional learning component of SEED as having great potential to alter practice, however, and believe this aspect of the model should be closely examined and supported in future years.

Additional Measures

SEED also calls for teachers to be assessed based on parent feedback and whole-school measures of student learning or student feedback. We found that districts generally did not allocate resources towards developing instruments with which to gather feedback from parents or students. Districts cobbled together measures of parent feedback. Most districts opted to include school-wide measures of student learning instead of student feedback. Again, we found that these components of SEED were overshadowed by districts' emphases on SLOs and observation. We believe these components of SEED are vitally important and should be strengthened in future years.

School Administrator Evaluation

Overall, we found that districts had implemented school administrator evaluation in 2012-13. However, because of the emphasis on teacher evaluation, school administrators reported limited experience with administrator evaluation. One assistant superintendent said that administrator evaluation had gone "not as well as I would have liked. Teacher evaluation was the priority."

Although all districts implemented the components of administrator evaluation, they began to implement the SEED administrator evaluation quite late (i.e., December-January). Some principals received more observation and feedback from their evaluators than others. Veteran principals and assistant principals reported that they had not interacted much with this evaluation system during their pilot. "I didn't pay attention to it, to tell you the truth," said one high school assistant principal.

In fall 2013, we surveyed school administrators on their experiences with SEED's administrator evaluation. Almost half of the administrators surveyed (47%) agreed or strongly agreed that they received feedback about how to improve their leadership practice. A smaller subset (36%) indicated that their evaluator(s) had the time and resources to evaluate them accurately (see Table 10).

Table 10. Administrator Perceptions of SEED (Fall 2013 Survey, *n*=22)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
I received feedback from my evaluator(s) about how to improve my leadership practice.	18%	0%	35%	29%	18%
My evaluator(s) had the time and resources to evaluate me accurately according to the SEED model.	18%	18%	29%	24%	12%

Administrator evaluation was implemented in a limited fashion for a variety of reasons. These included a lack of professional development and considerable time and energy focused on teacher evaluation. Principals asked for additional training, citing that they had had only two overviews of the administrator evaluation. Although the majority of districts implemented SEED for administrators, they did so on a compressed timeline. Mid-year conferences in some districts were held in May.

For example, one district had no mid-year check-ins for principals. Principals generally reported being observed once or twice, but some were observed doing things that were not central to their work as instructional leaders. They reported that they did not get feedback based on these observations.

Although administrator evaluation was implemented lightly, participants reported some benefits from it. One assistant superintendent commented that, "we've really had substantive conversations [about administrator evaluation] due to SEED's pressure." She further noted that principals were "receptive to ratings" from SEED and "harder on themselves than we [district administrators] are."

Variations in Implementation

Given that implementation tends to vary by setting and participant, it was important for us to examine whether this was the case in SEED's enactment. We found variation in implementation at several levels. First, we found variation in implementation over time, with stronger implementation later in the pilot year. Second, we found variation by site, with stronger implementation at the elementary level and in districts with leaders who supported the SEED model and had prioritized evaluation in recent years. Third, we found stronger implementation with non-tenured teachers as compared to tenured teachers. Finally, we found weaker implementation in districts facing challenges, such as low performance and multiple concurrent initiatives, than in districts where educators did not have to divide their attention as much.

Variation Over Time

As of our first round of data collection, in October-January, 2012, we found that SEED had been implemented only partially. Due to the hurried roll-out of SEED and substantial work involved with setting SLOs and IAGDs, districts had implemented only portions of the SEED model as of late 2012. With a few exceptions, most districts delayed the launch of SEED for administrators until December or later. Moreover, many administrators did not begin to conduct classroom observations under SEED until November-December. This stems from lack of time. Not only were administrators spending substantial time in goal setting with teachers, they reported that they were also required to pass a *Teachscape* reliability examination prior to observing teachers. Administrators reported that preparing for and completing this examination took more than a day's time, which administrators found challenging to capture. Indeed, in the fall of 2012, 45% of the principals interviewed expressed strong concerns about the *Teachscape* assessment, noting the amount of time required to prepare and take the test, the absence of feedback from the results, and low quality videos.

In fall 2012, participants also expressed strong concerns about the lack of communication and opportunities to learn about SEED. Almost all classroom teachers and specialists we interviewed reported that they lacked clarity about key aspects of SEED. Participants, particularly teachers, reported in the first round of data collection (fall, 2012) that they had not received enough training regarding SEED's components and, in particular, SLOs. Classroom teachers in many districts reported that they had received only two training sessions of between one and two hours each. Additionally, teachers in several districts reported that training was delayed, trainers were not able to answer their questions, or training was not conducive to teacher learning because it relied too much on lecture format and offered little or no follow-up. Classroom teachers and specialists indicated that the great majority of training they received was provided by their principal. Principals reported receiving more training than teachers, in some cases attending three or four-day training sessions in the summer or early fall. Newly hired principals, however, sometimes missed this training, putting them at an initial disadvantage in understanding the model and implementing it with fidelity.

Teacher and specialist participants reported that they were provided with few opportunities to develop a deep understanding of SEED. Teachers and specialists reported that principals often clarified or communicated about SEED via e-mail. Principals did not meet one-on-one with some teachers and specialists to set goals, choosing to handle this process by holding group sessions instead. In some cases, teachers perceived group goal setting as positive, for they felt it fostered collaboration. In other cases, teachers perceived this as negative; they felt it forced uniformity of goals when their diverse classrooms warranted more varied SLOs and IAGDs. Most teachers and specialists reported that they wanted more coaching on what SEED demanded of them and how to create an SLO.

In the second and third rounds of data collection, participants reported increased clarity and decreased stress related to SEED. This appeared to be due, in part, to improved communication from districts and the state, but also to greater familiarity with the SEED policy and procedures, as well as decreased emphasis on SEED in some sites after the initial goal setting phase in the fall. Participants also developed a better understanding of the SEED model through the process of implementing it. Anecdotal evidence from pilot districts suggests that their second year of SEED implementation (2012-13) is proceeding more smoothly than the pilot year.

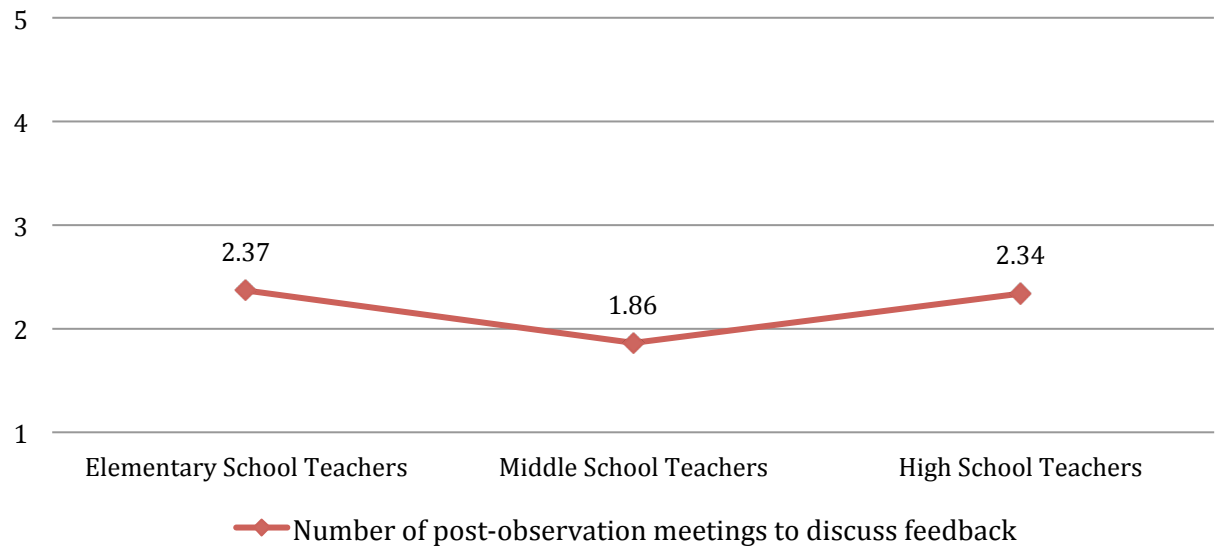
Variations across Grade Levels

A one-way analysis of variance (ANOVA) was used to test whether teachers at different grade levels reported similar experiences with SEED implementation. There were no significant differences in the number of formal or informal observations during the SEED pilot year across grade levels, F (2, 415) = .95, p = .388, and F (2, 409) = .38, p = .682, respectively. Teachers also did not report any statistically significant differences in the extent to which observations and feedback occurred more frequently under SEED than in previous years, F (2, 418) = 2.40, p = .092.

Teachers reported significant differences across grade levels in the number of meetings to discuss feedback about their observations, F (2, 413) = 7.21, p = .001 (see Figure 10). Tukey's post-hoc comparisons indicate that middle school teachers (M=1.86, 95% CI [1.62, 2.10]) reported significantly fewer meetings to discuss feedback about their observations than elementary school teachers (M=2.37, 95% CI [2.20, 2.54], p = .001) and high school teachers (M=2.34, 95% CI [2.16, 2.53], p = .004) (Figure 10).[3] The difference between elementary school teachers' responses and high school teachers' responses was not statistically significant at $p < .05$.

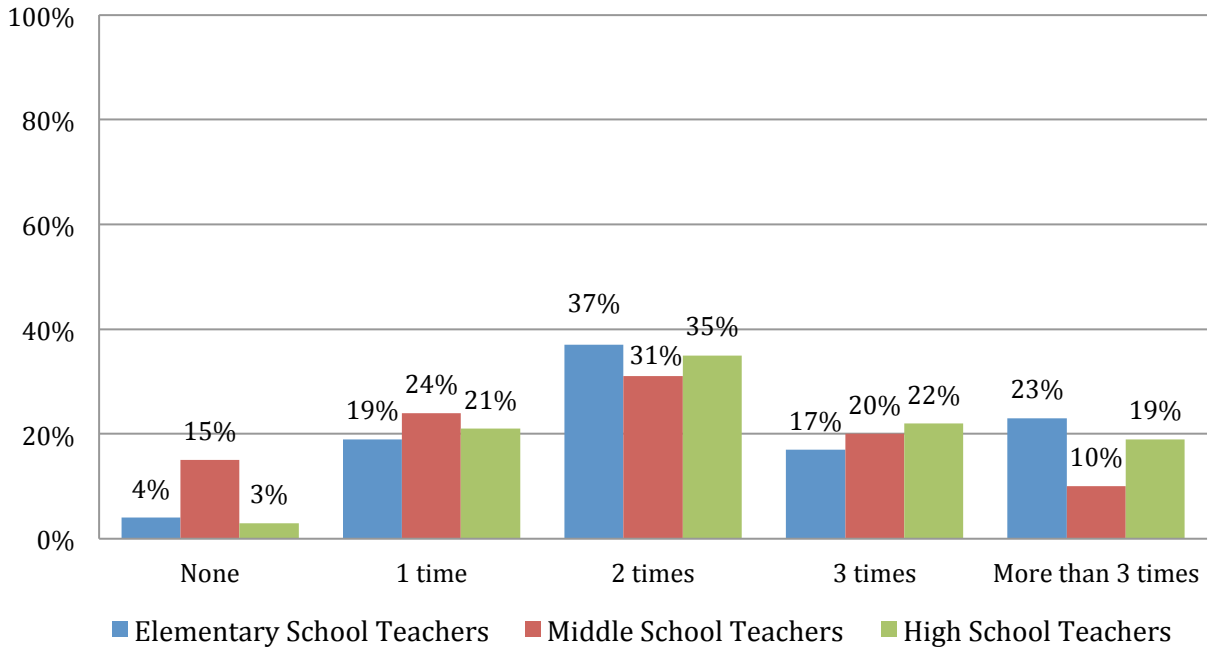
[3] Mean scores fall on a Likert-type scale where 1=Strongly Disagreed and 5=Strongly Agreed.

Figure 10. Means Plot for Number of Post-Observation Meetings across Grade Levels (n=533)



Forty percent of elementary school teachers and 41% of high school teachers reported three or more meetings with their evaluator to discuss feedback from observations; only 30% of middle school teachers reported meeting the number of meetings recommended in SEED (see Figure 11). Another 37% of elementary school teachers reported two post-observation meetings, 35% of high school teachers reported two post-observation meetings, and 30% of middle school teachers reported two post-observation meetings. This indicates a 15 to 16 percentage point difference in the teachers close to or fully receiving the number of recommended post-observation meetings to discuss feedback.

Figure 11. Distribution of Meetings to Discuss Observations by Grade Level (*n*=533)



Of the 67 teachers who indicated that they did not have a mid-year conference, almost half (45%, *n*=30) taught elementary grades (vs. 33% of the sample), 24% (*n*=16) taught middle grades (vs. 19% of the sample), 18% (*n*=12) taught high school (vs. 27% of the sample), and 13% (*n*=9) declined to indicate grade level. The difference across groups is approaching significant, with χ^2 (2, 418) = 5.49, *p* = .064. Teachers without a summative conference were evenly distributed across grade levels, with 28% (*n*=15) teaching elementary grades, 28% (*n*=15) teaching middle grades, and 30% (*n*=16) teaching high school; 15% (*n*=8) declined to indicate a grade level. There is no significant association between grade level and having a summative conference, with χ^2 (2, 421) = 2.76, *p* = .252.

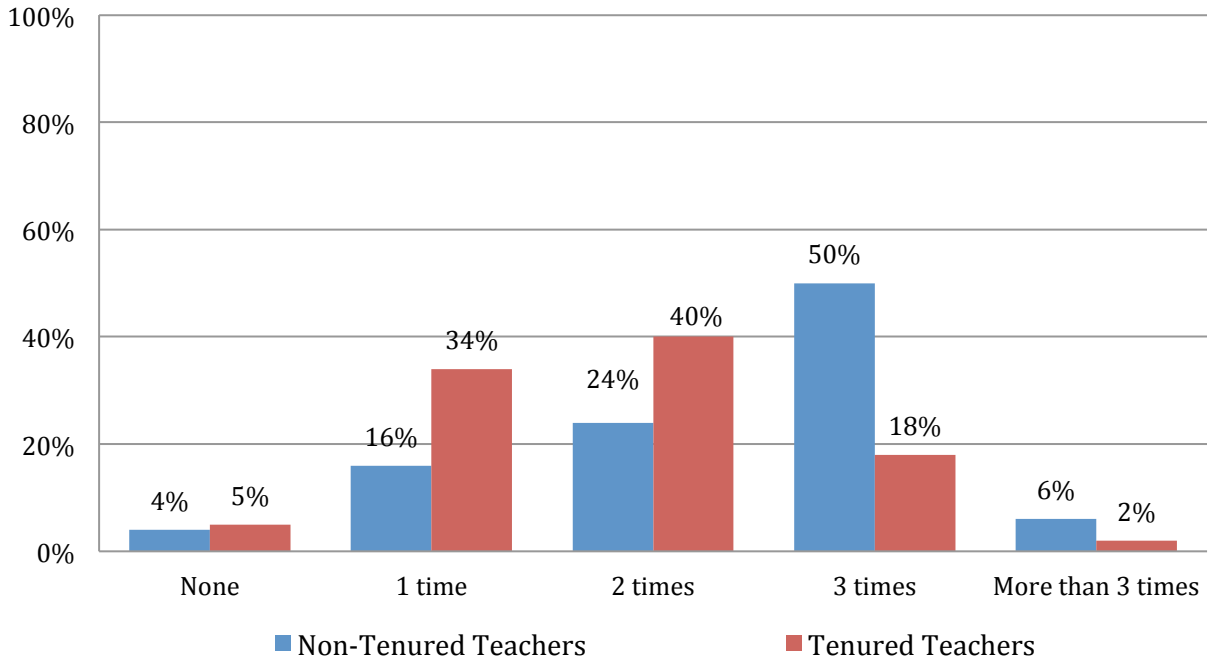
Variations by Tenure Status

Independent-samples *t* tests were used to test whether there were differences in teacher experiences with SEED implementation across tenure status. There were no significant differences between non-tenured and tenured teachers in the number of informal observations or the number of post-observation meetings to discuss feedback.

Non-tenured teachers were more likely than tenured teachers to report that observations and feedback occurred more frequently under SEED than in previous years, with *t* (131.63)=2.37, *p* = .019. Non-tenured teachers also reported significantly more formal observations than tenured teachers, with *t* (432) = 5.71, *p* < .001. Fifty-six percent of non-tenured teachers reported having three or more formal observations, as

recommended by SEED policy; in contrast, only 20% of tenured teachers reported three or more formal observations (see Figure 13). As noted above, this is due at least in part to a conscious decision by administrators to manage their evaluation load by observing non-tenured teachers more than tenured teachers.

Figure 13. Distribution of Formal Observations by Tenure Status (*n*=533)



Of the 67 teachers who indicated that they did not have a mid-year conference, 75% (*n*=50) were tenured (vs. 67% of the sample), 16% (*n*=11) were not tenured (vs. 15% of the sample), and 9% (*n*=6) declined to indicate tenure status. Similarly, of the 54 teachers who indicated that they did not have a summative conference to discuss their summative rating, 72% (*n*=39) were tenured, 17% (*n*=9) were not tenured, and 11% (*n*=6) declined to indicate tenure status.

Variations by District

Wide variations by district also were apparent. Although we caution readers against drawing firm conclusions on these statistics because of the small sample size for some districts, there were significant differences in the number of observations teachers in different districts reported receiving. For example, the percentage of teachers reporting that they had received three or more formal observations ranged from 8% in one district to 71% in another. Similarly, the proportion of teachers reporting fewer than two observations ranged from 7% in one setting to 60% in another. In general, lower-performing districts with multiple initiatives underway experienced more difficulty in implementing SEED with fidelity. In these districts, teachers were more likely than in

higher-performing districts to report not having a mid-year conference or summative conference.

In large part, these implementation challenges were connected to the large number of concurrent reform initiatives occurring in these districts. These other initiatives demanded leaders' and teachers' time, thus reducing a key resource necessary for implementing SEED.

Closing Remarks

On the whole, districts and schools implemented SEED with a relatively high degree of fidelity. They invested particular energy in implementing the goal setting and observation components of SEED. We further found that schools with the greatest challenges (i.e., very low student achievement and multiple reform initiatives) seemed to have the most difficulties in using the model to drive improvements in practice. Overall, implementation proceeded much as we would expect during the pilot phase of any new major initiative.

Findings on Educator Experiences

In addition to implementation, a second area of interest to the state was how educators in the 14 districts experienced the pilot implementation of SEED. Thus, we gathered data on teachers' and leaders' views of and experiences with this reform. This section presents data in response to research questions investigating these areas. Specifically, we present findings in response to the following questions:

EDUCATOR EXPERIENCES WITH SEED:

1. How did educators in pilot sites experience SEED?
2. What variations occurred? What explains these variations?

Introduction

Over the course of the year and three rounds of data collection, educators expressed a variety of perspectives on SEED. Teachers raised concerns in the first round of data collection, which were to be expected as they learned about and implemented a new initiative at a relatively fast pace, but over time teachers expressed relatively more acceptance and less anxiety. Ultimately, educators in the pilot districts expressed a positive or neutral view of SEED. Because of the rigors of learning and implementing a new system, most educators viewed SEED as a system focused mainly on their evaluation as opposed to their development. Given the pilot status of the initiative, this is not surprising.

In the following sections, we present educators' general views of SEED, as reflected in interviews and surveys. We then discuss their attitudes toward components of SEED, including observations, the instructional practice rubric, post-observation conferences, and goal setting (i.e., SLOs and IAGDs). We present data on variations in educators' attitudes and then turn to administrators' attitudes towards the administrator version of SEED.

General Views of SEED

Participants holding a range of roles expressed a positive or neutral view of the SEED model. Very few participants rejected the entire model out of hand, even if they critiqued parts of the policy or expressed frustration with how it was implemented during the pilot year.

Teachers' Views

Teachers[4] reported considerable anxiety and confusion regarding SEED in the first round

[4] Recall that we use the term "teacher" to encompass all personnel evaluated under the teacher version of SEED. Specialists are highlighted by name when their views diverged from those of classroom teachers.

of data collection. Over time, these sentiments diminished, although they did not completely disappear. In fact, many teachers have consistently voiced support for the ideas behind SEED. Teachers in particular appreciate the fact that SEED promises to increase the frequency of classroom observations by school leaders and strengthen the feedback they receive regarding their instruction. They believe the focus on student performance growth, rather than mastery, is a positive element of SEED as is the opportunity to adjust their goals mid-year. Teachers reported that the fact that administrators are also evaluated is a strong point of the model. Many school leaders in pilot districts emphasized to teachers that their own evaluations depended upon the performance of their teachers, and teachers appreciated this fact.

A considerable number of teachers felt that SEED has good potential. Forty-two percent of teachers surveyed in fall 2013 felt that with sufficient resources, such as time and staffing, they would be able to use SEED to improve teacher practice at their school. One-third of teachers (34%) were neutral on this question. Sixty-one percent of teachers reported that they understood how they were evaluated and 78% of teachers indicated that they felt comfortable being evaluated by their supervisor in SEED's pilot year.

Most principals reported that teacher feedback on SEED has been positive or neutral. According to surveys from fall 2013, 90% of administrators agreed or strongly agreed that teachers understood how they were evaluated under SEED and 85% agreed or strongly agreed that teachers were comfortable being evaluated by their evaluator. Principals indicated that some teachers have expressed appreciation for how SLOs have focused their attention on the progress of all students. Many complain that the forms used during the pilot year were cumbersome. In a very small number of schools principals reported that most teachers had a negative experience with SEED. One principal explained that in his building, "some [teachers] hate it" because SEED "started out so roughly…it put a bad taste in people's mouth at the beginning."

School Administrators' Views

Overall, principals in pilot districts support the SEED model and embrace its focus on instruction and its function as a lever for professional growth. Many administrators note that SEED's approach to evaluation, with its inclusion of classroom observations and goal setting, is not substantially different from what they had done in the past. They welcomed SEED's focus on student performance.

A number of administrators expressed concern, however, about performance labels (e.g., "proficient") being a distraction to the more nuanced and detailed feedback from an evaluation. They further argued that the labels may be in conflict with the notion of support and, at the extreme, threaten to de-professionalize teaching. Their greatest concern, foreshadowed in the prior section, is the large quantity of their time that SEED demands.

Principals reported that the pilot version of SEED increased their workload substantially. One principal, for example, said she was less available for non-evaluation matters than in previous years. Another principal observed that she had less time to think, plan, and innovate as a leader. Several principals reported that the amount of required paperwork reduced their time in the classrooms, holding one-on-one conferences, or participating in team meetings, thus diminishing their instructional leadership activities. Principals added that they did more work, in some cases much more, after school and at home than in previous years. One principal spoke for many by saying, "The volume of this system is overwhelming. It took a huge chunk away from other duties." Said another principal, SEED "got in my way…it's cumbersome, it slowed me down in completing evaluation." One principal kept track of the hours he spent on SEED and found that, on average, he devoted more than 60% of each day to evaluation, which he felt compromised his ability to focus on developing the school's faculty as a whole.

Principals in low-performing districts reported particular challenges in implementing SEED. One principal explained that having several initiatives happening all at once "puts everything in flux." This further limited the time principals in such settings could devote to SEED.

In some ways, the reported increase in principals' workload is a positive finding, as one goal of SEED was to increase principals' attention to and work on instructional leadership. As discussed below, it is clear that SEED caused administrators to spend more time in classrooms observing teachers. This is a benefit of SEED that should be cultivated. From another perspective, this finding is cause for concern; if principals feel overwhelmed, the quality of their work—whether in the area of instructional leadership or outside it--may suffer. It is clear that SEED changes administrators' jobs, sometimes substantially. The state and districts should help school administrators manage their workloads in ways that promote school and, ultimately, student success.

Principals also outlined several ways in which SEED was beneficial to their leadership. Specifically, most principals identified the rubric and summative meetings as particularly positive aspects of SEED. These are discussed in more detail below. A few principals reported that SEED's administrator evaluation piece was positive. In general, principals called for more professional development and time to learn how to use SEED in ways that maximize its benefits. As one principal said, "SEED has a lot of plusses. Make sure everyone has time to learn it and use it."

District Administrators

The majority of district leaders reported that SEED had been a positive experience for educators in their district and were cautiously optimistic about the new model. Despite "angst" over goal setting, one district leader said, "SEED has been largely positive." She added, "we haven't heard a ton" of feedback from teachers regarding the rubric or summative meetings.

In contrast, another district leader reported that there was a great deal of tension (particularly around goal setting) and low morale in his district that teachers have attributed in part to the SEED pilot. Multiple district leaders reported that educators in their district had negative experiences with *My Learning Plan*. One district leader reported that teachers found that "*My Learning Plan* is a pain," even though they supported SEED more broadly.

One district leader identified the next level of work for SEED within her district: "Everyone is stuck on the data thing. People still aren't seeing the link between practice and data. This is the bow we need to tie for them. We need to make that connection. They are scared of the SLOs because they are used to observation" and thus find it less intimidating.

Summary

Overall, participants reported that they viewed SEED favorably. They generally agreed that educators should be assessed in part based on growth in student learning and that observations should also contribute to their evaluations. On the whole, participants, whether teachers or leaders, reported that SEED is similar to their prior evaluation systems. The main similarities lie in SEED's emphasis on observation and teacher-created goals. The main differences are in (1) the more explicit instructional standards, embedded in the rubric, (2) an increased number of observations, (3) the use of student learning objectives, and (4) the summative, numerically-based rating.

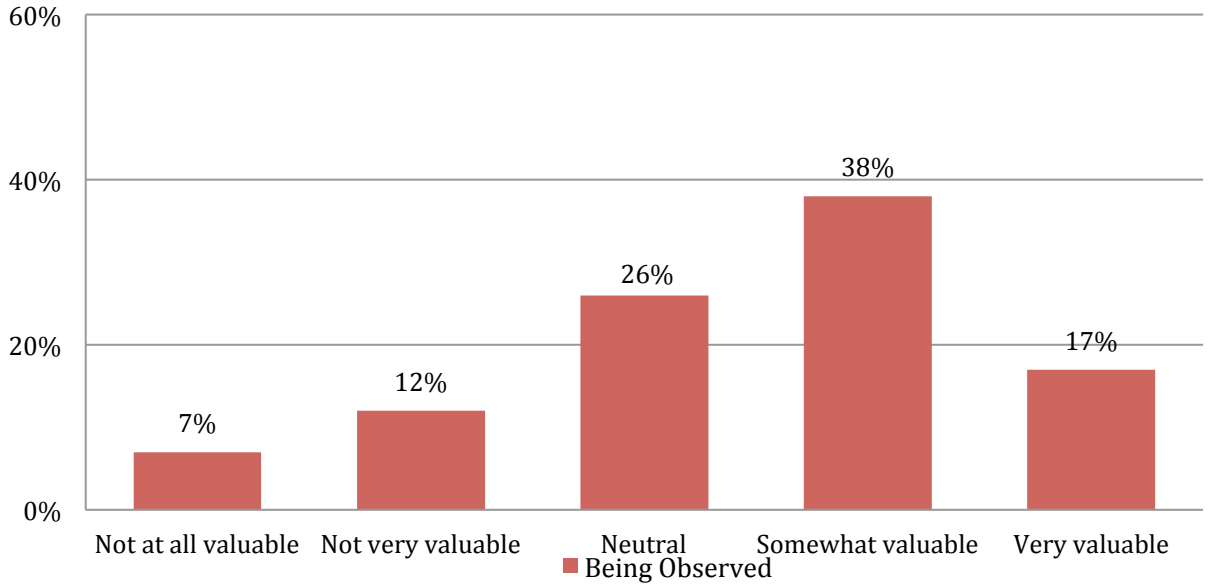
Despite their familiarity with its components, educators felt that SEED is complex. Teachers were more unified in this perspective than principals or district leaders, who had received more training on the model. In most focus groups, teachers reported that SEED was "overwhelming" or that they felt "overwhelmed" by the new system, especially in the first round of data collection (fall, 2012). For example, teachers and principals report that establishing SLOs was very time-consuming. Principals and assistant principals often had to educate teachers about how to write SLOs through the goal-setting process.

Educators reported that the time between trainings and action on tasks (e.g., SLO development) was very short. Timelines were reported to be rigid, despite delays in training and the finalization of forms. In some cases, educators reported that teachers and specialists received training on a component of SEED and were expected to implement the component immediately. The vast majority of participants, from district officials to teachers to trainers, reported that the pace of implementation was too rushed to maximize SEED's benefits.

Observations

Participants reported that the observation component of SEED is generally similar to what they had experienced under previous evaluation systems. The main differences were an increased number of required observations, the use of a standardized rubric to guide observations, and the resulting summative rating. At the conclusion of the pilot year, 55% of teachers surveyed reported that being observed under SEED was somewhat or very valuable (Figure 14).

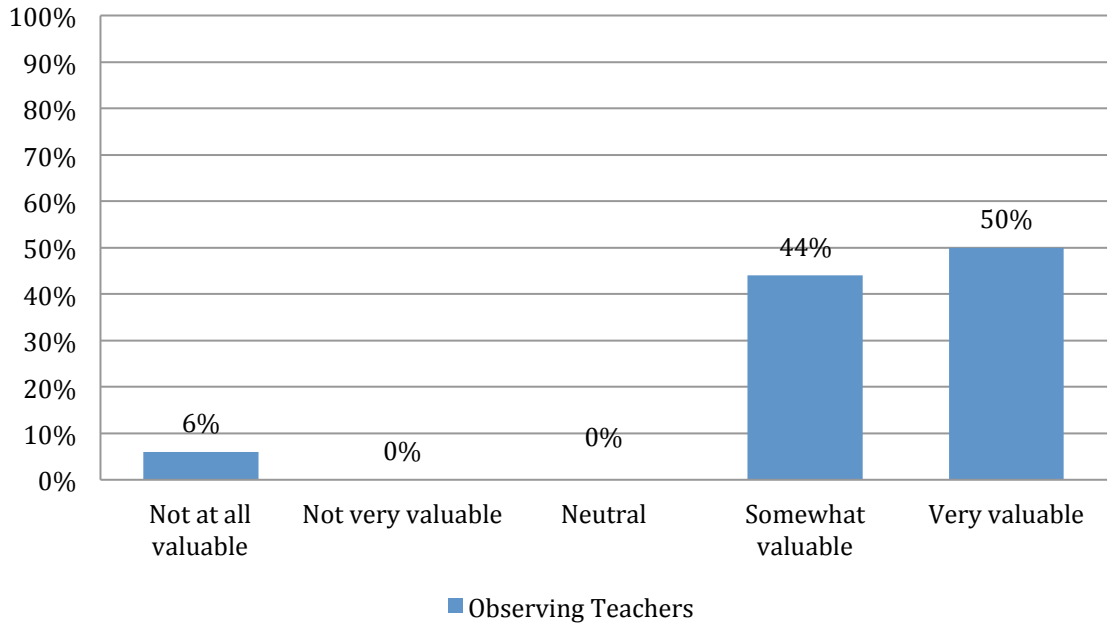
Figure 14. Teachers' Reported Value of Time Spent Being Observed Under SEED (Fall 2013 Survey, *n*=533)



When we examine the responses of teachers who reported being observed more than in prior years, we find that a slightly larger percentage--58%--found observations valuable. In interviews, teachers echoed this sentiment, with a majority of teachers saying that they found the experience of being observed helpful to them. In particular, they were grateful for feedback on their instruction.

School administrators reported observations were quite valuable to them. Although all administrators reported that they struggled to complete the required observations, the overwhelming majority--94%-- of administrators reported that observing teachers under SEED was somewhat or very valuable to them (Figure 15).

Figure 15. Administrators' Reported Value of Time Spent Observing Teachers Under SEED (Fall 2013 Survey, *n*=22)



Rubric

SEED requires that teachers be evaluated using a rubric with multiple performance categories. This is a marked departure from prior practice in most Connecticut districts. On average, teachers expressed neutral or positive views of the observation rubric. A third of teachers (34%) agreed or strongly agreed that the observation rubric accurately describes a continuum of teacher quality and 32% of teachers were neutral on this question (Table 11).

Table 11. Frequencies for responses to "The observation rubric accurately describes a continuum of teaching quality" (Fall 2013 Survey)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
Teachers (*n*=533)	16%	19%	32%	28%	6%
School Administrators (*n*=22)	5%	11%	16%	42%	26%

Administrators' views of the rubric were more positive than those of teachers. Two-thirds (68%) of administrators surveyed indicated that they believed that the observation rubric accurately described a continuum of teaching quality. In interviews, principals reported that the rubric helped them make evaluation meaningful. In the past, principals reported that observations were, in the words of one principal, "very, very

subjective." She added that, in her view, *Teachscape* taught administrators how to conduct more objective observations.

Another benefit of the new observations cited by school leaders is that the rubric encouraged a "common language." Under SEED, good instruction "seems more clear" to teachers and administrators, said one principal. "This was really helpful," she said. "I am sorry I got it in the last year of my tenure (career)." The rubric "takes pressure off me. Numbers don't lie. It's helped me to have those harder conversations… I like to make people happy…and it's helped me to be direct and make hard decisions."

School leaders commented that the rubric lacked specifics in some domains (e.g., 3A) and raised questions about its applicability in certain situations, for example, with specialists. As would be expected, administrators also reported that it took them a considerable amount of time to become familiar with the rubric and how to match evidence to different domains. Some administrators also struggled with the large amount of information generated by the many indicators included in the rubric. They said they needed to streamline the information for each teacher in order to make feedback useful, rather than overwhelming.

Leaders of many schools further reported that teachers generally responded positively to the rubric. They observed that many teachers had to recalibrate their views of their own performance. Said one district leader, "teachers have had to come to terms with lower scores," with some teachers struggling with the sentiment that "I used to be 'proficient' and now I'm 'developing.'" District and school leaders reported that teachers had responded neutrally or positively to the rubric.

Conferences

Multiple data sources suggest that post-observation conferences were one of the most valuable aspects of SEED in the pilot year. Overall, 57% of teachers surveyed found talking with evaluator to be a valuable experience (Figure 16). As we will discuss in detail below, teachers who spent more time in post-observation conferences under SEED compared to prior evaluation systems found it to be of even greater value. Specifically, 69% of teachers who spent more time during SEED debriefing with evaluators after observations found this to be somewhat or very valuable.

Similarly, school administrators reported that post-observation conferences were valuable. The vast majority (94%) of administrators reported that such meetings were valuable or very valuable. Fifty percent of respondents indicated that these conferences were very valuable. As might be expected, a lower percentage (45%) reported that writing up observations was somewhat or very valuable.

Figure 16. Value of Time Spent In Post-Observation Under SEED (Fall 2013 Survey, *n*=533)

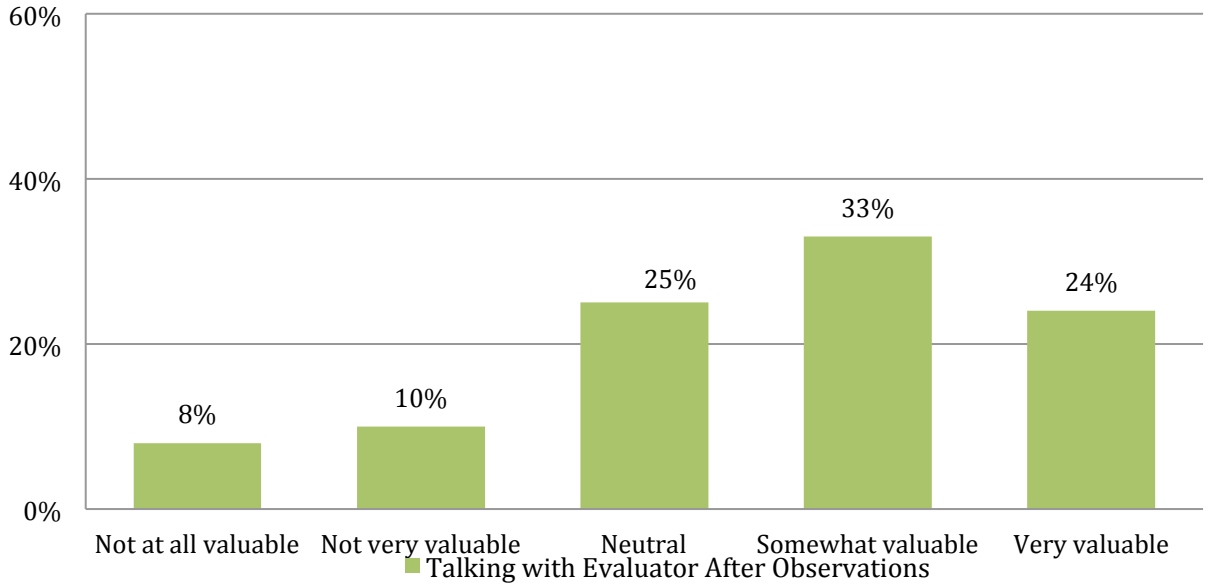
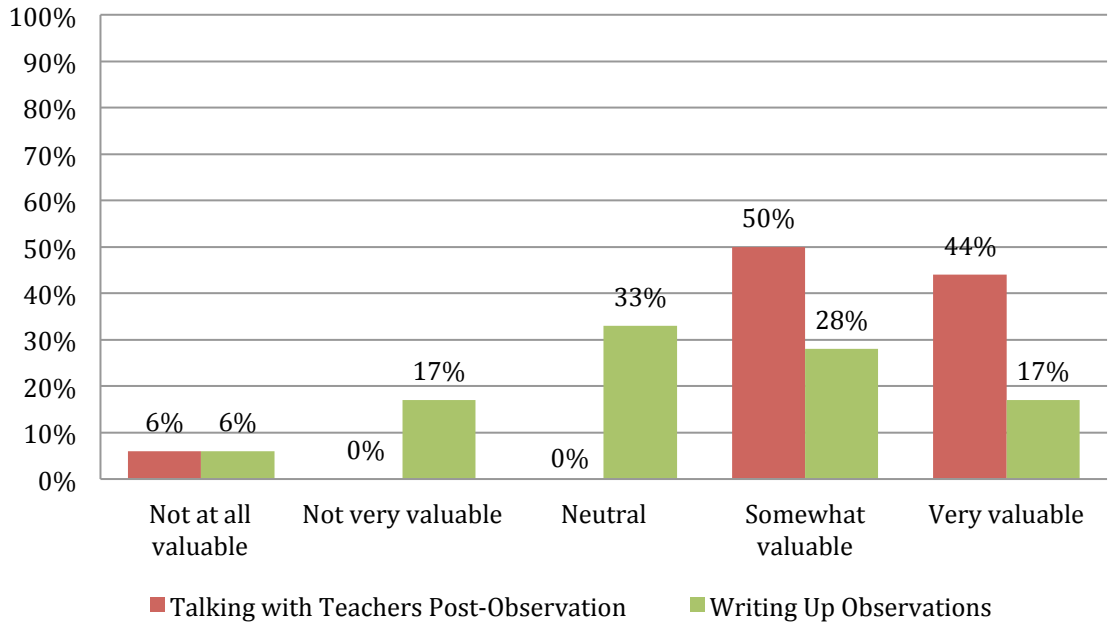
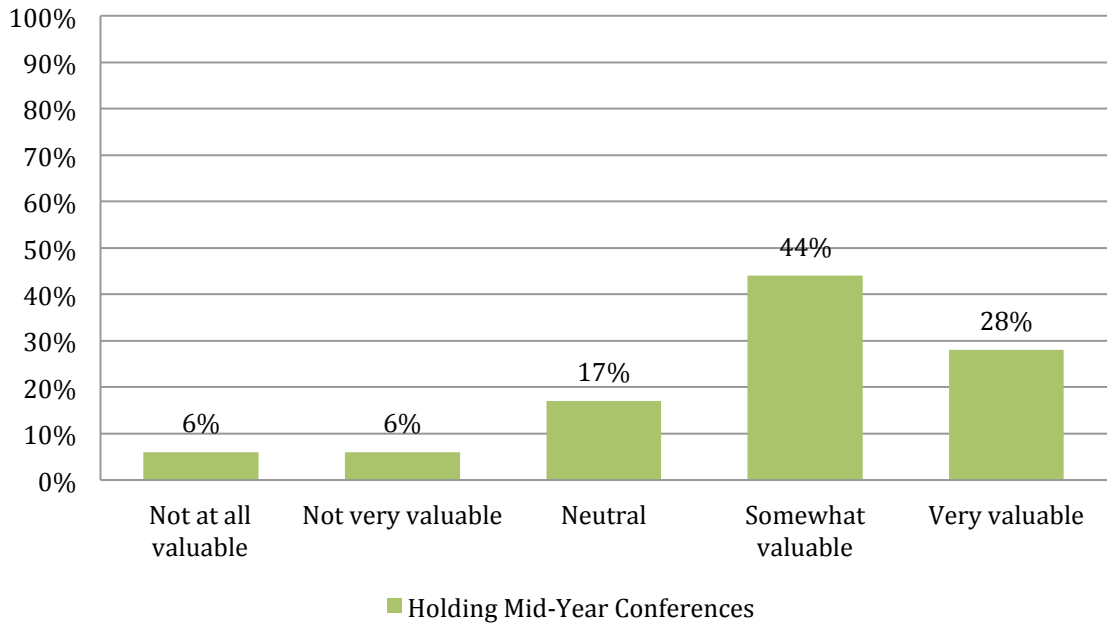


Figure 17. Administrators' Reported Value of Time Spent in Teacher Post-Observation Tasks Under SEED (Fall 2013 Survey, *n*=22)



As noted in prior sections, teachers reported that mid-year conferences were brief and perfunctory and did not result in much learning. However, administrators felt these conferences were valuable. Seventy-two percent of administrators reported that the mid-year conferences they held were somewhat or very valuable, perhaps because they gained information from teachers about their practice during these sessions (Figure 18).

Figure 18. Value of Administrator Time on Mid-Year Conferences (Fall 2013 Survey, *n*=22)



Summative Conferences

On the whole, a majority of teachers reported that their summative rating was accurate. More than half (58%) agreed or strongly agreed that their summative rating for the pilot year was accurate (see Table 12). Seventeen percent disagreed or strongly disagreed, with one-quarter expressing a neutral position. The administrator survey also included this item. When asked whether they felt teachers' summative ratings were accurate, 47% agreed, one-quarter were neutral and the balance disagreed or strongly disagreed (Table 13).

Table 12. Frequencies for Teachers' Perceptions of the Validity of SEED Measures (Fall 2013 Survey, *n*=533)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
I feel that my summative rating from 2012-2013 is accurate.	7%	10%	25%	44%	14%

Table 13. Frequencies for Administrators' Perceptions of the Validity of Teacher SEED Measures (Fall 2013 Survey, *n*=22)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
Teachers' summative ratings from 2012-13 are accurate.	10.5%	15.8%	26.3%	47.4%	0.0%

Teachers found summative conferences to be neutral to valuable. Forty-three percent of teachers reported that such meetings were somewhat or very valuable, while 32% of teachers were neutral on this question.

In contrast to this, survey data revealed that the great majority--89%--of administrators found the summative meetings to be somewhat or very valuable (fall 2013 survey). In interviews, principals reported that summative conferences were equally or more productive than in the past. The fact that the rating was, in one principal's words, "all based on data" reduced the frequency of arguments with teachers over summative results according to principals. Most principals reported that the summative ratings generally supported their assumptions about teacher performance.

One principal said that the fact that the summative *My Learning Plan* document showed all evidence from throughout the year helped make summative conferences data-driven. She said that under SEED the summative conference "was a very, very good conversation. It was much more productive, focused than in years past" since the conversation was focused on the evidence rather than based on broad reflective questions. Another principal felt that summative conferences "went very, very well." While she felt that *My Learning Plan* was burdensome in other respects, she felt that it added value to the summative conference: "Summative is the easiest piece of all," because, "*My Learning Plan* compiles everything."

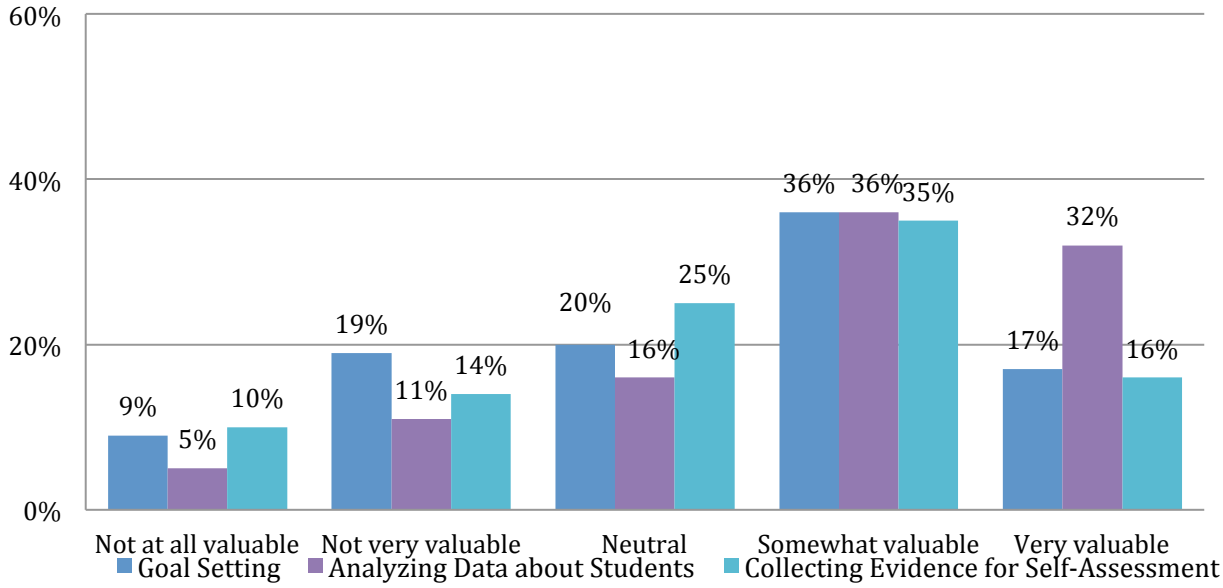
Goal Setting

In the early months of SEED's implementation, goal setting was a source of confusion and anxiety for many teachers and some school administrators. In our first round of interviews, teachers, in particular, reported feeling great strain about SLOs and IAGDs. At the conclusion of the pilot year, teachers reported much less stress about goal setting. Survey data indicated that goal setting, in general, and the analysis of student data, in particular, were ultimately valued by teachers in pilot sites.

On the whole, 39% of teachers surveyed in fall 2013 reported that SLOs were useful to them as professionals. Twenty-nine percent of teachers were neutral on this question.

Overall, teachers were more sanguine about the processes involved in goal setting. Teachers reported finding goal setting, the time spent analyzing data about their students, and collecting evidence for their self-assessment valuable, on average (see Figure 19). Fifty-three percent of teachers reported that goal setting was somewhat or very valuable in the fall 2013 survey. More than two-thirds of participants (68%) indicated that analyzing student data was somewhat or very valuable and 51% reported that collecting evidence for self-assessment was similarly valuable.

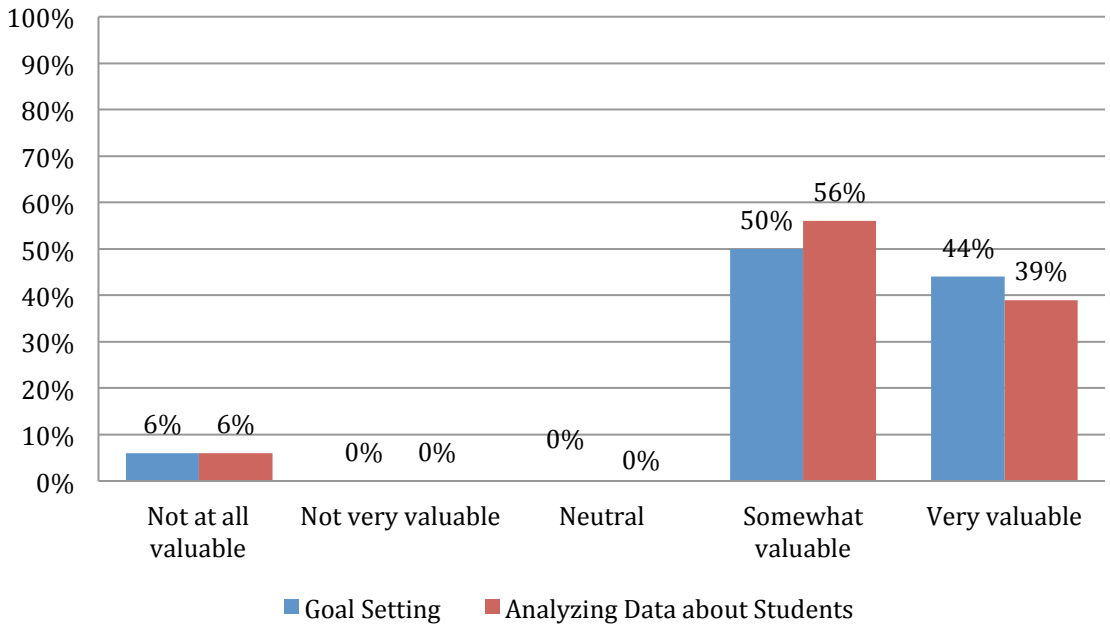
Figure 19. Value of Time Spent on Improvement-Related Tasks (Fall 2013 Survey, *n*=533)



Moreover, teachers who spent more time analyzing data about their students found that time to be valuable. As depicted in Figure X below, 67% of teachers who spent more time during SEED analyzing data about their students found this to be somewhat or very valuable.

Administrators reported that they found goal setting and having teachers analyze data about their students within SEED to be valuable. As shown below (Figure 20), 94% of administrators surveyed reported that the goal-setting process was somewhat or very valuable and 95% responded in this manner about that the data analysis process.

Figure 20. Value of Administrator Time on Goal Setting Tasks (Fall 2013 Survey, *n*=22)



Relative Value of Activities

To gain a sense of SEED's potential effects when implemented as the model intends, we examined the responses of teachers who reported that they had spent more time on various evaluation tasks than under prior evaluation systems.

When we examine only those teachers who indicated that they spent more time on various teacher evaluation and development tasks, we find that a modest to large percentage of teachers indicated that this additional time on SEED-related tasks was somewhat or very valuable (see Figures 21 & 22). Two activities in particular were valuable to a large proportion of teachers: 69% of teachers who spent more time during SEED talking with evaluators after observations found this to be somewhat or very valuable and 67% of teachers who spent more time during SEED analyzing data about their students found this to be somewhat or very valuable. Moreover, among teachers who reported being observed more than under previous evaluation systems, 58% reported that being observed was valuable. Of teachers who reported spending more time on goal setting, 52% found it valuable or very valuable. Fifty-one percent of teachers who spent more time gathering data about their students for self-assessment found this process valuable.

Figure 21. Value of Additional Time Spent on Teacher Evaluation (Fall 2013 Survey, *n*=533)

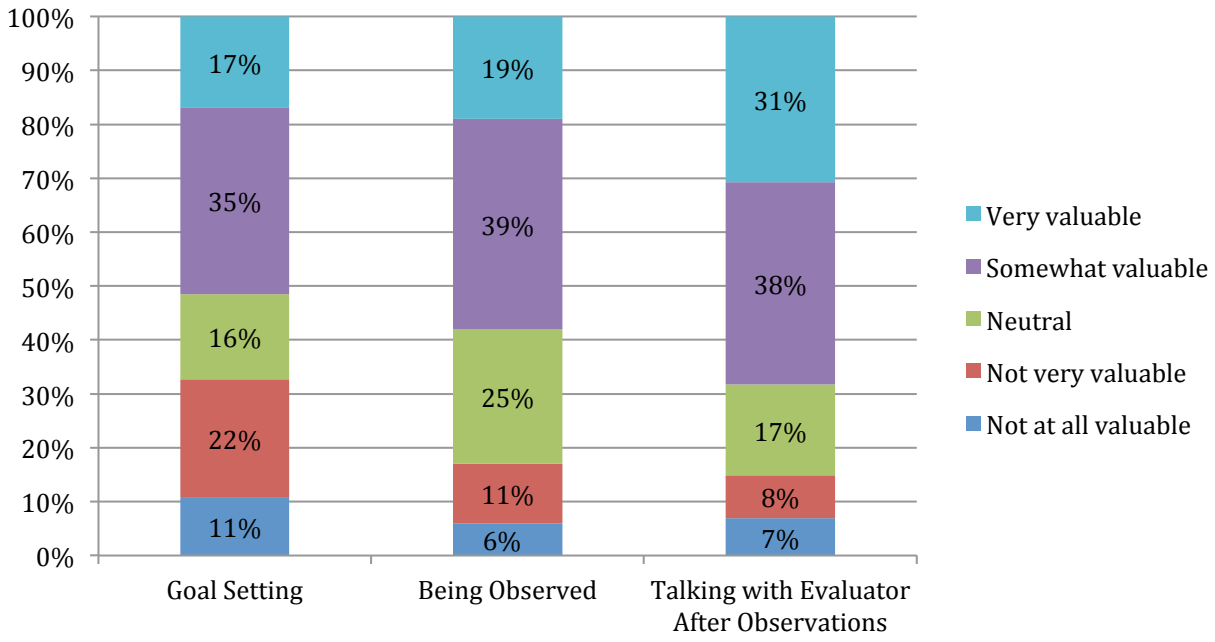
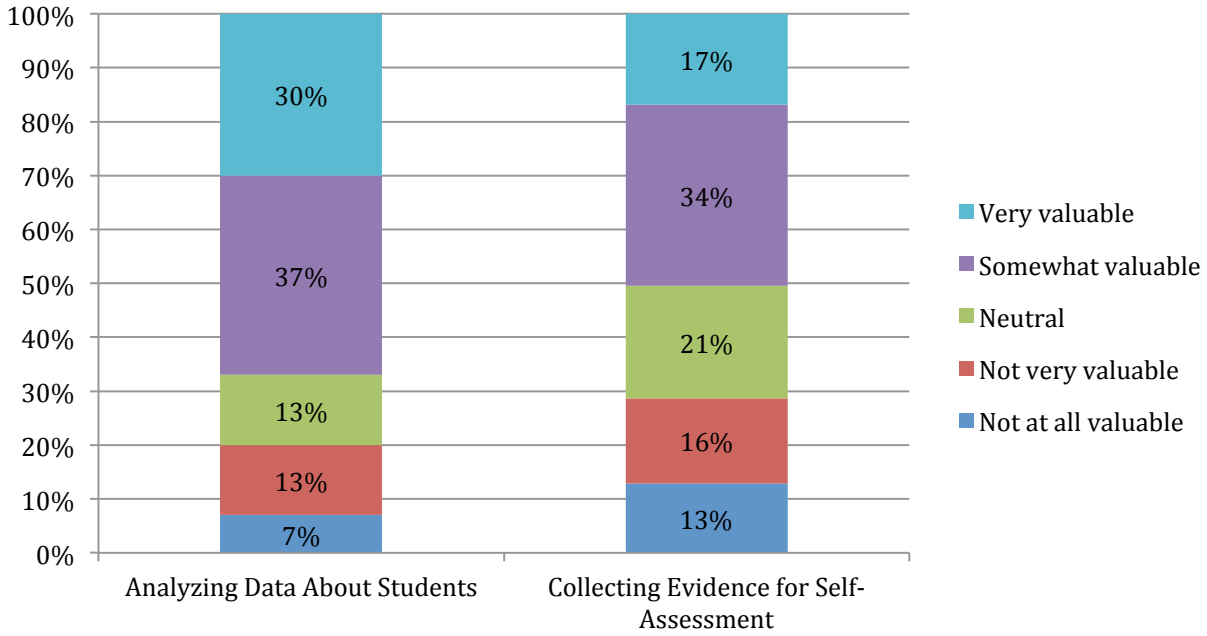


Figure 22. Value of Additional Time Spent Using Data (Fall 2013 Survey, *n*=533)



Overall, administrators reported that the time they spent on most SEED tasks was somewhat valuable (see Table 14). These tasks include setting goals with teachers, analyzing student data with teachers, conducting mid-year conferences, observing teachers, talking with teachers about their practice after observations, discussing teachers' performance with them during summative meetings, and constructing professional development plans. Administrators, on the whole, found the time they spent writing up observations and generating summative ratings to be neither valuable nor not valuable.

Table 14. Distribution of Time Spent on Evaluation and Development Tasks (Fall 2013 Survey, *n*=18)

		95% CI	
	*Mean**	Lower	Upper
Goal Setting with Teachers	4.00	3.60	4.40
Analyzing Student Data with Teachers	3.88	3.52	4.24
Holding Mid-Year Conferences	3.88	3.48	4.28
Observing Teachers	4.18	3.61	4.75
Talking about Teacher Practice with Teachers After Observations	3.94	3.51	4.37
Writing Up Observations	4.06	3.34	4.78
Generating Teachers' Summative Ratings	4.38	3.84	4.91
Discussing Teachers' Performance with Them in Summative Meetings	3.82	3.49	4.16
Constructing Professional Development Plans	3.65	3.32	3.97

* On a scale where: 1 = A lot less time than last year; 2 = A bit less time than last year; 3 = About the same amount of time as last year; 4 = A bit more time than last year; 5 = A lot more time than last year.

These findings suggest that when implemented with fidelity, SEED's evaluation activities --post-observation debriefs and data analysis, in particular—are valued by teachers.

Variations

On the whole, we found significant differences in teachers' experiences by tenure status, school level, and district.

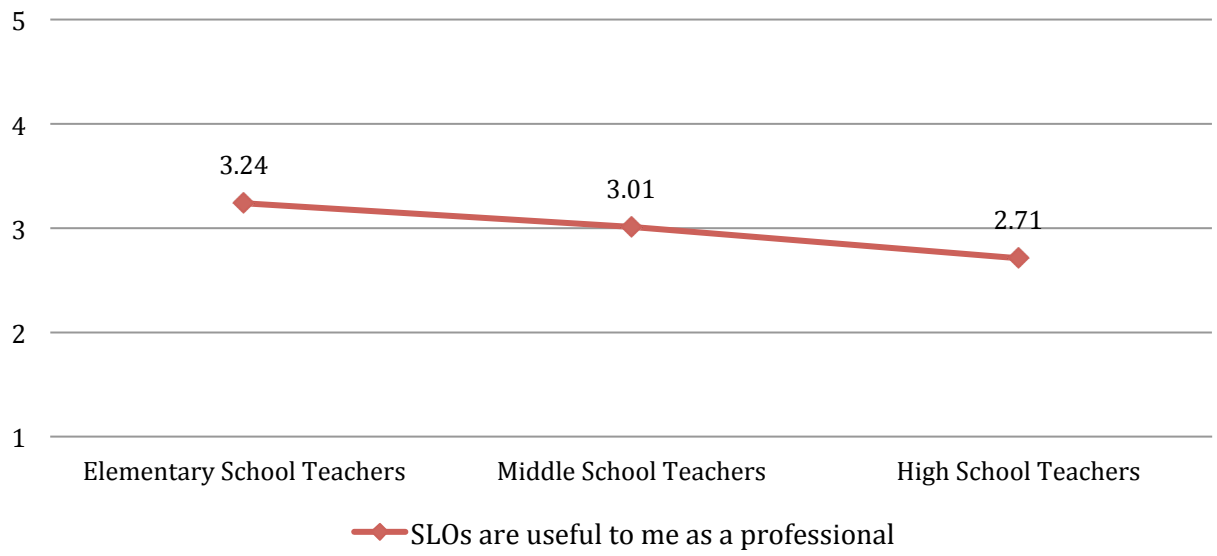
Variations by Grade Level

In surveys, we found differences in teacher experiences by grade level. Elementary teachers reported more positive views on SEED's components than did their middle school or high school counterparts. Across the three types of schools, teachers reported significant differences in the extent to which Student Learning Objectives (SLOs) were useful to them as professionals, F (2, 418) = 10.46, $p < .001$, and the extent to which feedback from classroom observations was useful to them as professionals, F (2, 416) = 8.70, $p < .001$ (see Figures 23 & 24). These findings are consistent with a long line of research suggesting that reforms take root more easily in

elementary schools than their secondary counterparts. Moreover, secondary schools' larger size and departmental structure may present particular challenges to the implementation of teacher evaluation systems.

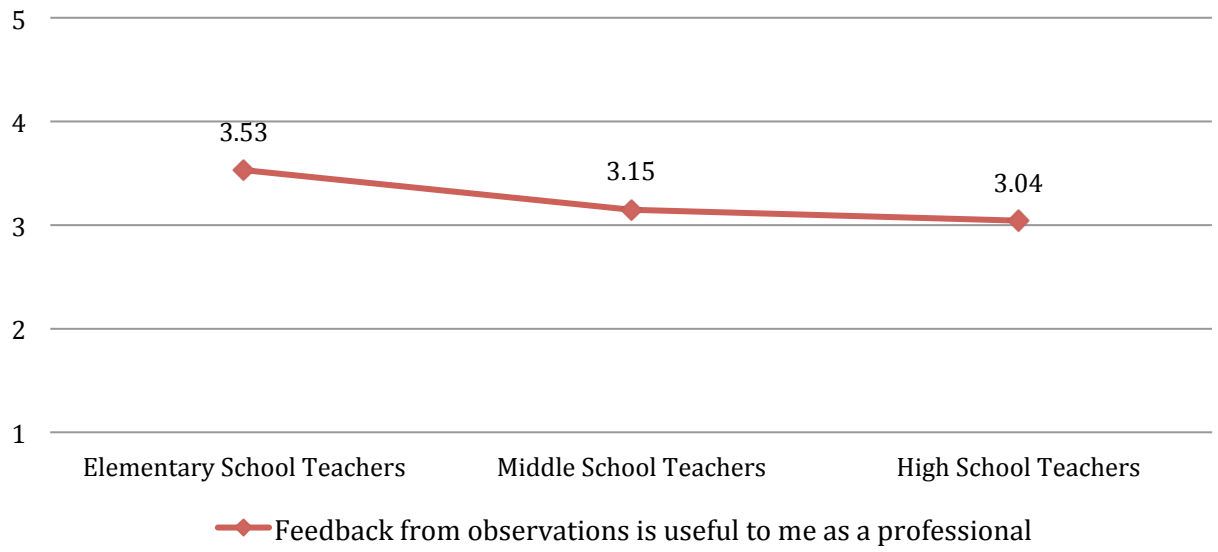
Dunnett's T3 post-hoc comparisons of the three grade levels indicate that elementary school teachers (*M*=3.24, 95% CI [3.10, 3.39]) found SLOs to be significantly more useful than high school teachers (*M*=2.71, 95% CI [2.52, 2.89]), $p < .001$. Comparisons between middle school teachers (*M*=3.01, 95% CI [2.80, 3.22]) and teachers at the other two grade levels were not significant at $p < .05$.

Figure 23. Means Plot for Usefulness of SLOs across Grade Levels (Fall 2013 Survey, *n*=533)



Post-hoc comparisons also indicate that elementary school teachers (*M*=3.53, 95% CI [3.39, 3.68]) found feedback from classroom observations to be significantly more useful than middle school teachers (*M*=3.15, 95% CI [2.91, 3.39]), $p = .022$, and high school teachers (*M*=3.04, 95% CI [2.84, 3.24]), $p < .001$. Middle school teachers and high school teachers did not indicate significant differences in responses at $p < .05$.

Figure 24. Means Plot for Usefulness of Feedback from Observations across Grade Levels (Fall 2013 Survey, *n*=533)

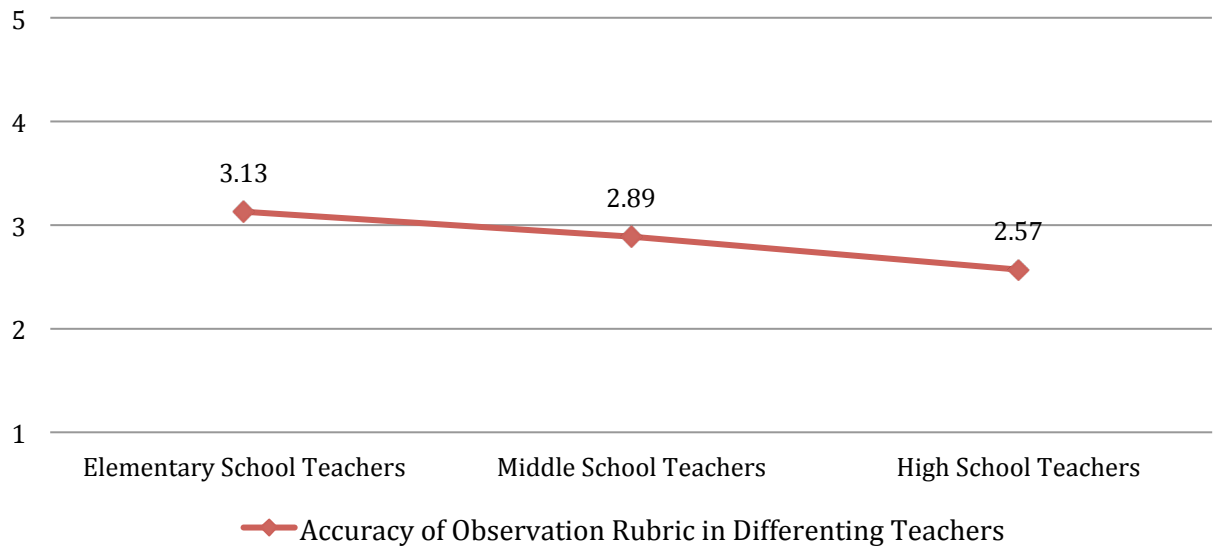


Moreover, a one-way ANOVA tested whether there were significant differences in teacher perceptions of the validity of SEED measures across grade levels. There were no differences in the extent to which teachers indicated that they feel their summative ratings from 2012-13 were accurate, $F(2, 408) = .01$, $p = .990$.

Teachers reported significant differences in the extent to which they believed that the observation protocol accurately describes a continuum of teaching quality, $F(2, 408) = 9.58$, $p < .001$ (see Figure 25). Tukey's post-hoc comparisons indicate that elementary school teachers (M=3.13, 95% CI [2.97, 3.29]) agree significantly more with this statement than high school teachers (M=2.57, 95% CI [2.37, 2.77]), $p < .001$. The difference between middle school teachers (M=2.89, 95% CI [2.67, 3.11]) and high school teachers is statistically significant at the .10 level ($p = .078$); the difference between elementary school teachers and middle school teachers is not significant at $p < .05$.

One-way ANOVAs indicate that there were significant differences across grade levels in the extent to which teachers valued the time spent on goal setting, being observed, talking about their practice with their principal/evaluator after observations, analyzing data about their students, collecting evidence to support self-assessment of professional growth, receiving summative ratings, and discussing summative ratings. In each of these teacher evaluation and development tasks, elementary school teachers reported valuing the task significantly more than high school teachers; middle school teachers also valued time spent on goal setting, analyzing data about their students, and collecting evidence to support self-assessment of professional growth significantly more than high school teachers. There were no significant differences between elementary school teachers and middle school teachers in the extent to which they valued time spent on any of these activities.

Figure 25. Means Plot for Teacher Perceptions of SEED Rubric Validity across Grade Levels (Fall 2013 Survey, *n*=533)



Differences across Tenure Status

Independent-samples *t* tests indicate that there was no difference between non-tenured and tenured teachers in the extent to which they believed the observation rubric accurately describes a continuum of teaching quality, t (429) = 1.42, p = .156, or the extent to which they felt their summative rating from 2012-13 was accurate, t (424) = .92, p = .360.

Non-tenured teachers found, overall, that their time spent on SEED activities was more valuable than tenured teachers, with the exception of the time spent on analyzing data, for which there was no significant difference. The average non-tenured teacher found goal setting, being observed, talking about their practice with the principal/evaluator after being observed, collecting evidence to support self-assessment of professional growth, and discussing their summative rating to be "somewhat valuable," whereas tenured teachers gave these activities a neutral mean score. This may be due to the fact, discussed in prior sections as well as below, that non-tenured teachers received a larger "dose" of SEED; specifically, they received more observations and more conferences, on average.

Moreover, on the spring survey, tenured teachers reported significantly less confidence in their evaluator's ability to evaluate them fairly than non-tenured teachers, with $t(548) = 4.66$, $p < .001$. Only 21% of tenured teachers agreed that their evaluator had sufficient knowledge, time, and resources to implement SEED, compared to 39% of non-tenured teachers. This may be due to the fact that principals observed non-tenured teachers more than tenured teachers, on average. For example, as of spring 2013, the majority of tenured teachers (73%) reported that they had been formally observed once or not at all, compared to only 43% of non-tenured teachers – the majority of whom had been formally observed twice or more.

Variations by District

Overall, we saw differences in how teachers viewed SEED across different districts. There were significant differences across districts in many indicators of teacher experiences, including how well teachers understood how they were evaluated under SEED [$F(8, 466) = 3.13$, $p = .002$], the extent to which they believe the observation rubric describes a continuum of teaching quality [$F(8, 439) = 5.49$, $p < .001$], the value of talking about their practice with their evaluator after being observed [$F(8, 452) = 3.52$, p = .001], and the extent to which teachers felt that their summative ratings from 2012-13 were accurate [$F(8, 434) = 2.95$, $p = .003$]. There was one pilot site with particularly higher mean ratings of teacher experiences and a few sites with particularly lower mean ratings across these four indicators, suggesting that there is a consistent effect related to district context.

Specialists raise strong concerns about SEED

At each stage of SEED implementation, specialists raised strong concerns about SEED. Most specialists, including psychologists, interventionists, instructional coaches, music teachers, and social workers, felt that the instructional practice rubric did not capture many aspects of their work. Moreover, many specialists reported crafting SLOs based on a small group of students and sometimes even one student. They raised questions about whether they should be assessed based on the progress of such a small number of students. Specialists were also confused about how to draft SLOs, and they spent a great deal of time trying to write them. These educators were especially interested in receiving examples of SLOs. Overall, specialists expressed a strong desire that SEED be differentiated to meet their needs as educators whose positions are structured quite differently from those of teachers of core subjects.

Administrator SEED

On the whole, school administrators reported varied responses on SEED's accuracy as a tool to measure administrator performance, but generally viewed their SEED ratings as accurate. On average, the administrators in our sample neither agreed nor disagreed that SEED accurately measures the performance of administrators (M=3.34, 95% CI [2.64, 4.07]), that the indicators of school leadership quality included in SEED capture what effective administrators do (M=2.76, 95% CI[2.17, 3.36]), or that their summative ratings from 2012-13 are accurate (M=3.29, 95% CI[2.83, 3.76]). However, there was a range of responses to these survey items. Specifically, 30% of administrators surveyed agreed or strongly agreed that SEED accurately measures administrator performance; 35% of respondents were neutral on this question; and 36% of respondents disagreed/strongly disagreed (see Table 15). Forty-seven percent of respondents indicated that SEED's indicators capture what effective administrators do, while just 18% disagreed/strongly disagreed on this item. In contrast to these fairly evenly distributed responses, 59% of administrators reported agreeing or strongly agreeing that their final SEED rating was accurate.

Table 15. Frequencies for Administrators' Perceptions of the Validity of SEED Measures (Fall 2013 Survey, *n*=22)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
I believe SEED accurately measures the performance of administrators.	24%	12%	35%	24%	6%
The indicators of school leadership quality included in SEED capture what effective administrators do.	6%	12%	35%	41%	6%
I feel that my summative rating from 2012-2013 is accurate.	24%	0%	18%	35%	24%

We asked school administrators additional questions about their views on SEED for administrators. On average, administrators responded on four indicators that they believed SEED would influence practice neutrally: the administrators we surveyed neither agreed nor disagreed, on the whole, that they received appropriate levels of support based on their summative ratings, that they received appropriate levels of recognition based on their summative ratings, that they felt more pressure to have students make academic progress while using SEED, or that the SEED model is going to help administrators improve their leadership practice (see Table 16).

Table 16. Distribution of Administrator Outcomes of SEED Implementation (Fall 2013 Survey, *n*=22)

			95% C.I.	
	N	*Mean*	Lower	Upper
I feel that I have received appropriate levels of support based on my summative rating.	17	3.47	2.84	4.10
I feel that I have received appropriate levels of recognition based on my summative rating.	17	3.29	2.67	3.92
I have felt more pressure to have students make academic progress since we began using SEED.	17	2.94	2.30	3.58
The SEED model is going to help administrators improve their leadership practice.	17	3.29	2.67	3.92

Closing Remarks

On the whole, we found that participants holding a range of roles expressed a positive or neutral view of the SEED model. Very few participants rejected the entire model out of hand. Teachers and school leaders found observations, post-observation conferences, and analyzing student data to be valuable aspects of SEED. The state and districts should build upon these early positive results, bolstering these aspects of SEED while responding to educators' criticisms of other components of the model.

Findings on SEED Implementation Outcomes

Consistent with our charge, we examined the first-year implementation of SEED within the 14 pilot districts. In this section, we discuss our findings related to the preliminary outcomes of the SEED pilots for teachers and school administrators. Outcomes from a pilot model should be interpreted with due caution. A pilot program is intended to test a model and is not typically implemented with complete fidelity. The results of evaluations of pilot programs are meant to inform the system, rather than to permit summative conclusions about the program. We present data that respond to the following questions:

OUTCOMES OF SEED

1. To what extent did educators report changing their practices as a result of the SEED?
2. To what extent was there variation in evaluation ratings within schools and districts at the conclusion of the SEED pilot year?
3. What variations occurred? What explains these variations?

In general, we found that some teachers and leaders reported changed practice as a result of SEED. Elementary school and non-tenured teachers were significantly more likely to report that SEED had resulted in changes to their practice or had the potential to do so in the future. School administrators reported mixed views on the extent to which SEED had changed their practice or would do so in the future. Summative performance ratings indicate the majority of teachers were rated as *proficient* (73%) or *exemplary* (23%), with a much smaller percentage rated as *below standard* (<1%) or *developing* (4%). Some variation in these figures occurred across districts, with the majority of variation existing between the top two rating categories.

In the sections below, we first discuss educators' views on whether teachers and leaders changed their practices as a result of participating in SEED. We then discuss variations in reported changes to practice. We then turn to summative ratings, presenting aggregate ratings distributions and their variations. We conclude with a summary.

Changes in Teacher Practice

One of the main goals of SEED is to improve teachers' and school leaders' practice. In fact, this goal is enshrined in initiative's title: the System for Educator Evaluation and Development. At the conclusion of the pilot year, participants reported modest changes to teacher practice as a result of SEED.

Teachers' Views on SEED's Effects on Their Practice

Teachers expressed mixed views on whether SLOs changed their practice, but were more positive about the influence of observations on their teaching.

As shown in Table 17, 36% of teachers agreed or strongly agreed that setting SLOs led them to make changes in their teaching practice while 39% disagreed or strongly disagreed. On the

same item, 55% of administrators agreed or strongly agreed that SLOs led teachers to change their practice.

Related to goal setting, 34% of teachers reported that they covered less content in 2012-13 in order to focus on their content and skills embedded in their SLOs. Twenty-eight percent were neutral on this question and 38% disagreed/strongly disagreed with this statement. Twenty-six percent of teachers reported that they felt more accountable for student performance because of SEED, whereas 48% disagreed or strongly disagreed with this statement.

Table 17. Teacher Views on SEED (Fall 2013 Survey, *n*=480)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
Setting SLOs led me to make changes in my teaching practice.	12%	27%	26%	32%	4%
Feedback from classroom observations was useful to me as a professional.	10%	12%	23%	47%	8%
Feedback from classroom observations led me to make changes in my teaching practice.	10%	15%	31%	37%	7%
Changes in my practice as a result of SEED were positive.	10%	14%	41%	31%	4%
I felt more accountable for growth in my students' learning under SEED than I did in previous years.	21%	27%	27%	21%	5%
I covered less content than usual when teaching in order to focus on the content embedded in my student performance goals.	9%	29%	28%	23%	11%

Compared to their views on SLOs, teachers were more positive about the effects of observation on their practice. Almost half of teachers surveyed (44%) agreed or strongly agreed that feedback from observations prompted them to change their practice while 25% disagreed or strongly disagreed. This may be related to the fact that a relatively large percentage (55%) reported that feedback from observations were useful to them as professionals, while 23% disagreed that this was the case.

Of course, a teacher's practice can change for the better or the worse. When asked whether the changes brought about by SEED were positive, 35% of teachers surveyed agreed/strongly agreed, while 41% were neutral on this question, and 24% disagreed/strongly disagreed.

In interviews, teachers expressed similarly mixed views on SEED's impact on their instruction. Some reported focusing disproportionately on the content and skills embedded in their SLOs, as reflected in the survey findings reported above. Some of these teachers felt this was a good thing, as their SLOs reflected the most important content and skills for their students to master. Others felt this constrained their teaching and made it more difficult for them to address other content and skills with sufficient depth. Some teachers felt that the rubric helped focus their teaching, giving them structure to evaluate their own instruction. Others, particularly specialists, felt that the rubric was unrealistic given their students' needs and skills. On the whole, teachers reported that SEED had caused them to change their practices slightly.

Looking to the future, teachers expressed some optimism that they would receive support based on their evaluation ratings, and somewhat less faith that they would receive recognition based on these outcomes (Table 18). Almost half (43%) agreed or strongly agreed with the statement that they would receive appropriate levels of support based on their summative rating, while 33% were neutral on this question and almost a quarter disagreed/strongly disagreed with it. Less than a third (29%) agreed/strongly agreed that they would be appropriately recognized for their rating; this lower number could be due to the fact that it was a pilot year or that only teachers who received relatively high ratings felt they would be recognized.

Table 18. Teacher Views on Outcomes of SEED Implementation (Fall 2013 Survey, *n*=480)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
I feel that I will receive appropriate levels of support based on my summative rating.	10%	14%	33%	35%	8%
I feel that I will receive appropriate levels of recognition based on my summative rating.	14%	21%	36%	24%	5%
With sufficient resources, such as time and staffing, we would be able to use SEED to improve teacher practice at this school.	11%	13%	34%	37%	5%

On one of the most important items, 42% of teachers agreed or strongly agreed that with sufficient resources school staff could use SEED to improve teacher practice at their school. This item probes teachers' optimism about SEED's impact in their setting. One third of teachers were neutral on this question, and 24% of teachers surveyed disagreed/strongly disagreed. Below, we explore these differences in more detail.

Administrator Views on SEED's Effects on Teacher Practice

Leaders' views on whether SEED changed teacher practice were similarly varied.

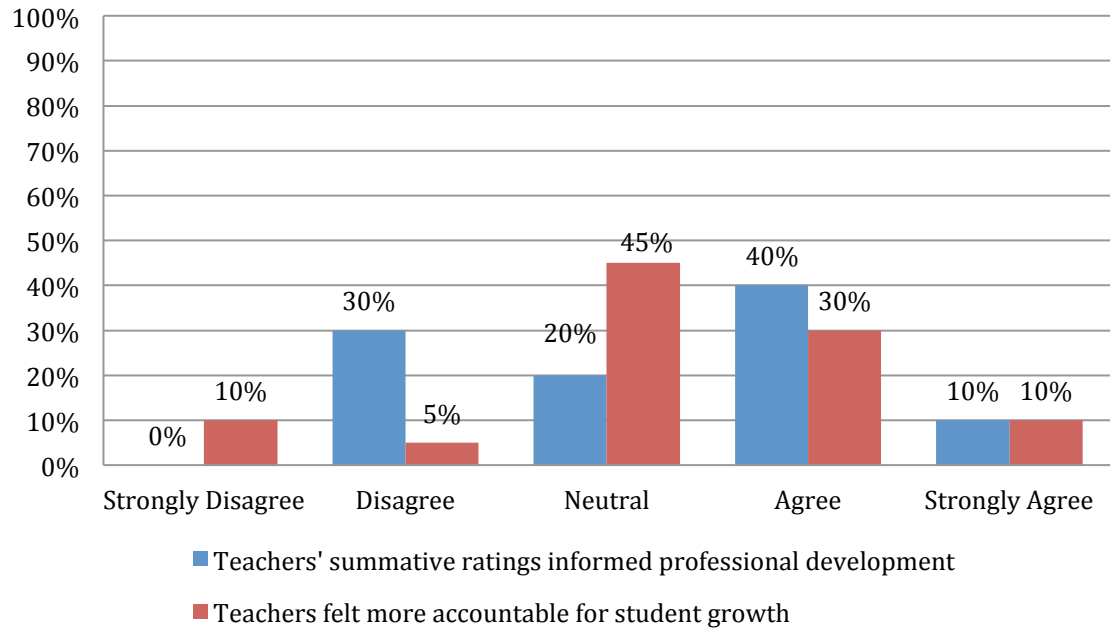
On average, administrators rated five indicators that they believed SEED would influence practice neutrally, neither agreeing nor disagreeing with each item (see Table 19).

Table 19. Distribution of Administrator Perceptions of SEED Outcomes (Fall 2013 Survey)

			95% C.I.	
	n	*Mean*	Lower	Upper
Teachers' summative ratings have been used to inform individual-level professional development.	20	3.11	2.68	3.54
Teachers felt more accountable for growth in their students' learning under SEED than they did prior to SEED.	20	3.06	2.61	3.50
Teachers at this school changed what they were teaching students in order to align with the specific goals for their evaluation.	19	3.00	2.56	3.44
Teachers focused heavily on the content embedded in their student performance goals.	20	3.28	2.92	3.64
Setting SLOs led teachers to make changes in their teaching practice.	20	3.17	2.71	3.62

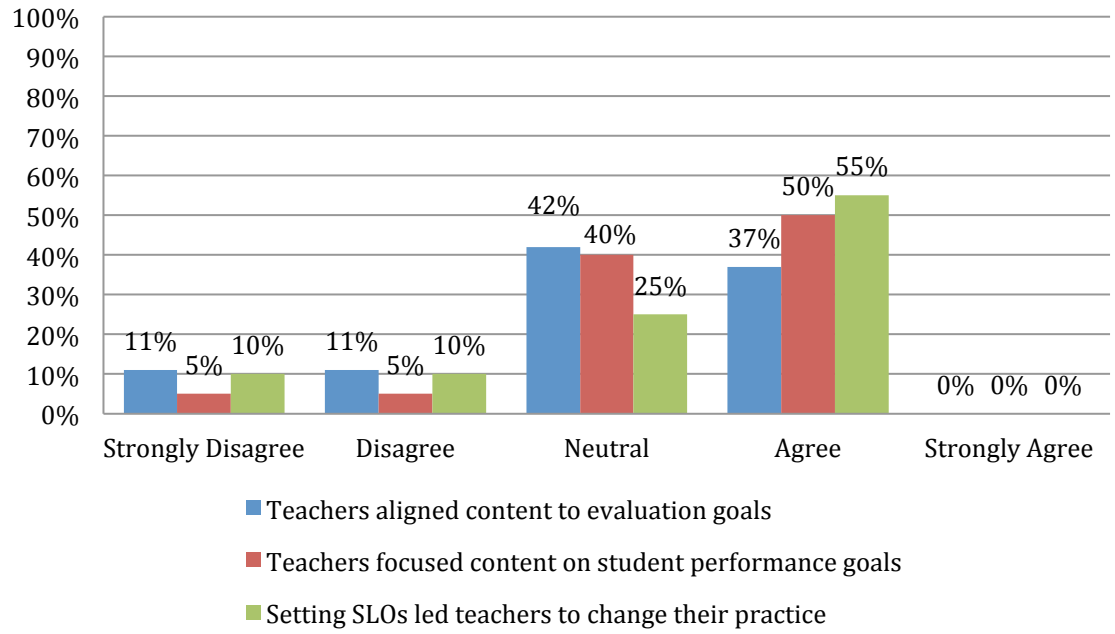
However, a substantial subset of administrators reported changes that they attribute to SEED (see Figures 26 and 27). Half (50%) of the administrators surveyed indicated that teachers' summative ratings have been used to inform individual-level professional development and 40% indicated that SEED caused teachers to feel more accountable for growth in student learning.

Figure 26. Administrator Views on the Effects of SEED on Professional Development and Accountability (Fall 2013 Survey, *n*=19)



Many administrators also reported that teachers altered their instruction and/or curricular content to fit their evaluation goals under SEED (Figures 26 and 27). A little more than one-third (37%) of administrators indicated that teachers at their schools changed what they were teaching students in order to align with the specific goals for their evaluation. Half (50%) indicated that teachers focused heavily on the content embedded in their student performance goals and 55% indicated that setting SLOs led teachers to make changes in their teaching practice.

Figure 27. Administrator Views on the Effects of SEED on Teachers' Instructional Practice



School leaders also weighed in on SEED's potential to improve teacher practice. The majority (53%) agreed or strongly agreed that the SEED model will help teachers improve their practice (Table 20). Almost three-quarters (74%) agreed or strongly agreed that SEED could improve teachers' practice with sufficient resources (Table 20). Although we caution readers due to the small sample size of principal respondents, recall that 42% responded similarly to the same item on the teacher survey.

Table 20. Administrator Views on the Effect of SEED on Teacher Practice (Fall 2013 Survey, *n*=19)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
The SEED model will help teachers improve their practice.	5%	21%	21%	42%	11%
With sufficient resources, such as time and staffing, we can use SEED to improve teacher practice at this school.	11%	5%	11%	53%	21%

Interviews corroborated these survey findings. Some principals reported that they observed changes in teachers' practices due to SEED. Two principals in different districts reported that SEED's instructional practice rubric had helped facilitate communication between school leaders and teachers. One principal added that he felt that SEED had helped teachers to know better how to analyze and use data.

A principal in a struggling school reported that SEED had had an impact on the lower performing teachers in his school. He reported: “at a basic level, some people for the first time saw the correlation between student performance and instruction.” Similarly, another principal perceived a high degree of focus and ownership of student achievement by the teachers. To support this assertion, she cited the fact that DRA scores in the school demonstrated growth across the board unlike any previous year.

Several principals reported that teachers learned how to better use data within the context of evaluation. One principal reported that teachers paid “heightened attention” to their data. Another principal reported that teachers started to understand SLOs and the importance of having goals, assessments, and data that are aligned. Some teachers in his school didn’t have data matched to their SLOs. At the end of the first year, the principal reported that teachers realized they needed better data. Another principal reported that teachers were accustomed to using data because their school had been a SIG school. However, under the School Improvement Grant, a school wide data facilitator had “owned” and tracked the data. In contrast, he recounted, SEED had shifted ownership of the data to individual teachers.

Other leaders reported that they felt that SEED had not changed instruction. Teachers, they reported, were not using data or assessments differently or improving their instruction. In some cases, principals reported that this was because teachers had been meeting in grade level teams and looking at data prior to SEED. They believed that SEED did not diverge substantially from prior practice. In other cases, principals reported that SEED’s late start and insufficient training meant that it had little influence on the practice of teachers. This was amplified in districts with multiple other initiatives underway or in schools with administrator turnover.

Most district leaders expressed optimism that SEED would lead to changes in teachers’ practice. “Ultimately it will change teaching,” said one leader. She added, “practice hasn’t changed yet but I think it will. They have a heightened awareness of the components of good instruction. Their awareness is there,” and change, she predicted, would follow.

Changes in Leaders’ Practices

On average, administrators rated four indicators that they believed SEED would influence practice neutrally: the administrators we surveyed neither agreed nor disagreed, on the whole, that they received appropriate levels of support based on their summative ratings, that they received appropriate levels of recognition based on their summative ratings, that they felt more pressure to have students make academic progress while using SEED, or that the SEED model is going to help administrators improve their leadership practice (see Table 21).

Table 21. Administrator Views on the Outcomes of SEED Implementation: Mean Scores (Fall 2013 Survey)

			95% C.I.	
	n	*Mean*	Lower	Upper
I feel that I have received appropriate levels of support based on my summative rating.	17	3.47	2.84	4.10
I feel that I have received appropriate levels of recognition based on my summative rating.	17	3.29	2.67	3.92
I have felt more pressure to have students make academic progress since we began using SEED.	17	2.94	2.30	3.58
The SEED model is going to help administrators improve their leadership practice.	17	3.29	2.67	3.92

Put another way, school leaders were divided on whether they thought SEED could improve their practice. Thirty percent of school leaders agreed or strongly agreed that SEED is going to help administrators become better leaders, while 41% were neutral on this question and 30% disagreed/strongly disagreed (Table 22). Administrators' divided responses could be related to their perception that their evaluators have the knowledge but not the time or resources to evaluate them accurately. While 71% of school administrators felt their evaluators possessed this knowledge, only 36% thought their evaluators had the time and resources to carry out evaluation as SEED prescribes (Table 22).

Table 22. Administrator Views on the Outcomes of SEED Implementation: Frequency Scores (Fall 2013 Survey)

	Strongly Disagreed	Disagreed	Neutral	Agreed	Strongly Agreed
The SEED model is going to help administrators improve their leadership practice.	18%	12%	41%	24%	6%
My evaluator(s) had the knowledge to evaluate me accurately according to the SEED model.	18%	0%	12%	47%	24%
My evaluator(s) had the time and resources to evaluate me accurately according to the SEED model.	18%	18%	29%	24%	12%

In interviews, school leaders reported slight changes in their practice as a result of participating in SEED. One principal reported that SEED helped her focus her progress monitoring with teachers, enabling her to ask teachers about individuals or target groups. Another principal reported that observations made him more aware of the range of practices in his building. He said his "eyes were opened" to how his teachers think. SEED has made him "a better observer" and opened up many more conversations with teachers about instruction, data, and students.

District administrators also reported that SEED had been somewhat beneficial to them. One assistant superintendent said that SEED "helped me to become clearer about what principals should look for in the classroom…I've learned a lot."

Variations

Consistent with findings presented in the previous sections on implementation and educator experiences, we found significant differences in teachers' views. On average, teachers in elementary schools and non-tenured teachers expressed more positive views on SEED's current and potential effects on their practice based on their participation in SEED's pilot year.

Differences across Grades in Teachers' Views

Teachers reported significant differences across grade levels in the extent to which they felt they would receive appropriate levels of recognition based on their summative ratings, F (2,409) = 4.79, p = .009 (see Figure 28), and the extent to which they believe that with sufficient resources, such as time and staffing, they would be able to use SEED to improve teacher practice at their schools, F (2, 411) = 4.35, p = .014 (see Figure 29).

Tukey's post-hoc comparisons indicate that elementary school teachers (M=3.03, 95% CI [2.88, 3.18]) believed they would receive appropriate levels of recognition based on their summative ratings more than did high school teachers (M=2.65, 95% CI [2.46, 2.86]), p = .007. The differences between middle school teachers (M=2.80, 95% CI [2.57, 3.03]) and teachers at elementary school and high school levels were not significant at $p < .05$.

Elementary school teachers (M=3.25, 95% CI [3.09, 3.40]) also believed that with sufficient resources, such as time and staffing, they would be able to use SEED to improve teacher practice at their school more than high school teachers did (M=2.92, 95% CI [2.73, 3.10]), p = .018. The difference between middle school teachers (M=3.23, 95% CI [3.02, 3.44]) and high school teachers was nearly statistically significant, with p = .063. The difference between elementary school teachers and middle school teachers was not significant at $p < .05$.

Figure 28. Means Plot for Teacher Beliefs about Receiving Recognition across Grade Levels

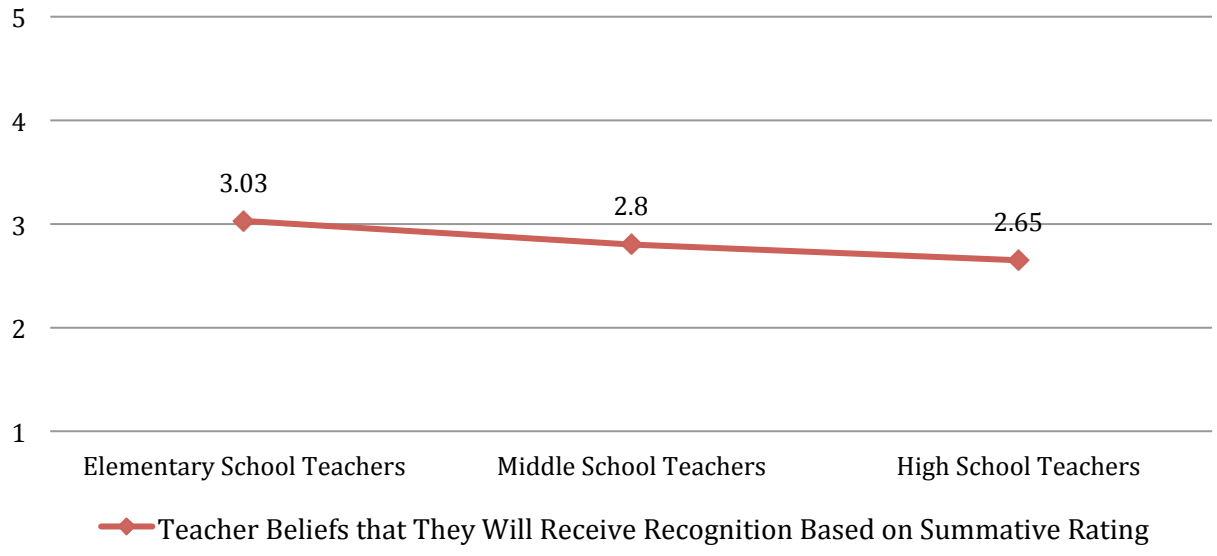
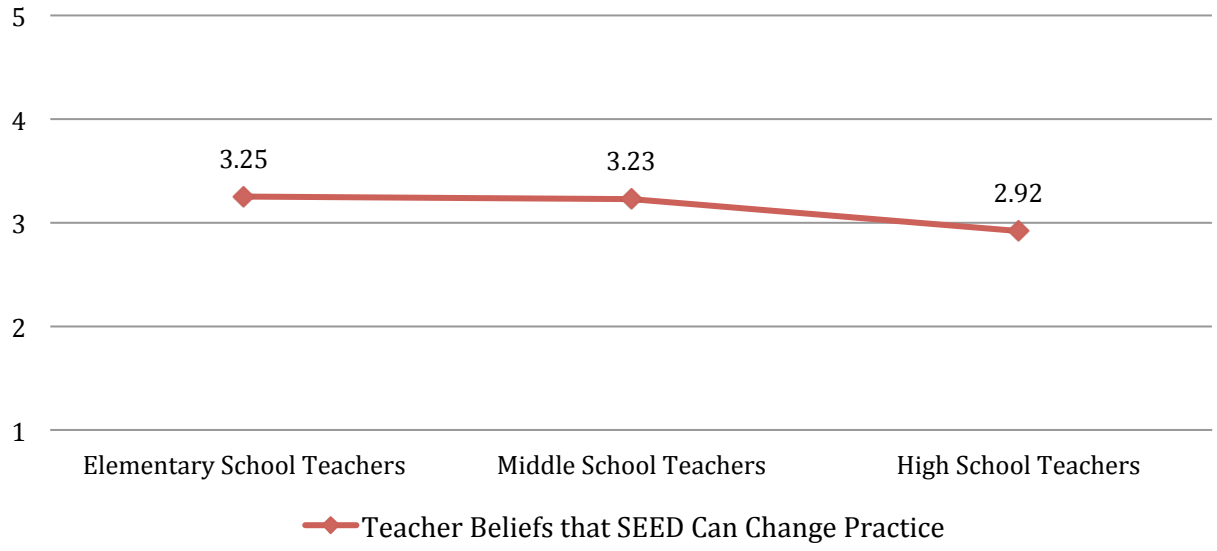


Figure 29. Means Plot for Teacher Beliefs that SEED Can Change Practice across Grade Levels



One-way ANOVAs indicated no significant differences across grade levels in the extent to which teachers felt they would receive appropriate levels of support based on their summative ratings, F (2, 411) = 1.17, p = .141; felt more accountable for growth in their students' learning under SEED than in previous years, F (2, 417) = 2.82, p = .061; or covered less content than usual in order to focus on the content embedded in their student performance goals, F (2, 414) = .14, p = .871.

Differences across Tenure Status

Non-tenured teachers indicated more positive beliefs about SEED's ability to influence practice for all indicators except for the extent to which they covered less content than usual in order to focus on the content embedded in their student performance goals (see Table 23).

Table 23. Differences in SEED Implementation Outcomes by Tenure Status

	Non-Tenured				Tenured							
	n	*Mean*	95% CI		*n*	*Mean*	95% CI		*df*	*t*	*p*	*r*
I feel that I will receive appropriate levels of support based on my summative rating.	80	3.44	3.23	3.65	349	3.08	2.97	3.20	427	2.62	.009	.13
I feel that I will receive appropriate levels of recognition based on my summative rating.	78	3.10	2.89	3.32	349	2.78	2.66	2.90	425	2.37	.018	.11
I felt more accountable for growth in my students' learning under SEED than I did in previous years.	80	2.91	2.67	3.16	356	2.53	2.41	2.65	121	2.79	.006	.25
I covered less content than usual when teaching in order to focus on the content embedded in my student performance goals. (REVERSED survey item)	81	3.07	2.84	3.31	351	3.02	2.89	3.14	430	0.40	.690	.02
With sufficient resources, such as time and staffing, we would be able to use SEED to improve teacher practice at this school.	80	3.45	3.26	3.64	349	3.05	2.94	3.17	427	3.05	.002	.15

Differences by District

Similar to our findings for implementation and experience, we again found variations in participants' views on whether SEED had changed educator practice or had the potential to do so in the future. Mean teacher response varied significantly by district on whether they felt more accountable for student learning growth under SEED; covered less content under SEED than in previous years; and felt they would receive appropriate support or recognition based on their summative evaluation rating. District-level means also varied significantly on whether,

with appropriate resources, SEED would help schools improve teacher practice. On the whole, districts that worked on linking evaluation with professional development in recent years recorded higher means on these items than did their counterparts with less experience in this area.

Ratings

Per SEED, each educator receives a summative, annual performance rating based on scores on the Teacher Practice Related Indicators and Student Outcomes Related Indicators. District aggregate evaluation ratings are reported to the Commissioner of Education by June 30 of each year. Summative teacher evaluation ratings from participating pilot districts are presented in Table 24. Due to incomplete or missing results, administrator ratings were not available. We caution readers against over-interpreting these data, as they represent scores from SEED's pilot year. Two of the fourteen pilot districts did not submit evaluation ratings. One of the participating districts was able to report ratings for only approximately 60% of its teachers. We did not have access to disaggregated data, such as by school, by school level (e.g., high school), by teacher tenure status or subject area.

On average, less than 1% of teachers were rated as *below standard* and 4% were rated as *developing* in the pilot districts.[5] Nearly three-quarters (73%) were rated *proficient* and 23% were rated as *exemplary*. Three-quarters (9 of 12) of the districts reported no teachers rated as *below standard.* There was some variation in aggregate ratings across districts, with small districts subject to greater fluctuation. Excluding one anomalous district (District A), the percentage of teachers rated *proficient* ranged between 56% and 92%. Figure 30 presents the numerical data via line plots, and indicates a general pattern of most teachers (96%) being rated *proficient* or *exemplary,* with the most variability across districts occurring between these two categories.

[5] These two figures may underestimate actual numbers because, for legal purposes, pilot districts tended to use prior evaluation systems for those teachers previously performing well below standard. It is unclear from our data how many, if any, teachers were excluded from these pilot ratings.

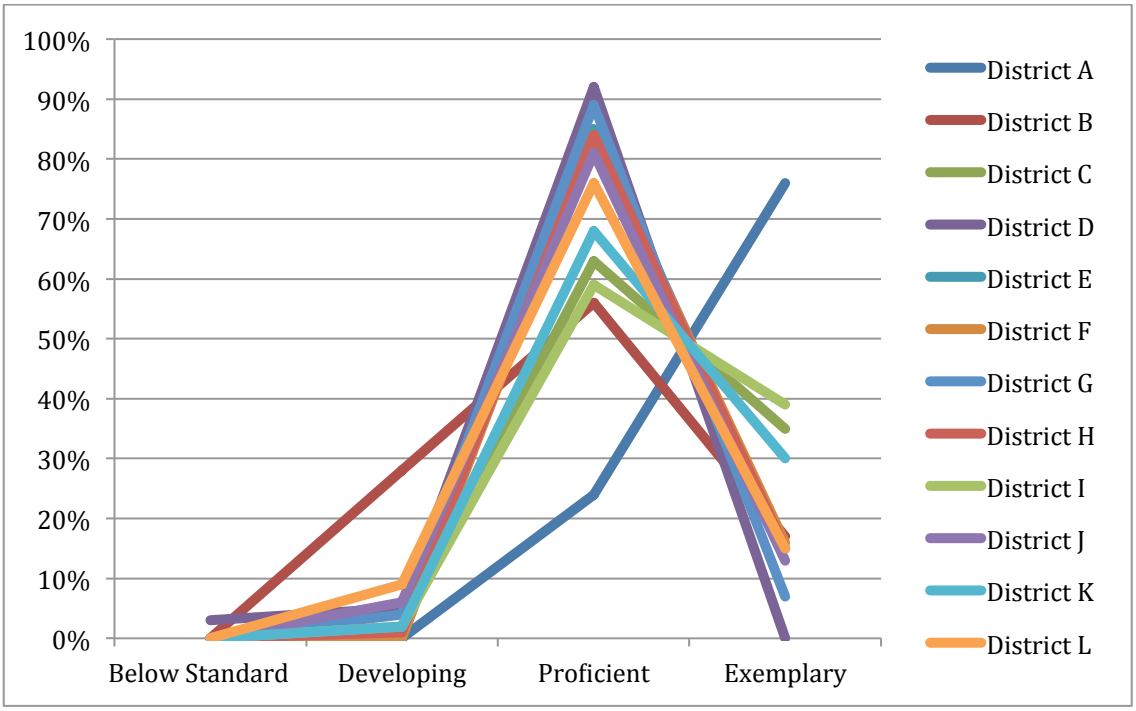
Table 24. Summary of Teacher Evaluation Ratings in SEED Pilot Districts, 2012-2013

	Below Standard	*Developing*	*Proficient*	*Exemplary*
District A	0%	0%	24%	76%
District B	0%	28%	56%	17%
District C	0%	2%	63%	35%
District D	3%	5%	92%	0%
District E	0%	0%	85%	15%
District F	0%	0%	84%	16%
District G	0%	4%	89%	7%
District H	0%	1%	84%	15%
District I	0%	2%	59%	39%
District J	0%	6%	81%	13%
District K	<1%	2%	68%	30%
District L	<1%	9%	76%	15%
Overall	*<1%*	*4%*	*73%*	*23%*

Notes:

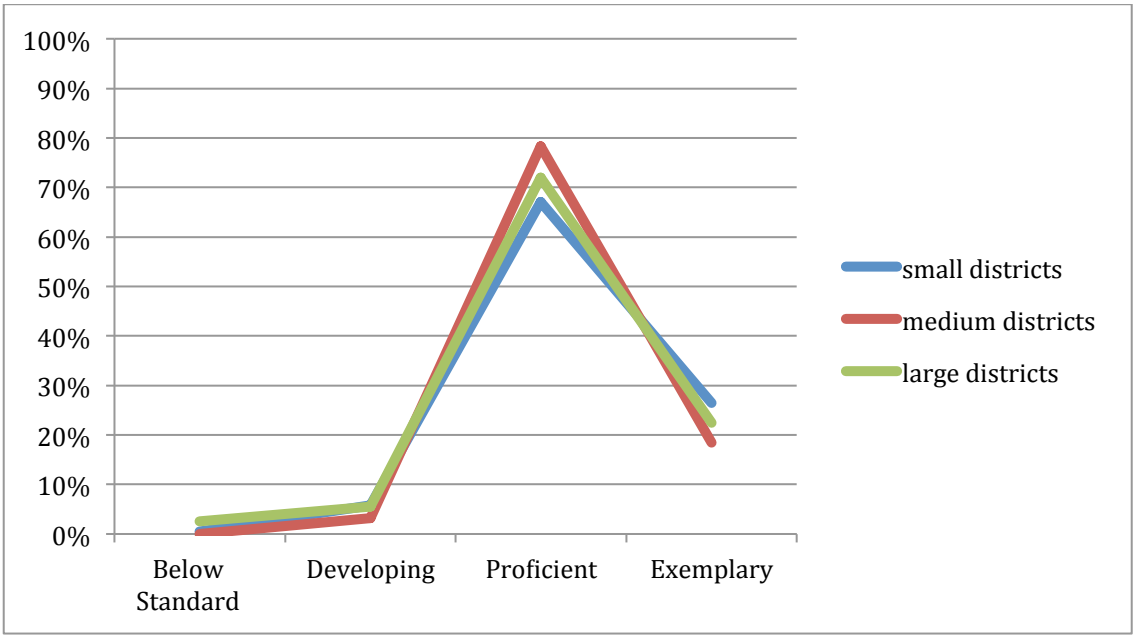
1. Incomplete evaluations (i.e., evaluations where one or more component rating was reported as "0" or "NULL" or left blank) are excluded from these data. In District I, almost 40% of the records supplied were excluded due to incomplete evaluation data.
2. These data represent final ratings as submitted by the districts. With the exception of incomplete evaluations referenced above, no additional verification of these final ratings was performed.
3. Due to rounding, row totals may not sum to 100%.

Figure 30. Line Plots of Aggregate Teacher Evaluation Ratings in SEED Pilot Districts, 2012-2013



Finally, we classified Districts A through L by size under the categories small, medium, and large. Small districts evaluated up to 100 teachers, medium districts evaluated between 101 and 250 teachers, and large districts evaluated more than 250 teachers. As Figure 31 shows, there was no discernible relationship between size of district and summative ratings among these twelve pilot districts. This finding is far from generalizable, however, given the small sample size.

Figure 31. Line Plots of Average Teacher Evaluation Ratings By District Size in SEED Pilot Districts, 2012-2013



The aggregate results above are consistent with survey and interview data, where district and school leaders reported that the majority of teachers were rated proficient. Principals reported that in most cases they were not surprised by teachers' final ratings. In instances when they were surprised, specialists whom the principals viewed as highly skilled had set high goals based on a very small sample size and were not able to meet their goals.

Principals reported that teachers generally were not dismayed by their ratings. Again, teachers who set extremely high goals that they were not able to achieve were the exception to this rule.

Most leaders reported that SEED ratings aligned with their implicit sense of teachers' performance. However, more teachers were rated "proficient" or "developing" than prior evaluation outcomes would suggest. Principals reported that the implementation of SEED had caused both teachers and principals to recalibrate their understandings of the meaning of as one principal recounted, "good, proficient, and effective." This calibration required teachers to come to terms with lower ratings than they had received in the past. That being said, one superintendent predicted that "a lot of people will score proficient because it's a pilot, slow roll out and people are gun shy [about rating teachers low]."

In some cases principals reported that SEED ratings conformed to his sense of teachers' performance in the area of classroom management but were lower than he had expected in the area of instruction. This principal reported that the standards for 'exemplary' were higher than he was accustomed to. Several other principals reported that the SEED rubric enabled them to rate low-performing teachers accurately whereas their prior evaluation system failed to anchor their evidence to an external standard (the rubric) and, as a result, they felt less justified in assigning a low but accurate rating.

What will result from the ratings? Leaders mentioned that they plan to differentiate professional development in the coming year. SEED ratings will also be used to identify who is going to be a complementary evaluator or be placed in a "quasi-leadership" position. District leaders in one district said that these teacher leadership positions are necessary for "our high flyers, for retaining these people. We need something different for our high flyers." Lastly, district leaders reported that they will use ratings to identify teachers whose performance should be more closely monitored in the coming year.

We offer one final note on how the pilot ratings data interface with time demands of conducting observations, and how they can be used by administrators for planning purposes. SEED guidelines require teachers rated as *below standard* or *developing* to receive three formal and up to five informal observations per year. The vast majority of teachers who scored *proficient* or *exemplary* should receive a combination of three formal in-class observations/reviews of practice (of which one must be a formal in-class observation). First and second-year teachers should receive three formal and three informal observations. The percentages will change across rating categories from year to year, and from school to school, but administrators can use prior year's ratings data to roughly estimate the evaluation schedule for the following year.

Closing Remarks

In summary, there is some evidence that teachers and administrators reported changed practice as a result of SEED. Elementary school and non-tenured teachers were statistically more likely to report that SEED had resulted in changes to their practice or had the potential to do so in the future. School administrators reported mixed views on the extent to which SEED had changed their practice or would do so in the future. Across pilot districts, 73% of teachers were rated as *proficient* and about one-quarter were considered *exemplary*. Less than one percent fell into the *below standard* category and about 4% were identified as *developing*. There was some variation in the ratings across districts.

Recommendations

We gathered extensive data from multiple stakeholders at three stages in the implementation of SEED's pilot program. Based on analyses of these data, we recommend that the State Department of Education carry out the following:

1. **Provide additional opportunities for all educators to learn about SEED.** There is a need for additional professional learning opportunities for all educators with respect to SEED. Professional learning needs fall into two categories: one is better understanding and implementing the technical aspects of SEED (setting goals, conducting observations, and providing post-observation feedback) and the other is improved development of educators (by providing individualized and targeted professional development). We strongly recommend that both administrators *and* teachers receive this training, rather than relying on a train-the-trainer model as was the case in the pilot year of SEED.

2. **Build the skills of evaluators, in particular.** Teacher survey and interview data indicated substantial variability in the perceived skill level of evaluators. This variance occurred within and across districts, and even within schools. Such a finding points to the need for administrators to develop and refine their supervisory skills. We suggest there be processes for identifying evaluators in need of improvement and then offering specialized training to these individuals. The SDE, along with professional organizations and regional consortia, should provide professional development to evaluators in using the CCT Rubric for Effective Teaching, conducting formal and informal observations, and providing verbal and written feedback. These professional learning opportunities are critical to the success of any teacher evaluation reform.

3. **Increase the use of complementary observers.** Educators reported that SEED places significant time demands on school administrators. We recommend that districts consider including complementary observers within their teacher evaluation systems. Including complementary evaluators not only reduces the time demands on principals and assistant principals, but also enhances the professional role of teachers by providing additional leadership roles for teachers. Under the moniker of peer assistance and review, such systems have been adopted by several high-profile districts in other states (e.g. Cincinnati, Montgomery County, MD) and have shown promising outcomes in terms of teacher performance and student learning. The SDE can help by sharing effective models with those in the field and providing the required training for complementary observers.

4. **Provide additional guidance on Student Learning Objectives and Indicators of Academic Growth and Development.** As SEED has begun in earnest, we recommend that the SDE continue to provide clear guidance on the identification of valid IAGD targets. The SMART goals heuristic is helpful although does not dictate what performance level or student growth target is both "attainable" and sufficiently challenging. Some teachers or administrators are selecting far too challenging targets while others are choosing far too

easy. The setting of IAGD growth targets is in most cases inherently arbitrary. Should 100% of students score a 70% on an exam or should 70% of students score at 100%? If half the students fall below a certain performance level at the beginning of the year, what percentage should reasonably be expected to meet it by the end of the year? Further, what is meant by growth differs based on the measures available (e.g., contrast pre and post measures from a vertically-scaled assessment with static measures of performance on a locally developed test). There are checks and balances built into the system, such as the mid-year check-ins, that are intended to provide at least one opportunity to adjust growth goals. However, the selection of IAGD targets is an area that deserves close attention as the SEED model evolves.

5. **Clarify and contextualize SEED to district and school personnel.** Teachers in pilot districts expressed trepidation over the number and magnitude of new reform initiatives they faced, and viewed teacher evaluation as adding "one more thing" to their already busy agenda. It would be helpful if educational leaders and policy makers demonstrated how educator evaluation interfaces with the implementation of other major initiatives such as the Common Core State Standards and Smarter Balanced Assessments. Moreover, make clear to districts any flexibility they have with regard to implementing SEED or the Core Requirements. Administrators are interpreting SEED with more rigidity than the model actually requires.

6. **Disseminate promising practices.** We recommend a coordinated effort by SDE and/or regional district consortia to identify and disseminate "promising practices" relative to SEED and educator evaluation. Connecticut educators represent an extraordinarily rich source of ingenuity and talent and should be actively engaged in the continuous improvement of evaluation as they adapt the model to their own district contexts and explore new processes. In some settings we found evidence of schools using SEED to promote deep conversations about teaching and learning. For example, one school piloted the use of video cameras to facilitate the evaluation of teachers. Teachers reported that this approach reduced their anxiety to "perform" in front a live observer and facilitated conversations with their evaluator by helping them understand the evaluator's feedback.

7. **Provide additional assistance to low-performing districts to support SEED's implementation.** We believe that the state should provide additional assistance to low-performing districts to support the implementation of SEED. This recommendation is supported by our finding that lower-performing districts with multiple initiatives underway encountered particular challenges in implementing SEED. Providing additional support and assistance to these districts may help them integrate SEED with other initiatives and increase the chance that the reform spurs positive change in their schools and classrooms.

8. **Continue to track SEED's implementation and effects.** We recommend that the state continue to gather data from educators at regular intervals to inform the continuous improvement of the model. In addition, we recommend tracking the model's utility in rating

the annual performance of educators and administrators. Specifically, we recommend that the relationship between SEED and student achievement be examined.