What's Happening

November 2015

A descriptive study of the pilot implementation of student learning objectives in Arizona and Utah

Reino Makkonen
Jaclyn Tejwani
Fernando Rodriguez, Jr.
WestEd

Key findings

This study examines the pilot implementation of student learning objectives (SLOs) in two states that used different approaches to achieve the same goals. SLOs are set by teachers and their principal to measure classroom-specific student achievement growth for teacher evaluations. The study found:

- In Arizona participating teachers' end-of-year SLO scores differentiated between high- and low-performing teachers and had low but statistically significant positive correlations with the results from classroom observations and student surveys of teacher performance.
- In Utah participating teachers' SLO scores after a half-year implementation varied little, with most teachers rated as meeting expectations. Teachers' survey responses indicated that the SLO process was beneficial both to students and to their own professional growth, but the teachers did not see the pilot as positively affecting their instruction or their knowledge of effective ways to assess students.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL WEST
Regional Educational Laboratory At WestEd

U.S. Department of Education
Arne Duncan, *Secretary*

Institute of Education Sciences
Ruth Neild, *Deputy Director for Policy and Research*
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Joy Lesnick, *Acting Commissioner*
Amy Johnson, *Action Editor*
OK-Choon Park, *Project Officer*

REL 2016–102

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

November 2015

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0002 by Regional Educational Laboratory (REL) West at WestEd. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Makkonen, R., Tejwani, J., & Rodriguez Jr., F. (2015). *A descriptive study of the pilot implementation of student learning objectives in Arizona and Utah* (REL 2016–102). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from: http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

Seeking a consistent way to evaluate teachers based on how much their students learn, many states and districts are incorporating student learning objectives (SLOs) into their set of evaluation tools. SLOs, which are classroom-specific targets for student achievement growth, may be based on a wide range of student assessments (Gill, Bruch, & Booker, 2013). This study describes results of the pilot implementation of this teacher-evaluation approach in Arizona and Utah.

SLOs consist of differentiated learning expectations for students, who are grouped according to their academic performance at the beginning of the school year. SLOs encompass an array of instructional activities that generally align with best teaching practices—chiefly, setting clear expectations for student learning, using formative assessment data to track student progress, and adjusting and differentiating instruction based on that progress. Some SLOs also articulate the instructional strategies to be used in advancing student learning. SLOs are set by individual teachers and their evaluator, often the principal or vice-principal. In an SLO teacher evaluation process, the teacher is evaluated and given an overall score based on how well the SLOs are achieved.

This descriptive study, a collaboration between the Arizona Department of Education and the Utah State Office of Education as part of the Regional Educational Laboratory West Educator Effectiveness Alliance, aims to help state and district leaders determine the appropriate level of guidance and oversight to provide in support of the SLO teacher-evaluation process. Findings from the study will help education officials better understand what teachers are focusing on in writing their SLOs, how well SLOs differentiate between high- and low-performing teachers (a key concern among evaluation reform advocates), and the benefits of the process (according to teachers). Although the contexts differed in the two states' pilots, both implemented the SLO teacher-evaluation process with the same aims: to improve student achievement and to fulfill the state's required student accountability component in teacher evaluations. Other states with this goal are likely to find the information in this report useful as well.

In Arizona, SLOs from 363 teachers in four volunteer pilot districts tended to target student proficiency growth on vendor-developed tests and to include few specifics about instructional strategies. Despite similar types of proficiency goals and the use of similar types of vendor-developed tests, teachers' end-of-year SLO scores from their principals varied, spanning performance levels (and distinguishing high- and low-performing teachers). The SLO scores from principals also identified high-performing teachers in low-performing schools and vice versa, picking up within-school variation that schoolwide test results cannot reveal. Teachers with higher SLO scores were also rated higher by their principals in classroom observations and by their students in surveys (the correlations between these measures were positive and statistically significant, ranging from 0.17 to 0.27).

Utah's much smaller SLO pilot started mid-year, with 82 teachers (about half of them special education teachers) volunteering to participate. The Utah pilot teachers tended to define their SLO-focused instructional strategies and use their own classroom tests or rubrics, with goals geared toward students demonstrating knowledge (through project completion) or a physical skill. The resulting SLO scores for Utah pilot teachers varied little, with 89 percent of them meeting expectations.

In both pilots, teachers' end-of-year SLO scores from their principals differed for elementary and secondary school teachers, with elementary school teachers scoring better on average in Arizona and secondary school teachers scoring better in Utah. This finding suggests that each state's SLO process might need to be adjusted to better account for variations in teacher performance at the two school levels. For example, the states might consider offering separate SLO training and guidance to elementary and high school principals.

In end-of-year surveys, participating Utah teachers generally perceived the SLO process as worthwhile and beneficial to their students and to their own professional growth. However, they did not perceive the SLO pilot as having positively affected their instruction or their knowledge of effective ways to assess students. (A low response rate precluded survey analysis for the Arizona pilot.)

Study researchers did not directly observe the SLO process (for example, teacher and evaluator goal-setting discussions or the assessment of students) in either state, relying instead on a review of pilot documentation and results.

Contents

Summary i

Why this study? 1

What the study examined 3

What the study found 6
Arizona's student learning objectives pilot 6
Utah's student learning objectives pilot 11

Implications of the study findings 16

Limitations of the study 16

Appendix A. Data and methodology A-1

Appendix B. Correlations between Arizona teacher measures, by grade span B-1

Appendix C. Response frequencies for Utah's pilot teacher survey C-1

Notes Notes-1

References Ref-1

Boxes
1 Previous research shows benefits and concerns about use of student learning objectives in teacher evaluation 2
2 Pilot student learning objective process in Arizona 3
3 Pilot student learning objective process in Utah 4

Figures
1 Arizona pilot teachers' student learning objective scores spanned performance levels 8

Tables
1 Arizona and Utah engaged different teacher samples for their pilot implementation of student learning objectives 6
2 Arizona pilot teachers set different types of goals in their student learning objectives 7
3 Arizona pilot teachers used different types of assessments in their student learning objective 8
4 Arizona pilot teachers' student learning objective scores varied according to the type of goal set 9
5 Arizona pilot teachers' student learning objective scores varied by district 10
6 Arizona pilot teachers' student learning objective scores varied by school letter grade 10
7 Arizona pilot elementary school teachers had higher student learning objective scores than pilot secondary school teachers 11
8 Correlations indicate that Arizona pilot teachers with higher observation and student survey scores tended to have higher student learning objective scores 11
9 Utah pilot teachers set different types of student learning objective goals 12

10	Most Utah pilot teachers met or exceeded expectations on their student learning objectives	13
11	A higher proportion of Utah pilot special education teachers than of pilot general education teachers met their student learning objective expectations	13
12	Utah pilot teachers' student learning objective scores varied by school grade span	14
13	Utah pilot teacher survey respondents had varied perceptions of the student learning objective process	15
A1	Coding scheme for the focus of the student learning objective	A-3
A2	Coding scheme for the goals of the student learning objective	A-5
B1	Correlations between evaluation measures differed for Arizona elementary and secondary school pilot teachers	B-1
C1	Utah pilot teacher survey respondents taught different subjects	C-1
C2	Utah pilot teacher survey respondents taught different grade spans	C-1
C3	Utah pilot teacher survey respondents had varying years of experience at their schools	C-1
C4	Utah pilot teacher survey respondents had varying years of teaching experience	C-1
C5	Utah pilot teachers' agreement with statements about the student learning objective process varied	C-2
C6	Utah pilot general and special education teachers had different views on the student learning objective process changing their instruction	C-2
C7	Utah pilot general and special education teachers had different views on the student learning objective process benefiting students	C-3
C8	Utah pilot elementary and secondary school teachers had different views on the potential accuracy of the student learning objective process	C-3
C9	Utah pilot elementary and secondary school teachers had different views on their gaining assessment knowledge through the student learning objective process	C-3
C10	Utah pilot elementary and secondary school teachers had different views on the student learning objective process informing their professional growth	C-4
C11	Utah pilot elementary and secondary school teachers had different views on the student learning objective process benefiting students	C-4
C12	Utah pilot elementary and secondary school teachers had different views on the student learning objective process being worthwhile overall	C-4
C13	Utah pilot teachers at different scoring levels had different views on student learning objective results accurately reflecting their contributions to their students' growth	C-5
C14	Utah pilot teachers who revised their targets mid-pilot had different views on the student learning objective process improving teacher conversations	C-5

Why this study?

Responding to new state laws and federal grant requirements, nearly two-thirds of U.S. states have changed their teacher evaluation policies since 2009 (Jerald, 2012). Federal grant applications for Race to the Top in 2009 and 2010 required states to design comprehensive evaluation systems with multiple measures of teacher performance (Duncan, 2010). In addition, applications for flexibility in meeting Elementary and Secondary Education Act provisions require states to describe their plans to reform teacher evaluation to strengthen the focus on the quality of instruction and student results (U.S. Department of Education, 2012).

As part of this reform, many states have been developing evaluation systems over the past three years to gauge the contributions of individual teachers to their students' learning growth. Thirty states are adopting—or planning to adopt—the approach of using student learning objectives (SLOs), but definitions of SLOs vary widely (Lacireno-Paquet, Morgan, & Mello, 2014). This study describes results of pilot implementations of this teacher evaluation approach in two states in the Regional Educational Laboratory West Region, Arizona and Utah.

This descriptive study explores variations in SLO goals, types of assessments, and end-of-year scores from principals, along with teacher perceptions of the process, during pilot implementation in two states

SLOs are goals set collaboratively by individual teachers and their evaluator, often the principal or vice-principal. Goals are expressed as differentiated learning expectations for groups of students, clustered by the teacher according to students' beginning-of-year performance. A teacher's annual SLOs could include, for example, moving 50 percent of students in the lowest performing category into the average category and 30 percent of average students into the highest performing category, as judged by a year-end assessment. SLOs also address how student learning will be measured, for example, through proficiency scores on statewide standardized tests or through a performance rubric. Students' progress in meeting the expectations set by the SLOs is tracked over a defined period—say, a semester or a school year (Marion, dePascale, Domaleski, Gong, & Diaz-Biello, 2012). Teachers are evaluated based on the degree to which their students meet the objectives and are given an overall score by the evaluator.

Previous research, driven by the need to inform states on how to link student achievement and teacher evaluation, has explored the pros and cons of using SLOs to evaluate teachers (box 1). This descriptive study extends that research by exploring variations in SLO goals, types of assessments, and end-of-year scores from principals, along with teacher perceptions of the process, during pilot implementation in two states.

Findings from this study will help education officials better understand what teachers are focusing on in writing their SLOs, how well SLOs can provide differentiated results about teacher effectiveness, and the benefits of the process (according to teachers). Arizona and Utah officials can use the findings to inform decisions on whether policy modifications are needed, as well as to determine the appropriate level of state oversight and support to provide as more districts adopt or adapt the SLO approach. Findings will also be of interest to the many other districts and states considering incorporating SLOs into teacher evaluation.

Box 1. Previous research shows benefits and concerns about use of student learning objectives in teacher evaluation

The use of student learning objectives (SLOs) can provide benefits beyond teacher evaluation. The SLO process orients teachers toward a number of instructional activities that generally align with effective teaching practices (Lachlan-Haché, Cushing, & Bivona, 2012), chiefly, setting clear performance expectations for students, using formative assessment data to track student progress, and adjusting and differentiating instruction based on that progress. In Denver, Colorado, where SLOs were a key part of the district's performance pay system in 1999, more than 60 percent of participating teachers expressed the belief that SLOs improved their instructional practices (Proctor, Walters, Reichardt, Goldhaber, & Walch, 2011). Instructional benefits were also reported in other recent implementations of SLOs in performance pay systems in Austin, Texas (Schmitt & Ibanez, 2011), and in Charlotte-Mecklenburg, North Carolina (Community Training and Assistance Center, 2013).

For teacher evaluation purposes, SLOs are flexible enough to be implemented across content areas and grade levels, using existing or newly developed methods of assessment. Teachers may find the approach more credible than some other approaches because of their direct involvement in setting goals and assessing students (Goe & Holdheide, 2011). Yet the SLO process's fundamental reliance on teachers' and principals' professional judgments also makes it difficult to ensure that SLOs are sufficiently valid and reliable for teacher evaluations (Gill et al., 2013). Also, because of the potentially high degree of variability in the type and quality of assessments used to measure SLO achievement, results may not be comparable across classrooms or schools (Goe & Holdheide, 2011; Gill et al., 2013).

Nonexperimental evaluations of SLO initiatives have shown promising correlations between teacher SLO scores and other student achievement indicators. For example, in both Denver and Charlotte-Mecklenburg, teachers who developed higher quality learning objectives for their students (rated according to district-created rubrics) also had better student achievement gains on standardized tests than teachers who developed lower quality objectives (after prior student achievement was controlled for; Community Training and Assistance Center, 2004, 2013). In Austin, students whose teachers met at least one of their SLOs were more likely to score higher on standardized assessments, on average, than students whose teachers did not meet any of their SLOs (Schmitt & Ibanez, 2011).

A common concern expressed about the use of SLOs in teacher evaluation involves a potential for lack of rigor, or ambitiousness, in the goals teachers might set. Indeed, evidence suggests that teachers tend to achieve their SLOs; a recent literature review about the use of SLOs published by Regional Educational Laboratory Mid-Atlantic found that "more than half the teachers met their targets in all the (examined) locations" (Gill et al., 2013, p. 11). In Denver, 89–93 percent of teachers met one or more SLOs over the four years of the study, with less rigorous objectives more likely than more rigorous ones to be achieved (Proctor et al., 2011). A similar theme was reflected in a recent study of initial SLO implementation in Tennessee, which found that choices of measures or assessments in the SLO process were often based on the teachers' and principals' beliefs about which assessments would produce the highest scores (Tennessee Department of Education, 2012).

The use of SLOs to measure teacher effectiveness presumes that scores will vary and that not everyone will get the same score—that SLOs will distinguish low-performing teachers who would benefit from training or remediation. Differentiation among teachers is shown by variability among their SLO scores. This is a key concern among those advocating for teacher evaluation reform, who cite evidence showing that many evaluation systems tend to rate the vast majority of teachers as satisfactory (Weisberg, Sexton, Mulhern, & Keeling, 2009).

What the study examined

During the 2013/14 school year Arizona and Utah pilot tested new SLO-based teacher evaluation processes with a sample of participating teachers (see table 1 later in the report for the sample characteristics). The Arizona Department of Education conducted a full-year SLO implementation with 363 participating teachers in four volunteer pilot districts as part of the pilot implementation of its multiple-measure teacher evaluation model. In January 2014 the Utah State Office of Education also began a pilot of its SLO-based teacher evaluation process for 82 volunteering teachers across 10 districts. Although the pilot implementations differed, both states used SLOs in teacher evaluation with the same aims—to improve student achievement and to fulfill the state's required student accountability component in teacher evaluations (Arizona Department of Education, 2013a; Utah State Office of Education, 2014a). The pilot SLO process is described for Arizona in box 2 and for Utah in box 3.

Box 2. Pilot student learning objective process in Arizona

Arizona's pilot student learning objective (SLO) process began in August 2013 and included multiple steps for teachers and evaluators in 2013/14, with each step documented on a standard Arizona Department of Education reporting form. First, teachers were asked to use data from the previous year (including, for example, students' prior grades, tests from other subjects, portfolios, and performance rubrics) to determine how prepared their incoming students were to start the class. Then participating pilot teachers used data analysis and professional judgment to rank students and group them into high, adequate, and low levels of preparedness (Arizona Department of Education, 2013b).

Next, teachers identified (and evaluators approved) the state, district, or classroom assessments to be used for the SLO process, and students were administered an initial baseline assessment. In approving the assessments, evaluators used a state rubric to ensure that each SLO assessment was aligned to state content standards, appropriately rigorous for the grade level, and valid and reliable (or at least "feasible for the SLO process at this time"; Arizona Department of Education, 2013b, pp. 21–23).

Pilot teachers then established (and evaluators approved) a classroom SLO and a targeted SLO. The classroom SLO was a goal for the full class. To set the classroom SLO, teachers collaborated with other teachers and their evaluator to set a proficiency target (a cutscore) reflecting where they expected the students to be at the end of the year (Arizona Department of Education, 2013a, 2013b, p. 26). The Arizona Department of Education suggested a 75 percent proficiency target. The state guidance recommended that "all teachers within the LEA [local education agency] instructing in the same course set the same proficiency level" for their 2013/14 classroom SLO (Arizona Department of Education, 2013b, p. 26).

The targeted SLO was a "growth goal" focused only on students beginning at the low level of preparedness, that is, "at risk students who will need targeted support or intensive instruction to master the course content standards" (Arizona Department of Education 2013b, p. 8), a group that should include "at least one-third of the class" (p. 13). For the targeted SLO, separate end-of-year target cutscores were set for each student in the low preparedness group. The state's general expectation for growth in 2013/14 was that each student in the low preparedness subgroup would gain at least 65 percent of the possible increase over his or her

(continued)

Box 2. Pilot student learning objective process in Arizona *(continued)*

baseline score, with the individual target calculated from a state-provided formula as follows: Baseline score + [(Highest possible score – Baseline score) × .65] (p. 29).

At the end of the year, teachers were to administer and score the selected student assessments and review the results with their evaluator, who, in turn, would assign a score, ranging from 1 (low) to 4 (high), for each teacher's performance on the classroom and targeted SLOs. For the classroom SLO, teachers received 4 points if 90–100 percent of the class met the cutscore, 3 points for 80–89 percent, 2 points for 60–79 percent, and 1 point for less than 60 percent. For the targeted SLO, scores were first calculated by the teacher at the student level—4 points if the student surpassed his or her target by more than 5 percent of the cutscore, 3 points if within 5 percent of the cutscore, 2 points if 5–10 percent below the cutscore, and 1 point if more than 10 percent below the cutscore. The mean of the subgroup's points represented the teacher's final targeted SLO score (Arizona Department of Education, 2013b pp. 42–43).

This pilot was conducted to explore how to incorporate SLOs into a wider program of teacher evaluation. In summer 2012 the Arizona Department of Education developed a state model teacher evaluation process based on the Arizona Framework for Measuring Educator Effectiveness, approved by the State Board of Education the previous year. The framework consists of three general evaluation components—teaching performance, school-level data, and classroom-level academic progress data. The state model specifies that these components include observations of instructional practice in classrooms using the Danielson Framework for Teaching; online surveys of students, parents, and peer teachers; and measures of student academic progress based on teaching assignment (defined by state-created "rating tables"), with SLOs incorporated in the latter component in the 2013/14 school year (Arizona Department of Education, 2013a).

Box 3. Pilot student learning objective process in Utah

Utah state officials engaged in a lengthy student learning objectives (SLO) development period before teachers embarked on their pilot in January 2014. Starting in fall 2012, workgroups convened by the Utah State Office of Education created a statewide SLO template and rubric for assessing SLO quality and developed more than 100 sample SLOs for various grades and subjects (each including a classroom learning expectation and group performance goals), which participating educators were encouraged to use in developing their pilot SLOs (Utah State Office of Education, 2014b).

As in Arizona, teachers established a classroom-level student learning expectation for the end of the year based on state content standards. Teachers gathered and documented baseline performance data on their students, including, for example, students' prior standardized test results, grades, attendance rates, tutoring or remediation services, and "any other data that links classroom practices to student achievement" (Utah State Office of Education 2014b, p. 5). Teachers used the baseline data to classify students into low, average, and high performance groups and set differentiated targets for each group on evaluator-approved assessments, which were created by teachers, the district, or curricula vendors. A quality assessment, as defined by preliminary Utah State Office of Education guidance, is one that "measures the

(continued)

Box 3. Pilot student learning objective process in Utah *(continued)*

intended learning target at the appropriate depth of knowledge; engages students in meaningful subject matter; allows students and teachers to learn from the assessment; is accessible to a wide variety of students; is as fair as possible; and (yields) scores (that) provide a reliable estimate of student learning" (p. 4).

At the end of the pilot year, teachers assessed their students and recorded their growth compared with their group's target (Utah State Office of Education, 2014b, p. 5). Each teacher's completed pilot SLO form was expected to include the grade level or content area taught, the baseline data used to categorize performance groups, a description of the assessments used (accompanied by clear criteria or scoring rubrics), the expected and the observed number and percentage of students reaching each group's target, and the evaluator's final overall rating for the teacher based on the teacher's achievement of the goals embodied in the SLO (does not meet expectations, partially meets expectations, meets expectations, or exceeds expectations).

While Utah's SLO template included both a learning goal (essentially a classroom SLO) and targets for different student subgroups (targeted SLO), Utah's general education pilot teachers were awarded a single final overall score for their SLO. Although an overview of the state's initial guidance for assigning final SLO ratings to teachers is provided in the January 2014 *Utah SLO Toolkit* (Utah State Office of Education 2014b, p. 6), Utah State Office of Education officials did not specify how participating pilot administrators were to calculate a teacher's final overall SLO rating or ratings. The state planned to offer more explicit scoring guidance in 2014/15.

Like Arizona, Utah was examining SLOs as part of a state mandate for local education agencies to develop new teacher evaluation systems. In November 2011 the Utah State Board of Education approved the Public Educator Evaluation Requirements, which specifies that a local education agency's evaluation system "shall include valid and reliable measurement tools including, at a minimum: (a) observations of instructional quality; (b) evidence of student growth; (c) parent and student input; and (d) other indicators as determined by the [local education agency]" (Administrative Rule R277–531–3(3)).

Using data from the two pilots, this study explored three research questions:

1. How did teachers' SLO content (goals and assessments) and end-of-year scores vary within each state?

2. What associations were evident between teachers' SLO scores and their scores based on classroom observation and student surveys?

3. What were teachers' perceptions of the SLO process?

The study was able to address only research questions 1 and 2 for Arizona because of a low survey response rate and only research questions 1 and 3 for Utah because no observation or student survey data were available for pilot teachers.

What the study found

The two states implemented different SLO processes (see boxes 2 and 3) on different pilot timelines—an August 2013 start in Arizona and a January 2014 start in Utah—with different samples of teachers, who volunteered for the pilot in different ways. In Arizona four districts (identified in this report as A–D) volunteered their teachers to be part of the state's two-year, multiple-measure pilot in summer 2012. In Utah teachers volunteered individually, and many were special education teachers. Because certain key data were unavailable, as noted previously, the findings for each state's SLO pilot are presented separately. Summary information on the pilot teacher samples in the two states, compiled from SLO documents, is displayed in table 1.

Arizona's student learning objectives pilot

Arizona and Utah implemented different SLO processes on different pilot timelines with different samples of teachers, who volunteered for the pilot in different ways

SLOs were consistently assessment-focused, most often targeting proficiency growth on vendor-developed tests. The content of Arizona teachers' SLOs focused mainly on students' test performance and tended not to describe the content that students were to learn or the instructional strategies the teacher was to employ.[1] This approach aligned with the state guidance provided to pilot teachers, which described the classroom SLO as a "proficiency achievement objective" or "mastery score" and recommended that the targeted SLO be written as "a growth measure focusing on" the teacher's least prepared students (Arizona Department of Education, 2013b, pp. 6, 26, 8; see box 2 for a discussion of the two types of SLOs). The goals set by Arizona's pilot teachers on their classroom SLO tended to focus on having students meet either a particular benchmark (51.5 percent of teachers) or a specific proficiency score (35.5 percent), whereas the targeted SLO for

Table 1. Arizona and Utah engaged different teacher samples for their pilot implementation of student learning objectives

Sample	Arizona	Utah
Total pilot teachers	363	82
Teachers providing student learning objective documents	323	67
Grade level[a] taught (percent of pilot teachers providing documents)		
Elementary (PK–5)	48.6	32.8
Middle (6–8)	23.2	27.9
High (9–12)	25.7	31.1
Multiple grade spans	2.5	8.2
Subject area taught[a] (percent of pilot teachers providing documents)		
English language arts	22.0	0.0
Fine arts	6.2	14.9
Math	44.6	3.0
Physical education	5.9	0.0
Science	4.6	9.0
Social studies	5.0	10.4
Special education	4.3	55.2
Other	7.4	7.5

a. As reported on the student learning objectives form submitted for this study.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education and the Utah State Office of Education.

low-performing students tended to specify growth goals in terms of an increase in either percentage of correct answers on the year-end assessment (62.0 percent) or proficiency level (23.8 percent; table 2). (See table A2 in appendix A for further information about these goal categories). The most common types of SLO goals, such as growth and proficiency, did not vary based on the subject area taught by the teacher. Approximately 5 percent of the Arizona teacher sample (primarily special education and specialist teachers) included a second targeted SLO; however, the remainder of this discussion of Arizona's SLO pilot focuses only on the results of teachers' first targeted SLO.

To measure students' progress against the expectations described in the classroom and targeted SLOs, Arizona teachers in pilot districts B, C, and D tended to rely on vendor-developed content tests, while teachers in pilot district A tended to rely more on teacher-developed classroom tests (table 3). State or national standardized tests and district- or school-wide assessments were used less frequently. There was no relationship between teachers' SLO scores and the type of assessment used in the SLO.

The goals set by Arizona's pilot teachers on their classroom SLOs tended to focus on having students meet either a particular benchmark or a specific proficiency score, whereas the targeted SLOs for low-performing students tended to specify growth goals in terms of an increase in either percentage of correct answers on the year-end assessment or proficiency level

SLO scores appeared to distinguish between high- and low-performing teachers. Both classroom and targeted SLO scores awarded to Arizona's pilot teachers spanned performance levels (figure 1). However, compared with teachers' classroom SLO scores, teachers' targeted SLO scores—which were based on the progress of their least prepared subgroup of students—were skewed toward higher scores. Arizona teachers' classroom SLO scores were more balanced across the four scoring levels, with 4 being the most common score.[2]

Different types of SLO goals were associated with different teacher scores. Arizona pilot teachers' SLO scores varied by goal type. For example, a higher proportion of pilot teachers whose classroom SLO asked students to demonstrate a physical skill (such as performing a piece of music or running a mile in a specified amount of time) were awarded a 4 than teachers setting other types of SLO goals (table 4). Meanwhile, the classroom SLO scores for

Table 2. Arizona pilot teachers set different types of goals in their student learning objectives

Type of goal	Percent of pilot teachers (n = 323)
Classroom student learning objective	
Demonstrates knowledge through project completion	1.3
Demonstrates a physical skill	7.6
Meets growth target	4.0
Meets test range or benchmark	51.5
Meets proficiency level	35.5
Targeted student learning objective	
Demonstrates completion, no growth	1.3
Demonstrates growth	8.3
Meets test score, no growth	3.5
Meets percentage growth	62.0
Meets proficiency level growth	23.8
Project completion, no growth	1.1

Note: Includes only teachers' primary (first) classroom or targeted student learning objective. Goal categories are defined in table A2 in appendix A. Percentages may not sum to 100 because of rounding.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

Table 3. Arizona pilot teachers used different types of assessments in their student learning objectives

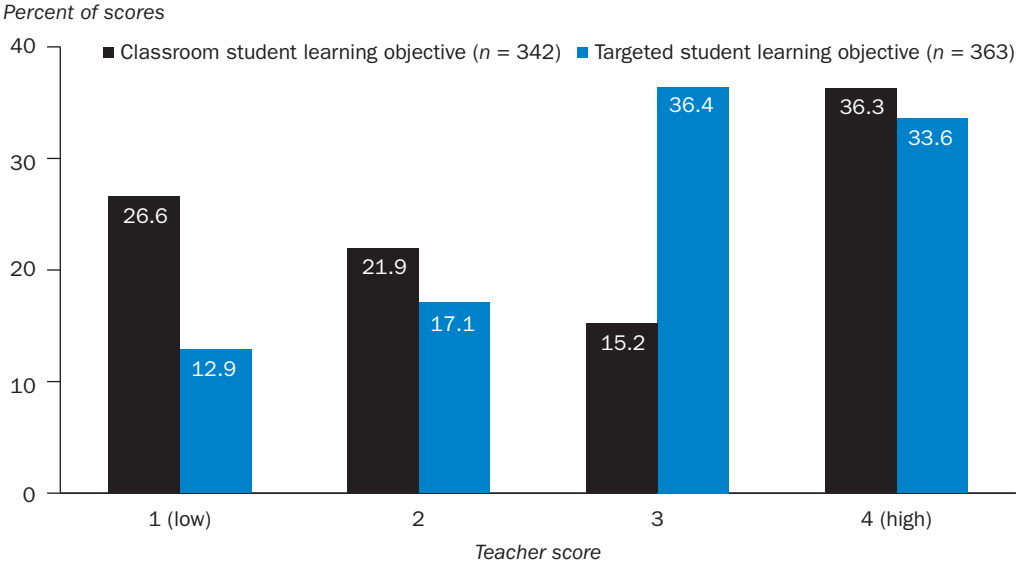
Percent of teachers

Type of assessment	Arizona (*n* = 323)	District A	District B	District C	District D
Classroom student learning objective					
Vendor-developed content	79.8	31.6	85.6	83.3	72.7
Teacher-developed classroom	10.6	52.6	7.2	3.3	12.1
National or state standardized	6.1	10.5	4.4	13.3	6.0
District- or school-wide	3.4	5.3	2.8	0.0	9.1
Targeted student learning objective					
Vendor-developed content	81.1	36.8	83.7	89.5	80.0
Teacher-developed classroom	11.4	52.6	9.8	2.6	8.6
National or state standardized	4.6	5.3	3.7	7.9	5.7
District- or school-wide	2.9	5.3	2.8	0.0	5.7

Note: Includes only teachers' primary (first) classroom or targeted student learning objective. Assessment categories are defined in appendix A. Percentages may not sum to 100 because of rounding.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

Figure 1. Arizona pilot teachers' student learning objective scores spanned performance levels



Note: This figure displays the variation in student learning objective (SLO) scores across Arizona's participating teachers, by SLO type; for example, 27 percent of teachers received a 1 on their classroom SLO, and 13 percent of teachers received a 1 on their targeted SLO. The figure includes only teachers' primary (first) classroom or targeted student learning objective score. Teachers' targeted student learning objective scores were rounded to the nearest integer.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

teachers who wanted their students to meet a particular test range or benchmark tended to span all four scoring levels. For targeted SLOs—with the average of the group's student-level results rounded to the nearest integer for comparison here—teachers seeking demonstrated growth from students in a physical skill or behavior tended to score 3 or 4 on their SLO,

Table 4. Arizona pilot teachers' student learning objective scores varied according to the type of goal set

Percent of scores

Type of goal	Teacher score			
	1	2	3	4
Classroom student learning objective				
Demonstrates a physical skill	14.3	9.5	9.5	66.7
Meets growth target	25.0	25.0	0.0	50.0
Meets test range or benchmark	21.3	21.3	18.4	39.0
Meets proficiency level	47.7	21.6	7.9	22.7
Targeted student learning objective				
Demonstrates growth	0.0	6.9	55.2	37.9
Meets test score, no growth	18.2	18.2	18.2	45.5
Meets percentage growth	18.0	17.0	32.5	32.5
Meets proficiency level growth	1.9	24.5	45.3	28.3

Note: Data cover 323 teachers. This table offers a further breakdown of the results displayed in figure 1. Only goal types with more than 10 associated scores are displayed, and teachers' targeted student learning objective scores were rounded to the nearest integer. The relationship between goals and scores was statistically significant for both classroom student learning objectives (chi square = 37.06, $p < 0.01$) and targeted student learning objectives (chi square = 33.48, $p < .01$). Goal categories are defined in table A2 in appendix A. Percentages may not sum to 100 because of rounding.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

as did 74 percent of teachers specifically referencing growth in proficiency on a particular assessment or rubric. As noted, the majority of pilot teachers sought a specific percentage of growth for their targeted SLO. Results varied among the pilot teachers who set this type of goal: 65 percent received scores of 3 or 4, and 35 percent received scores of 1 or 2.

The distribution of SLO scores awarded to Arizona pilot teachers varied by district but were not necessarily associated with the district's letter grade

SLO scores also varied across districts and by school performance level. Arizona assigns A–F letter grades for districts and schools based on student test scores on statewide exams, with essentially equal value placed on current-year achievement and longitudinal test growth (specifically, the test growth of all students as well as a school's lowest achieving students). In this study the distribution of SLO scores awarded to Arizona pilot teachers varied by district but were not necessarily associated with the district's letter grade for the 2013/14 school year (table 5). For example, while pilot districts A and C had the same state test–based letter grade of C, 62 percent of teachers in pilot district A received the highest possible score (4) on their classroom SLO, while only 21 percent of teachers in district C did. At the same time, teachers in pilot district D, which received the lowest 2013/14 letter grade among the pilot districts, also tended to receive lower SLO scores than teachers in other pilot districts.

SLO scores corresponded more closely to the state-assigned school letter grade. In general, pilot teachers in schools with higher grades tended to receive higher end-of-year SLO scores, while pilot teachers in schools with lower grades tended to receive lower scores. For example, while 79 percent of pilot teachers in schools graded A earned a 3 or 4 on their classroom SLO in 2013/14, 20 percent of teachers in schools graded D earned a 3 or 4 (table 6). There was a low positive correlation between teachers' SLO scores and their school's grade in 2014—approximately 0.3 ($p < .01$) for both classroom and targeted SLOs.[3] At the same time, teacher performance varied within schools. The SLO scores identified this within-school variation, recognizing high-performing teachers in low-performing schools, and vice versa.

Table 5. Arizona pilot teachers' student learning objective scores varied by district

Percent of scores

District	District letter grade	Teacher score			
		1	2	3	4
Classroom student learning objective (*n* = 342)					
District A	C	2.2	13.3	22.2	62.2
District B	B	24.2	22.9	16.9	36.0
District C	C	55.2	20.7	3.4	20.7
District D	D	53.1	28.1	3.1	15.6
Targeted student learning objective (*n* = 363)					
District A	C	8.2	8.2	34.7	48.9
District B	B	8.9	19.6	39.2	32.2
District C	C	2.9	14.3	37.1	45.7
District D	D	58.8	14.7	17.6	8.8

Note: Teachers' targeted student learning objective scores were rounded to the nearest integer. The differences in scores by pilot district were statistically significant for both classroom student learning objectives (chi square = 41.11, $p < 0.01$) and targeted student learning objectives (chi square = 77.81, $p < .01$). Percentages may not sum to 100 because of rounding.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

Table 6. Arizona pilot teachers' student learning objective scores varied by school letter grade

Percent of scores

School letter grade	Teacher score			
	1	2	3	4
Classroom student learning objective (*n* = 342)				
A	5.3	15.8	26.3	52.6
B	9.1	23.6	20.9	46.4
C	35.8	21.2	11.9	31.1
D	55.0	25.0	5.0	15.0
Targeted student learning objective (*n* = 363)				
A	0.0	15.8	31.6	52.6
B	0.8	13.2	41.2	44.7
C	16.3	18.8	36.1	28.8
D	54.5	22.7	18.2	4.5

Note: Teachers' targeted student learning objective scores were rounded to the nearest integer. The relationship of teachers' student learning objective scores with school grade was statistically significant for both classroom student learning objectives (chi square = 43.99, $p < 0.01$) and targeted student learning objectives (chi square = 64.39, $p < 0.01$). Percentages may not sum to 100 because of rounding.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

Elementary school teachers scored higher on SLOs than secondary school teachers. In Arizona's SLO pilot, participating elementary school teachers received higher scores on their classroom SLO and targeted SLO than did pilot secondary school teachers, on average. A higher proportion of secondary school teachers than elementary school teachers received a score of 1, and a lower proportion received a 4 (table 7).

SLO scores aligned with other teacher performance indicators. In addition to being rated on SLO achievement in 2013/14, Arizona's pilot teachers were rated by their administrators in classroom observations using the Danielson Framework for Teaching (Danielson Group,

Table 7. Arizona pilot elementary school teachers had higher student learning objective scores than pilot secondary school teachers

Percent of scores

	Classroom student learning objective score (n = 342)				Targeted student learning objective score (n = 363)			
Grade span	1	2	3	4	1	2	3	4
Elementary	19.6	19.6	16.8	44.0	6.2	12.4	37.6	43.8
Secondary	34.8	24.7	13.3	27.2	20.7	22.5	34.9	21.9

Note: Teachers' targeted student learning objective scores were rounded to the nearest integer. The relationship with grade span was statistically significant for both classroom student learning objectives (chi square = 16.34, $p < 0.01$) and targeted student learning objectives (chi square = 33.22, $p < .01$). Elementary school teachers scored 0.44 pooled standard deviation higher than secondary school teachers on their classroom student learning objectives ($p < .01$) and 0.62 pooled standard deviation higher on their targeted student learning objectives ($p < .01$).

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

Table 8. Correlations indicate that Arizona pilot teachers with higher observation and student survey scores tended to have higher student learning objective scores

Result	Classroom student learning objective score	Targeted student learning objective score	Average observation score[a]	Total observation points[b]	Average student survey score[c]
Targeted student learning objective score	0.64*				
Average observation score[a]	0.20*	0.09			
Total observation points[b]	0.25*	0.17*	0.96*		
Average student survey score[c]	0.24*	0.19*	0.27*	0.26*	
Student survey points[b]	0.22*	0.17*	0.23*	0.23*	0.94*

* Significant at $p < .05$.

a. Mean score across the teacher's 22 Danielson Framework for Teaching component ratings.

b. The state's evaluation model assigns a point value to each component measure's results to establish a uniform summative point scale.

c. Students rated more than 30 teacher behaviors on a four-point scale (1 = never, 2 = some of the time, 3 = most of the time, 4 = always). An average rating was calculated across all behaviors for each student, and then the teacher was assigned an aggregate score that was the proportion of the average ratings that were 3 or above.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

2011) and by their students on anonymous online surveys that assessed the extent to which the teacher's instruction engaged and challenged them.[4] Results for the three measures of teacher evaluation had positive (but low) rank correlations (table 8), suggesting that although the measures may capture separate dimensions of effective teaching, Arizona teachers who were observed to use more effective teaching methods tended to have higher SLO scores, as did teachers who were viewed by their students as being more engaging and challenging. The correlations between the different pilot measures were slightly stronger (and more often statistically significant) among secondary school teachers than among elementary school teachers (see table B1 in appendix B).

Utah's student learning objectives pilot

Utah's SLO pilot differed from Arizona's pilot in several ways. It began later than Arizona's, in January 2014, with only 82 participating teachers, over half of whom were special

education teachers. In addition, the participating Utah teachers were not assessed through classroom observations or student surveys during the SLO pilot (precluding any examination of research question 2 in Utah). However, more than two-thirds of Utah's pilot teachers completed surveys about their experience (allowing for an exploration of research question 3).

Most Utah pilot teachers' classroom SLOs involved student demonstrations, while most of the targeted SLOs sought growth in proficiency levels. Utah pilot teachers established both a learning goal for the class, a classroom SLO (what students were expected to be able to do at the end of the course or grade based on state standards), and targets for the teacher's student subgroups, targeted SLOs (see box 3 for a discussion of the two types of SLOs.) Some 73 percent of Utah teachers' classroom SLOs focused on demonstrating knowledge through project completion, and 23 percent focused on demonstrating a physical skill (table 9). Meanwhile, 62 percent of Utah's targeted SLOs sought growth in students' proficiency level, while 18 percent targeted a particular cutscore on the assessment (for example, a score of 75 percent or higher, with growth not specified), and 16 percent sought demonstrated growth in a physical skill or behavior.[5] (Further information about these goal categories is provided in table A2 in appendix A.)

Some 73 percent of Utah teachers' classroom SLOs focused on demonstrating knowledge through project completion, and 23 percent focused on demonstrating a physical skill

Utah's pilot teachers tended to use teacher-developed classroom tests in their SLOs, but general and special education teachers documented the SLO process in different ways. All 30 of the participating general education teachers with available documents included in their classroom SLO both a learning goal and instructional strategies for attaining it, and 90 percent (27) of them included expected target outcomes for each student subgroup. These practices aligned with the guidance in the state's template SLO form. Some 96 percent of these teachers indicated that they used a teacher-developed test to measure student progress for their SLOs.[6] The review of SLO documents suggested that special education teachers and general education teachers recorded their SLOs differently. Instead of documenting their SLOs using the state's template, participating special education

Table 9. Utah pilot teachers set different types of student learning objective goals

Type of goal	Percent of pilot teachers (n = 67)
Classroom student learning objective	
Demonstrates knowledge through project completion	73.3
Demonstrates a physical skill	23.3
Meets growth target	0.0
Meets test range or benchmark	0.0
Meets proficiency level	3.3
Targeted student learning objective	
Demonstrates completion, no growth	0.0
Demonstrates growth	15.5
Meets test score, no growth	17.5
Meets percentage growth	4.1
Meets proficiency level growth	61.9
Completes project, no growth	1.0

Note: Goal categories are defined in table A2 in appendix A. Percentages may not sum to 100 because of rounding.

Source: Authors' analysis of 2014 data provided by the Utah State Office of Education.

teachers relied on spreadsheets to track and quantify their students' progress on specific targeted skill-based measures that assessed, for example, fine motor skills, text comprehension, or number sense. These SLOs did not identify a learning goal, nor did they describe the teacher's instructional strategies. Instead, they were more likely to specify, for example, that over the course of the year eight students would combine to make a quantified amount of progress in their speech-language accuracy or complexity.

Utah pilot teachers tended to achieve their SLOs. Although Utah's SLO template included both a learning goal (classroom SLO) and targets for different student subgroups (targeted SLO), Utah's pilot teachers were awarded final overall scores for their SLOs.[7] Of the 82 participating Utah pilot teachers, 63 had a final SLO score included on their forms. Of these 63 teachers with a final overall SLO score, 30 special education teachers received a second score to reflect a second (separate) targeted SLO. In general, Utah's pilot SLO results skewed toward higher scores: 89 percent of Utah teachers met or exceeded expectations on their first SLO, and 87 percent of special education teachers with two SLOs met or exceeded expectations on their second SLO (table 10).[8]

In general, Utah's pilot SLO results skewed toward higher scores: 89 percent of Utah teachers met or exceeded expectations on their SLO

The distributions of first SLO scores for Utah's general and special education teachers are displayed separately in table 11, because special education teachers used a different (yes or no) scoring scheme in 2014. Throughout the remainder of this report, Utah pilot special educators' yes results are coded as 3 (meets expectations), while no results are coded as 1 (does not meet expectations). Very little variation was evident in special educators' SLO scores, with 97 percent meeting expectations.

Table 10. Most Utah pilot teachers met or exceeded expectations on their student learning objectives

Percent of teachers

Student learning objective	Does not meet or partially meets expectations	Meets or exceeds expectations
Student learning objective 1 (n = 63)[a]	11.1	88.9
Student learning objective 2 (special education teachers; n = 30)	13.3	86.7

Note: For this table, Utah's scoring scale was collapsed to two levels (due to the small size of the pilot sample). For special education teachers, yes scores were recoded as meets expectations, and no scores were recoded as does not meet expectations.

a. Includes general and special education teachers.

Source: Authors' analysis of 2014 data provided by the Utah State Office of Education.

Table 11. A higher proportion of Utah pilot special education teachers than of pilot general education teachers met their student learning objective expectations

Percent of teachers

Type of teacher	Does not meet expectations	Partially meets expectations	Meets expectations	Exceeds expectations
General education (n = 30)	0.0	20.0	46.7	33.3
	No		Yes	
Special education (n = 33)	3.0		97.0	

Source: Authors' analysis of 2014 data provided by the Utah State Office of Education.

Table 12. Utah pilot teachers' student learning objective scores varied by school grade span

Percent of scores

Grade span	Does not meet expectations	Partially meets expectations	Meets expectations	Exceeds expectations
Elementary	0.0	11.8	88.2	0.0
Secondary	3.0	12.1	60.6	24.2

Note: For special education teachers, yes scores were recoded as meets expectations, and no scores were recoded as does not meet expectations. Percentages may not sum to 100 because of rounding. Fisher's exact test: $p = .073$.

Source: Authors' analysis of 2014 data provided by the Utah State Office of Education.

Some 52 percent of participating teachers who responded to a survey about their experience with the student learning objective (SLO) pilot in Utah agreed that the SLO process was worthwhile overall, and 66 percent disagreed that they would prefer to be held accountable for schoolwide average test scores rather than SLOs

A higher proportion of secondary school teachers than of elementary school teachers exceeded expectations on their SLOs. In Utah the relationship between pilot teachers' SLO scores and the grade span they taught was the reverse of the one identified in Arizona. A higher proportion of Utah's participating secondary school teachers than elementary school teachers received a score of exceeds expectations (table 12), and although participating secondary school teachers scored higher on their SLO, on average, than elementary school teachers, the difference was not statistically significant ($p = .17$).[9]

Utah's SLO pilot teachers were not rated on any performance measure other than the SLOs (for example, observations or student surveys) in 2014, and school grades were unavailable during the study period because of a change in Utah's testing and accountability policy (Wright, 2014). There was no significant correlation between Utah pilot teachers' SLO scores in 2014 and their school's grade the year before (2012/13).

Participating teachers perceived the pilot SLO process as worthwhile for their students and their professional growth but did not see it as positively affecting their instruction or their knowledge of effective ways to assess students. At the end of the 2013/14 school year, Utah state education department officials invited the 82 teachers who had volunteered for the SLO pilot to complete an online survey about their experience. Some 56 teachers (68 percent) submitted responses. Two questions yielded majority opinions: 52 percent of respondents agreed that the SLO process was worthwhile overall, and 66 percent disagreed that they would prefer to be held accountable for schoolwide average test scores rather than SLOs. Other survey results were also indicative of teachers' positive perceptions of the SLO process. Higher proportions of respondents agreed than disagreed that the SLO process helped inform their professional growth, that the assessments used in their SLO appropriately measured their students' growth, that their final SLO score accurately reflected their contributions to their students' growth, and that their participation in the SLO process benefited their students (table 13).

Other survey results suggested that participating teachers did not view the SLO pilot as positively influencing teachers' knowledge, instruction, or professional interactions. Higher proportions of respondents disagreed than agreed that they had gained new knowledge about effective ways to assess students through the SLO process, that the SLO process had improved the quality of their conversations with their administrator or fellow teachers, and that they had changed their classroom instruction as a result of the implementation of SLOs (see table 13).

Table 13. Utah pilot teacher survey respondents had varied perceptions of the student learning objective process

Percent of survey respondents

Survey item	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Through the SLO process this year I gained new knowledge about effective ways to assess students.	30.4	44.6	25.0
I am confident that the assessment(s) used as part of my SLO(s) appropriately measured my students' growth.	46.4	32.1	21.4
The SLO process this year improved the quality of my conversations with my administrator about instruction.	21.4	33.9	44.6
The SLO process this year improved the quality of my conversations with my administrator about assessment.	23.2	33.9	42.9
The SLO process this year improved the quality of my conversations with my fellow teachers.	33.9	35.7	30.4
The SLO process helped inform my professional growth.	44.6	28.6	26.8
I have changed the instruction in my classroom as a result of the implementation of SLOs.	33.9	39.3	26.8
The final SLO score I received accurately reflected my contributions to my students' growth this year.	42.9	25.0	32.1
My participation in the SLO process has benefited my students.	44.6	19.6	35.7
Overall, the SLO process was worthwhile.	51.8	21.4	26.8
Rather than SLOs, I would prefer to be held accountable for schoolwide math and/or reading test scores.	7.1	66.1	26.8

SLO is student learning objective.

Note: Percentages may not sum to 100 due to rounding; $n = 56$. Full results from Utah's teacher survey are displayed in table C5 in appendix C.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Special education and elementary school teachers had more unfavorable perceptions of the SLO process than did general education and secondary school teachers. There were several significant differences in respondents' perceptions of the SLO process by teacher subgroup[10] (see tables C6–C14 in appendix C for details). Significantly higher proportions of special education teachers than general education teachers disagreed that they had changed their classroom instruction as a result of the implementation of SLOs (59 percent versus 19 percent, $p < .01$) and that their participation in the SLO process had benefited their students (24 versus 15 percent, $p < .05$). As noted, participating special education teachers did not include learning goals or instructional strategies in their SLOs.

In addition, participating elementary school teachers—who received a lower proportion of exceeds expectations scores on their SLOs than did secondary school teachers—more often disagreed that the SLO process can provide an accurate assessment of their teaching effectiveness (45 percent versus 28 percent of responding secondary school teachers, $p < .05$).[11] A significantly higher proportion of elementary school teachers than secondary school teachers also disagreed that the SLO pilot had helped inform their professional growth (40 percent versus 22 percent, $p < .01$), benefited their students (30 percent versus 14 percent, $p < .05$), and was worthwhile overall (30 percent versus 17 percent, $p < .01$).

Implications of the study findings

Although the two pilots differed in size and duration, teachers' SLO goals, assessments, and results varied across districts, schools, subjects, and grade spans in both states. The study offers some suggestive evidence that SLO scores distinguish high and low performers—identifying high-performing teachers in low-performing schools and vice versa—while also aligning with results from other measures of teacher performance. Moreover, in completing the SLO forms, participating teachers and administrators appeared to set goals for subgroups of students and monitored their progress, as intended by the states, and at the end of the pilot year surveyed teachers tended to see the process as beneficial to their students and their own professional growth. Together these findings suggest some benefits from initial SLO implementation.

At the same time, SLO scores differed between elementary and secondary school teachers in both pilots, with elementary school teachers scoring better on average in Arizona and secondary school teachers scoring better in Utah. It is possible that each state's SLO process (and its related training and guidance for raters) may need to be adjusted to better account for variations in teacher performance at the two grade spans. Utah officials might also consider providing more direct oversight of SLO documentation, given the different forms used by general and special education teachers to record their progress in the 2014 pilot. As states continue to guide and support administrators and teachers in implementing SLOs—for example, clarifying expectations around reliable scoring and what constitutes a high-quality SLO in different grades and subject areas—they might also consider reinforcing local work already under way in interim assessment and progress monitoring, which represent the foundational teaching concepts underlying SLOs.

The study offers some suggestive evidence that SLO scores distinguish high and low performers—identifying high-performing teachers in low-performing schools and vice versa—while also aligning with results from other measures of teacher performance

Limitations of the study

This exploratory study relied on analyses of SLO documents and of survey results and administrative data collected by state department of education representatives. Given the small number of districts and teachers participating in the study (in relation to the populations in each state) and the voluntary nature of recruitment, the study samples are not representative of any larger group. Certain key data were also unavailable from the states. Specifically, teacher survey data were unavailable for Arizona, while neither 2013/14 school performance grades nor results from any other teacher effectiveness measures were available for Utah. In addition, the two pilot contexts—for example, the number of participating teachers, the nature and timing of the SLO rollout, and the subject areas and grade spans taught by participating pilot teachers—differed substantially, limiting the study team's ability to compare the two SLO processes.

It is possible that differences in teachers' SLO scores may be influenced by variations in quality and rigor, but the content review did not assess this aspect of the SLO process. State officials in both states instructed participating teachers and administrators to apply the state's checklist or rubric as they implemented the process to ensure that each SLO was of acceptable quality before it was used to award teacher ratings. SLOs that did not meet state standards were to be revised. The study team assumed that each state's minimum quality standards were met and did not empirically confirm that SLOs were held to quality standards. Further, it is acknowledged that, even so, there will be variation in the quality of objectives.

Finally, this was a study of two pilot programs, and findings may differ in subsequent years as training, support, and experience with the processes increase.

Appendix A. Data and methodology

This appendix describes the data sources and the general methods used for this study, including a detailed explanation of the coding methods for student learning objective (SLO) content.

Data sources

This descriptive study uses data that Arizona and Utah representatives collected during pilot programs to include SLOs in teacher evaluations during the 2013/14 school year, including teachers' completed SLO documents/forms and final scores, results from end-of-year online surveys of participating pilot teachers in Utah,[12] and existing state administrative data (including school and district performance grades from the state). In Arizona 363 teachers from four volunteering districts participated in the SLO pilot (each had at least one SLO score), and SLO documents were made available for 323 (88 percent) of them. Utah's pilot teacher sample included 82 volunteering teachers from 10 districts, with final SLO documents available for 67 (82 percent), final SLO scores available for 63 (77 percent), and end-of-year survey responses from 56 (68 percent) of these teachers.

The study team's content analysis examined 282 classroom SLOs and 353 targeted SLOs from Arizona and 29 classroom SLOs and 87 targeted SLOs from Utah. Pilot teachers in Arizona included an average of 9 students in their targeted SLO groups, while teachers in Utah had an average of 14 (perhaps the result of the high proportion of special education teachers represented in Utah's SLO pilot).

Methodology

To explore the first part of research question 1 regarding the variation in SLO content, researchers reviewed, coded, and tabulated the information from the 323 Arizona SLO forms and from the 67 Utah forms. First, the assessments used in the SLO were categorized as national or state standardized tests, vendor-developed content assessments, district- or school-wide assessments, or teacher-developed classroom tests. Then the content of the SLO was analyzed qualitatively in two ways. First the SLOs were identified as assessment-focused (focused on how students will perform on an assessment, without mentioning the specific material they will learn), learning content–focused (focused on what students will learn in the class, without mentioning their specific performance on an assessment), or both. Next, the nature of the goals specified in the teachers' SLO was identified, such as demonstrating a physical skill (for example, playing a piece of music, acting a theatrical part, or running a specified distance in a specified time) or increasing a test score by a certain amount. In applicable cases, SLO were double-coded. Coding methods for addressing this question are explained in more detail in the next section.

To address the second part of question 1, regarding variation in SLO scores, the study team examined the final SLO scores and also compared them with school performance and teachers' grade span. The analysis involved cross-tabulating the distribution of participating pilot teachers' end-of-year SLO scores, first by the letter grades earned by their schools from the Arizona state accountability system and then by the grade span taught (elementary or secondary), using significance tests to assess whether teachers' pilot SLO scores were equally likely to appear across schools with different letter grades (A–F) or across the

elementary or secondary grade spans.[13] At the time of the study, school letter grades were not available for Utah due to changes in Utah's testing and accountability policy.

To explore question 2, regarding any associations between teachers' SLO scores and the results from other forms of teacher evaluation such as classroom observation scores and student survey ratings of their teachers from 2013/14, the study team used correlations, cross-tabulations, and significance tests for the Arizona teacher sample. Because none of the 82 SLO pilot teachers in Utah had available observation or student survey scores in 2013/14, the analysis for question 2 focused exclusively on Arizona's pilot teachers.

To address research question 3, regarding teachers' perception of the SLO process, the analysis relied primarily on data from state-administered surveys. In May 2014 state education officials in Arizona and Utah emailed participating SLO pilot teachers a link to an online survey. A total of 56 Utah SLO pilot teachers (68 percent) responded. But the response rate was far lower in Arizona (approximately 20 percent). As a result, Arizona did not provide survey data to the study team, and no information from the Arizona surveys is reported here. To address this question for Utah, the study team summarized the survey data using response frequencies (percentage of agree/disagree/neither). Based on early anecdotal evidence from Utah State Office of Education officials, the study team also explored whether Utah pilot teachers' perceptions of SLO differed by grade span, by experience level, by school size, or by final SLO score received and whether perceptions differed for special educators and among teachers who revised their SLO targets mid-pilot.

Data collection. To explore the variability in SLO content, teacher's SLO documents and completed forms were collected in May and June 2014 by two state education agency partners (Arizona Department of Education and Utah State Office of Education) of the Regional Educational Laboratory West Education Effectiveness Alliance. Certain categories of information were provided on both states' forms, including the subjects and grade levels taught by the teacher, the number of students included in the targeted SLO, the names of the assessments used to evaluate students' academic progress, the targeted learning goals or the proficiency or progress sought by the teacher, the specific state academic content standards taught to students to achieve the targets, the instructional strategies teachers intend to employ to ensure their students met the SLO targets, and the results of the SLO—that is, the score awarded to the teacher on the SLO by his or her evaluator at the end of the 2013/14 school year. State officials redacted teacher names before providing the SLO documents in PDF format to the study team. The information was entered into an Excel® spreadsheet so the data could be reviewed, coded, and tabulated. Basic descriptive statistics were generated on, for example, the subjects and grades taught and the average number of students included in the targeted SLO.

Coding. A coding system was developed to analyze the remaining content of each SLO document. Following Lacireno-Paquet, Morgan, and Mello (2014), the assessments used in teachers' classroom and targeted SLOs were categorized as one of the following types of measures: national or state standardized assessment, vendor-developed content assessment, district- or school-wide assessment, or teacher-developed classroom assessment. National or state standardized assessments include those that are used at the national or state level to measure student performance (for example, Arizona's Instrument to Measure Standards). The study team reviewed the Arizona Department of Education and Utah State Office of Education websites to determine the state tests used. Vendor-developed content

assessments measure student progress in a particular content area (but are not used as national or state standardized tests). District- or school-wide assessments are used across the district or school (but are not vendor-developed content tests), such as performance rubrics or report cards. Finally, teacher-developed classroom assessments are alternative assessments designed by the teacher to gauge student understanding of the class material (and are not specifically labeled to be used across the district or school), such as year-end subject exams or classroom projects.

The content of the SLOs was qualitatively analyzed in two ways, using coding schemes that reduced the narrative data into a manageable set of categories and allowed the study team to aggregate and quantify the data. First, using the findings from the Community Training and Assistance Center (2004) study in Denver, the Arizona and Utah SLOs were categorized as being assessment-focused or learning content–focused (see table A1 for details). For this study a third code, "both," was added to accommodate SLOs that included contents that were both assessment-focused and learning content–focused. The code "both" refers to SLOs that describe the content that will be learned in class as well as the specific performance expected on an assessment or rubric.

A second content coding scheme was created to analyze SLO content at a deeper level—to categorize the nature of the objectives. A preliminary review was first conducted on the Arizona classroom SLOs, with two members of the study team broadly reviewing the SLOs to become familiar with the content. The codes emerged through an inductive process (Miles, Huberman, & Saldana, 2014) in which the two study team members reviewed the first SLO goal and created a descriptive code to categorize the nature of that SLO, then reviewed the second SLO and either verified the first descriptive code or created a new code to capture the second SLO's content, and so on. This process was conducted for the first 30 Arizona classroom SLOs, at which point the study team members felt that they had reached the limit of different categories of SLOs.

The study team members then clearly defined the classroom SLO categories with basic decision rules for coding and provided examples for each code. The resulting codes for

Table A1. Coding scheme for the focus of the student learning objective

Code	Definition
Assessment-focused	Student learning objective (SLO) is focused on how students will perform on an assessment, without mentioning the specific material they will learn. Example: All students will achieve benchmark on Dynamic Indicators of Basic Early Literacy Skills assessment as measured by their final composite score of the year.
Learning content–focused	SLO is focused on what students will learn in the class, without mentioning their specific performance on an assessment. Example: Students will show knowledge gained in each unit of instruction: Maps and Geographic Tools, Earth's Processes, Human Environment Interaction, Population Distribution and Development, Migration and Culture.
Both	SLO mentions the content that will be learned in class and the specific performance expected on an assessment or rubric. Example: All students will demonstrate mastery of the scientific method, basic analytical and quantitative skills used in science, and basic physics and chemistry concepts, with an average of at least 70 percent accuracy on the chemistry/physics final exam for both semesters.

Source: Authors' compilation.

classroom SLOs included demonstration, growth, meets range, and proficient. The two study team members then coded the next five SLOs together to verify the coding scheme. They then coded 20 other SLOs independently and compared the results, finding disagreement on only 2. Each of these two SLOs had been coded the same on the first code, but one study team member assigned a second code to both. The study team members decided that both codes were needed for those SLOs and that double-coding would be appropriate. In addition, the study team members decided to break down the demonstration code into "demonstrates knowledge through project completion" and "demonstrates a physical skill."

Some language in the SLO code definitions was slightly modified due to this test round, but the overall structure of the coding scheme worked well with all of the Arizona classroom SLO data. The final codes for classroom SLOs included demonstrates knowledge through project completion, demonstrates a physical skill, meets growth target, meets test range or benchmark, and meets proficiency level (table A2). One study team member then completed the coding for the remaining SLOs, and the second verified the coding.

While the classroom SLOs stated what teachers thought their students would learn or be able to perform by the end of the performance period, the targeted SLOs tended to focus on the type of growth students would demonstrate. Therefore, study team members decided that a new coding scheme was required to accurately capture the nature of the targeted SLOs. The same inductive process used on the classroom SLO codes was applied to create and test the targeted SLO codes. Here, too, after the study team members had coded 20 SLOs separately, they identified 2 that they had coded differently. As with the classroom SLOs, they decided that both codes were appropriate and that those two SLOs should be double coded. Some language in the SLO code definitions was modified in a minor way following this test round. The final codes for targeted SLOs included demonstrates completion, no growth; demonstrates growth; meets test score, no growth; meets percentage growth; meets proficiency level growth; and completes project, no growth (see table A2). One study team member then completed the coding for the remaining SLOs, and the second verified the coding.

The Utah SLOs were received later and included more text than the Arizona SLOs, but the study team felt that the same codes would apply. Two study team members then coded 10 Utah SLOs together to ensure that the coding schemes fit with the Utah data as well. One study team member then coded the remaining SLOs, and the second verified the coding.

Table A2. Coding scheme for the goals of the student learning objective

Code	Definition	Example
Classroom student learning objective		
Demonstrates knowledge through project completion	Students will apply their knowledge in the completion of a final project or task, such as a lab investigation or final report.	All students will prepare a comprehensive and quantitative 10 year plan.
Demonstrates physical skill	Students will physically demonstrate a skill or procedure, such as bench press, typing, behavior, or music performance.	All grade-3 Fit Kids will demonstrate the ability to execute sit-ups, jumping jacks, and grapevines.
Meets growth target	A specified growth in percentage or score is provided, for example, 65 percent growth or an increase of 10 points.	All students will increase their baseline score by 5 growth points by April 2014.
Meets test range or benchmark	Students will score within the meets range on an assessment, or will meet benchmark; no score or percent is given.	All students will score within the meets range on the Galileo Math end-of-year comprehensive assessment by April 2014.
Meets proficiency level	Students will score at a specified level of proficiency on an assessment—for example, 60 percent proficiency or 70 out of 80 correct.	All students will achieve a proficiency score of 1233 by May 2014.
Targeted student learning objective		
Demonstrates completion, no growth	Students will complete a physical task (for example, a mile run); growth is not specified.	Targeted cadets are required to perform the 30 minute count drill sequence as well as give the commands.
Demonstrates growth	Any language relating to growth in demonstrating a physical skill or behavior that is not a written demonstration (for example, bench press, sight words).	Each student in the targeted group will reach his or her healthy fitness zone for at least one additional test from baseline to post-assessment.
Meets test score, no growth	Students will meet a cutscore on an assessment—for example, 75 percent or higher; growth is not specified.	The identified students will pass the Civil War Assessment at 65 percent.
Meets percentage growth	Includes a specific percentage of growth—for example, increase score by 10 percent.	All targeted students will grow 125 percent of the annual expected growth between Galileo Math pre-test and post-test by April 2014.
Meets proficiency level growth	Includes a specific reference to growth in proficiency—for example, Intensive to Strategic—and mentions a specific assessment or rubric, such as Dynamic Indicators of Basic Early Literacy skills (DIBELS).	Each student in the targeted SLO group will grow one proficiency level as measured by comparing the DIBELS composite score of Benchmark 1 to Benchmark 3.
Completes project, no growth	Students will successfully complete a project or assignment—for example, a 10 year plan; growth is not specified.	Targeted students will complete all assignments regarding *Of Mice and Men.*

Source: Authors' compilation.

Appendix B. Correlations between Arizona teacher measures, by grade span

Table B1 displays the correlations between several of the teacher measures piloted in Arizona in 2013/14, with separate results for elementary and secondary school pilot teachers. As shown, the correlations were slightly stronger (and more often statistically significant) among secondary school teachers than among elementary school teachers, particularly for the student survey results.

Table B1. Correlations between evaluation measures differed for Arizona elementary and secondary school pilot teachers

Teacher group and measure	Classroom student learning objective score	Targeted student learning objective 1 score	Average observation score[a]	Total observation points[b]	Average student survey score[c]
Elementary school teachers (n = 100–194)					
Targeted student learning objective score	0.64*				
Average observation score[a]	0.12	0.06			
Total observation points[b]	0.16*	0.00	0.94*		
Average student survey score[c]	0.17	0.03	0.12	0.11	
Student survey points[b]	0.15	0.01	0.05	0.05	0.93*
Secondary school teachers (n = 151–165)					
Targeted student learning objective score	0.59*				
Average observation score[a]	0.18*	0.13			
Total observation points[b]	0.18*	0.14	0.99*		
Average student survey score[c]	0.24*	0.20*	0.31*	0.31*	
Student survey points[b]	0.22*	0.18*	0.28*	0.28*	0.95*

* Significant at $p < .05$.

a. Mean score across the teacher's 22 Danielson Framework for Teaching component ratings.

b. The state's evaluation model assigns a point value to each component measure's results to establish a uniform summative point scale.

c. Students rated more than 30 teacher behaviors on a four-point scale (1 = never, 2 = some of the time, 3 = most of the time, 4 = always). An average rating was calculated across all behaviors for each student, and then the teacher was assigned an aggregate score that was the proportion of the average ratings that were 3 or above.

Source: Authors' analysis of 2013/14 data provided by the Arizona Department of Education.

Appendix C. Response frequencies for Utah's pilot teacher survey

This appendix contains tables of response frequencies for each of the Utah pilot teacher survey questions (tables C1–C5), as well as for questions where significant teacher subgroup differences were evident (tables C6–C14).

Table C1. Utah pilot teacher survey respondents taught different subjects

Pilot teachers	Career and technical education	Fine arts	Science	Social studies	Special education
Number	8	8	5	8	27
Percent	14.3	14.3	8.9	14.3	48.2

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C2. Utah pilot teacher survey respondents taught different grade spans

Pilot teachers	Elementary school	Middle school	High school
Number	20	17	19
Percent	35.7	30.4	33.9

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C3. Utah pilot teacher survey respondents had varying years of experience at their schools

Pilot teachers	Fewer than 5 years	5 10 years	More than 10 years
Number	26	16	14
Percent	46.4	28.6	25.0

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C4. Utah pilot teacher survey respondents had varying years of teaching experience

Pilot teachers	Fewer than 5 years	5–10 years	More than 10 years
Number	12	17	27
Percent	21.4	30.4	48.2

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C5. Utah pilot teachers' agreement with statements about the student learning objective process varied

Percent of survey respondents

Please indicate the extent to which you agree or disagree with the following statements based on your participation in the state's SLO process this year (2013/14).

Statement	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
In theory, I believe an SLO process has the potential to provide an accurate assessment of my teaching effectiveness.	48.2	33.9	17.9
The overall training I received on the new SLO process was adequate for me to participate in the SLO process this year.	46.4	35.7	17.9
Through the SLO process this year I gained new knowledge about effective ways to assess students.	30.4	44.6	25.0
I am confident that the assessment(s) used as part of my SLO(s) appropriately measured my students' growth.	46.4	32.1	21.4
My mid-year check-in with my administrator to discuss my SLO(s) was helpful.	39.3	21.4	39.3
The SLO process this year improved the quality of my conversations with my administrator about instruction.	21.4	33.9	44.6
The SLO process this year improved the quality of my conversations with my administrator about assessment.	23.2	33.9	42.9
The SLO process this year improved the quality of my conversations with my fellow teachers.	33.9	35.7	30.4
The SLO process helped inform my professional growth.	44.6	28.6	26.8
I have changed the instruction in my classroom as a result of the implementation of SLOs.	33.9	39.3	26.8
The final SLO score I received accurately reflected my contributions to my students' growth this year.	42.9	25.0	32.1
My participation in the SLO process has benefited my students.	44.6	19.6	35.7
Overall, the SLO process was worthwhile.	51.8	21.4	26.8
Rather than SLOs, I would prefer to be held accountable for schoolwide math and/or reading test scores.	7.1	66.1	26.8

Note: Strongly agree and agree responses and strongly disagree and disagree responses were combined for this table. Percentages may not sum to 100 due to rounding.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C6. Utah pilot general and special education teachers had different views on the student learning objective process changing their instruction

Percent of survey respondents

I have changed the instruction in my classroom as a result of the implementation of SLOs.

Type of teacher	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Special education teachers (n = 29)	17.2	58.6	24.1
General education teachers (n = 27)	51.9	18.5	29.6

Note: Fisher's exact test: p = .004. Percentages may not sum to 100 due to rounding.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C7. Utah pilot general and special education teachers had different views on the student learning objective process benefiting students

Percent of survey respondents

My participation in the SLO process has benefited my students.

Type of teacher	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Special education teachers (*n* = 29)	27.6	24.1	48.3
General education teachers (*n* = 27)	62.9	14.8	22.2

Note: Fisher's exact test: $p = .029$. Percentages may not sum to 100 due to rounding.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C8. Utah pilot elementary and secondary school teachers had different views on the potential accuracy of the student learning objective process

Percent of survey respondents

I believe an SLO process has the potential to provide an accurate assessment of my teaching effectiveness.

Grade span	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Elementary (*n* = 20)	25.0	45.0	30.0
Secondary (*n* = 36)	61.1	27.8	11.1

Note: Fisher's exact test: $p = .025$.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C9. Utah pilot elementary and secondary school teachers had different views on their gaining assessment knowledge through the student learning objective process

Percent of survey respondents

Through the SLO process this year I gained new knowledge about effective ways to assess students.

Grade span	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Elementary (*n* = 20)	10.0	60.0	30.0
Secondary (*n* = 36)	41.7	36.1	22.2

Note: Fisher's exact test: $p = .043$.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C10. Utah pilot elementary and secondary school teachers had different views on the student learning objective process informing their professional growth

Percent of survey respondents

The SLO process helped inform my professional growth.

Grade span	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Elementary (*n* = 20)	15.0	40.0	45.0
Secondary (*n* = 36)	61.1	22.2	16.7

Note: Fisher's exact test: $p = .002$.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C11. Utah pilot elementary and secondary school teachers had different views on the student learning objective process benefiting students

Percent of survey respondents

My participation in the SLO process has benefited my students.

Grade span	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Elementary (*n* = 20)	20.0	30.0	50.0
Secondary (*n* = 36)	58.3	13.9	27.8

Note: Fisher's exact test: $p = .019$.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C12. Utah pilot elementary and secondary school teachers had different views on the student learning objective process being worthwhile overall

Percent of survey respondents

Overall, the SLO process was worthwhile.

Grade span	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Elementary (*n* = 20)	20.0	30.0	50.0
Secondary (*n* = 36)	69.4	16.7	13.9

Note: Fisher's exact test: $p = .001$.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C13. Utah pilot teachers at different scoring levels had different views on student learning objective results accurately reflecting their contributions to their students' growth

Percent of survey respondents

The final SLO score I received accurately reflected my contributions to my students' growth this year.

Average SLO score received	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
2 – Partially meets expectations (*n* = 11)	9.1	27.3	63.6
3 – Meets expectations (*n* = 32)	46.9	28.1	25.0
4 – Exceeds expectations (*n* = 8)	75.0	12.5	12.5

Note: Fisher's exact test: $p = .031$.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Table C14. Utah pilot teachers who revised their targets mid-pilot had different views on the student learning objective process improving teacher conversations

Percent of survey respondents

The SLO process this year improved the quality of my conversations with my fellow teachers.

Mid pilot revisions	Agree or strongly agree	Disagree or strongly disagree	Neither agree nor disagree
Teachers who revised SLO targets mid-pilot (*n* = 13)	7.7	38.5	53.8
Teachers who did not revise their SLO targets (*n* = 43)	41.9	34.9	23.3

Note: Fisher's exact test: $p = .034$. Percentages may not sum to 100 due to rounding.

Source: Authors' analysis of 2014 survey data provided by the Utah State Office of Education.

Notes

1. Although Arizona's SLO form provided space for pilot teachers to identify the state standards and learning strategies they were targeting in their SLO, this information was provided on less than 20 percent of the forms submitted for review. However, in one smaller district (pilot district C), 91 percent of participating teachers included this information as part of their SLO.
2. The Wilcoxon-Mann-Whitney test indicated a statistically significant difference ($z = 2.9$, $p = 0.004$) between the distributions of Arizona teachers' classroom SLO and targeted SLO scores (with targeted SLO scores having higher rank). For the purposes of this analysis, Arizona teachers' targeted SLO scores were rounded to the nearest integer.
3. The rank correlations (with pilot school letter grades converted to ordinal numeric values, that is, D = 1, C = 2, B = 3, and A = 4) were 0.32 ($p < .01$) for classroom SLO scores and 0.34 ($p < .01$) for targeted SLO scores. Teachers' SLO scores did not have a significant relationship with their schools' letter grades from the prior year (2012/13).
4. For Arizona pilot districts, 2013/14 marked the second year of implementation of both the Danielson Framework for Teaching (which assesses four domains of teaching: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities) and the state's model student surveys (separate surveys are administered for grades 3–5 and grades 6–12, based on public domain items from Cambridge Education's Tripod Student Perception Survey; for details, see Arizona Department of Education, 2013a, pp. 68–73).
5. A higher proportion (42 percent) of the physical demonstration SLOs received "exceeds expectations" scores ($p < .01$); for example, only 6 percent of proficiency level growth SLOs received an "exceeds expectations" score.
6. There may be several reasons for the preponderance of teacher-developed tests in Utah. Not only did the SLO pilot include predominantly teachers working in areas that do not traditionally tend to be assessed through standardized tests, but state guidance also suggested that assessments be selected or developed "to best measure the knowledge and skills found in the learning goal" and be "accompanied by clear criteria or scoring rubrics to determine student learning" (Utah State Office of Education, 2014b, p. 4). Teachers were also asked on the SLO template to explain how they would use the information from the assessment "to differentiate instruction for all students" to achieve the learning goal (p. 14).
7. Utah State Office of Education officials did not specify how these scores were to be calculated or determined by evaluators. The state planned to offer more explicit scoring guidance in 2014/15.
8. A subgroup of 15 participating Utah pilot teachers revised their SLO targets mid-pilot. Although the final SLO scores among these teachers were 0.17 pooled standard deviation higher than the SLO scores among teachers who did not revise their targets, the difference was not significant ($p = .30$).
9. Specifically, participating secondary school teachers scored 0.29 pooled standard deviation higher, on average, than elementary pilot teachers.
10. The study team analyzed the survey results separately by teacher experience level, by grade span taught, by school size, by final SLO score received, for special education teachers, and for teachers who revised their SLO targets mid-pilot. Additional information about the subgroup breakdowns, including statistically nonsignificant results, is available from the study team on request.

11. Higher scoring teachers were more likely to agree that their final score accurately reflected their contribution to their students' growth (see table C13 in appendix C). Responding Utah pilot teachers' perceptions did not appear to vary according to their school's enrollment or their years of experience.
12. Although teacher surveys were administered in Arizona, the low response rate (approximately 20 percent) precluded reporting results from that state.
13. For these cross-tabulations, Arizona teachers' targeted SLO scores were rounded to the nearest integer.

References

Arizona Department of Education. (2013a). *Teacher evaluation process: An Arizona model for measuring educator effectiveness (2013/14)*. Retrieved December 10, 2013, from http://www.azed.gov/teacherprincipal-evaluation/files/2012/10/teacher-evaluation-v4.0-website-update-11_22_13-sl.pdf.

Arizona Department of Education. (2013b). *The student learning objective handbook*. Phoenix, AZ: Author.

Community Training and Assistance Center. (2004). *Catalyst for change: Pay for performance in Denver, final report*. Boston, MA: Author.

Community Training and Assistance Center. (2013). *It's more than money: Teacher Incentive Fund—Leadership for educators' advanced performance, Charlotte–Mecklenburg Schools*. Boston, MA: Author.

Danielson Group. (2011). *The Framework for Teaching*. Retrieved November 5, 2014, from http://www.danielsongroup.org/article.aspx?page=frameworkforteaching.

Duncan, A. (2010). Overview Information: Race to the Top Fund; Notice inviting applications for new awards for fiscal year 2010. *Federal Register, 75*(68), 18171–18185.

Gill, B., Bruch, J., & Booker, K. (2013). *Using alternative student growth measures for evaluating teacher performance: What the literature says* (REL 2013–002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. http://eric.ed.gov/?id=ED544205

Goe, L., & Holdheide, L. (2011). *Measuring teachers' contributions to student learning growth for nontested grades and subjects*. Washington, DC: National Comprehensive Center for Teacher Quality. http://eric.ed.gov/?id=ED520722

Jerald, C. D. (2012). *Movin' it and improvin' it! Using both education strategies to increase teaching effectiveness*. Washington, DC: Center for American Progress. http://eric.ed.gov/?id=ED535645

Lachlan-Haché, L., Cushing, E., & Bivona, L. (2012). *Student learning objectives as measures of educator effectiveness: The basics*. Washington, DC: American Institutes for Research.

Lacireno-Paquet, N., Morgan, C., & Mello, D. (2014). *How states use student learning objectives in teacher evaluation systems: A review of state websites* (REL 2014–013). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. http://eric.ed.gov/?id=ED544768

Marion, S., DePascale, C., Domaleski, C., Gong, B., & Diaz-Biello, E. (2012). *Considerations for analyzing educators' contributions to student learning in non-tested subjects and grades with a focus on student learning objectives*. Dover, NH: Center for Assessment.

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Proctor, D., Walters, B., Reichardt, R., Goldhaber, D., & Walch, J. (2011). *Making a difference in education reform: ProComp external evaluation report 2006–2010*. Prepared for the Denver Public Schools. Denver, CO: The Evaluation Center, University of Colorado.

Schmitt, L., & Ibanez, N. (2011). *AISD REACH program update: 2009–2010 Texas Assessment of Knowledge and Skills results and student learning objectives*. Austin, TX: Austin Independent School District, Department of Program Evaluation.

Tennessee Department of Education. (2012). *Teacher evaluation in Tennessee: A report on year 1 implementation*. Nashville, TN: Author. http://eric.ed.gov/?id=ED533726

U.S. Department of Education. (2012). *Laws & guidance, elementary & secondary education, ESEA flexibility.* Retrieved March 15, 2015, from http://www.ed.gov/esea/flexibility.

Utah State Office of Education. (2014a). *Utah's handbook for student learning objectives*. Salt Lake City, UT: Author.

Utah State Office of Education. (2014b). *Utah's SLO toolkit*. Salt Lake City, UT: Author.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009*). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project. http://eric.ed.gov/?id=EJ857720

Wright, B. (2014). State school board frets about ugly Utah report cards. *Utah Standard-Examiner.* Retrieved November 3, 2014, from http://www.standard.net/Education/2014/09/06/School-grades-going-down-to-raise-the-bar.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research