SREB

Toward Trustworthy and Transformative Classroom Observations

Progress, challenges and lessons in SREB states

Southern
Regional
Education
Board

SREB.org

Inside

Teachers aren't the only ones who need feedback. Please let us know how well this report met your needs. It will take less than one minute. Visit *http://tinyurl.com/sreb15E01*, or scan the code with your smart device.

Toward Trustworthy and Transformative Classroom Observations

Progress, challenges and lessons in SREB states

February 2015

SREB | Educator Effectiveness

Executive Summary

Classroom observations continue to carry significant weight in the evaluation system of most states and districts and remain a part of the evaluation experience of every teacher. This continued focus on classroom observation reflects the belief shared by policymakers and educators alike that an enhanced observation and feedback system for teachers can improve teaching and student learning.

Despite some initial reports of positive perceptions and experiences with new-generation classroom observations, many states are also facing technical, logistical and human capital challenges in implementing and sustaining a robust, reliable and fair observation system.

The amount of time required to train observers on a new framework and in the dialogue process may have been understimated. Ensuring that observers are consistent across multiple observations and accurately measure teaching quality based on relatively short observations is a challenge with no easy solution.

> *Without an effective system and culture of educator support, could new classroom observation systems become an "exercise in compliance" much like their predecessors?*

The following key recommendations from this report for states implementing classroom observation policies are based on successful practices seen in the region and emerging research on the topic:

1. **Create opportunities for educators to develop an accurate and shared understanding of the new teaching quality framework.** This means extending the learning process beyond the initial introduction to the observation rubric. High-quality tools such as an enhanced rubric with look-fors and a video library showing highly effective teaching in all subjects and grades could be worthwhile.

2. **Monitor train-the-trainer models and ensure all observers receive minimum training.** Train-the-trainer models must be monitored to ensure consistency of information. If strong oversight of trainings is not possible, consider creating standardized training materials (presentations, videos, etc.) that could deliver at least some key information with consistency.

    States that allow local control over teacher evaluation systems should monitor progress across districts to ensure that all educators have the minimum training and resources needed to implement a new classroom observation and feedback system. The state should facilitate knowledge- and resource-sharing between districts.

3. **Develop or continue some form of observer assessment and certification.** This will ensure that observation ratings and feedback are meaningful and reliable enough to guide teaching improvement. However, assessment and certification should not be viewed as a substitute for ongoing training and calibration.

4. **Train observers equally on giving written and oral feedback and on giving accurate ratings.** This ensures teachers benefit from classroom observation feedback. Consider the support teachers need to change their teaching practice in response to this feedback. Although teacher development is typically viewed as a separate domain from teacher evaluation, systems to evaluate teachers ultimately need to be linked toward a development system and culture of educator support.

5. **Continually monitor implementation, whether it's through systemic data collection or random statewide audits.** Efforts to evaluate the system throughout its development is key to system refinement and ensuring that observations are happening as intended with desired outcomes and impact on teaching quality.

Introduction

Classroom observation for improving teaching and learning

Classroom observation is a powerful component of teacher evaluation systems. It measures instructional practice, provides clarification on what effective teaching looks like and gives teachers the concrete and actionable feedback they need to improve teaching practice. Although student outcome measures seem to dominate current discussions about teacher evaluation, standardized test scores count toward the appraisal of one-fourth to one-third of teachers, while observations are part of the evaluation experience of every teacher. Observation carries significant weight in the evaluation system of most states and districts, contributing to half or more of the performance score in states using a percentage formula. It counts as much as student outcomes in states using a decision-matrix approach.

> *"The classroom observation component of new teacher evaluation models appears least controversial. … At the same time, many SREB states report facing technical, logistical and human capital challenges."*

The classroom observation component of new teacher evaluation models appears least controversial — and even well received — in SREB states that have begun system pilots or implementation. Part of the reason may be that the use of classroom observations for judging teaching quality is more familiar to educators. Beyond its familiarity, many educators welcome a classroom observation process that provides them with more specific and actionable feedback to guide teaching improvement.

At the same time, many SREB states report facing technical, logistical and human capital challenges in implementing a rigorous, evidence-based observation system. Every step — introducing a new framework, training observers and teachers, assuring quality and monitoring for continuous improvement — demands substantial state and local resources. Fidelity of system implementation requires evaluators and teachers to take on massive learning and change. States are aware of the many validity threats to observation ratings and multiple barriers to effective feedback. Without a strong system of support, some have noted, classroom observation could easily fall short of its potential and become an exercise in compliance.

SREB Classroom Observation Project

This report synthesizes a review study conducted by SREB in 2014 to learn about classroom observation implementation across the SREB region. We systematically reviewed state policy documents and publicly available information to gather and catalog information about each state's teacher evaluation system, classroom observation framework, observation instruments, training resources, assessment and certification policies, and system monitoring efforts. We also studied some district systems in states with decentralized decision making (Florida, Maryland, South Carolina, Texas).

Following document reviews, we interviewed state and district staff with close knowledge of classroom observation implementation in their agencies. Interviewees elaborated on the publicly available information and provided additional insight about their agency's successes, challenges, questions and future direction regarding classroom observation implementation.

Classroom Observations in SREB States

New classroom observation systems across SREB states generally have a dual purpose:

1. To accurately and fairly assess teaching effectiveness, and

2. To provide teachers with more regular, specific and actionable feedback to improve teaching practice.

Although the systems may have similar goals, their contexts vary. For example, some states lean toward a statewide educational system while others tend to allow decentralized decision. Thus, a range of approaches, models, tools and processes to improve classroom observations have emerged in the SREB region.

Key differences in context and characteristics of teacher evaluation systems in SREB states are shown in Table 1. The purpose of this report is to describe overall progress and common challenges states face in the areas of a) framing observations, b) training observers, c) ensuring observation quality, d) building a system and culture of educator support, and e) monitoring for continuous improvement. We highlight lessons from states that have had success in each area that other state and district leaders should consider in continuing efforts to develop classroom observation systems that would better teaching and learning.

TABLE 1: Overview of classroom observation in SREB states

State	Teacher evaluation system	Locus of evaluation system	Implementation in 2014-15	Weight of classroom observation
Alabama	Alabama's Educator Effectiveness System	Local	Pilot	30% of formula
Arkansas	Teacher Excellence and Support System (TESS)	State	Full – first year, overall rating will not be calculated	Dominant factor in a decision matrix
Delaware	Delaware Performance Appraisal System II (DPAS II)	State	Full	1 of 2 factors in a decision matrix
Florida	Florida's Teacher Evaluation System	Local	Full	50% of formula
Georgia	Teacher Keys Effectiveness System (TKES)	State	Full – first year	50% of formula
Kentucky	Professional Growth and Effectiveness System (PGES)	State	Full – first year	Significant factor in decision matrix
Louisiana	Compass	State	Full	50% of formula
Maryland	Teacher and Principal Evaluation (TPE)	Local	Full	Significant contributor[1]
Mississippi	Mississippi Teacher Evaluation System (M-STAR)	State	Full – first year	50% of formula
North Carolina	North Carolina Educator Evaluation System (NCEES)	State	Full	1 of 2 factors in decision matrix
Oklahoma	Teacher and Leader Effectiveness (TLE)	Local	Full – second year, overall rating will not be calculated	50% of formula
South Carolina	Improved Educator Support and Evaluation	Local	Full – first year, state model optional	50% of formula
Tennessee	Tennessee Educator Acceleration Model (TEAM)	State	Full	50-60% of formula
Texas	Texas Teacher Evaluation and Support System (T-TESS)	Local	Pilot	70% of formula
Virginia	Uniform Performance Standards and Evaluation Criteria	Local	Full	Informs performance-rating decisions[2]
West Virginia	West Virginia Educator Evaluation System	State	Full	Contributes to 80% of formula[3]

Sources: Review of state policy documents, and interviews with state education agencies in summer 2014

[1] In Maryland, classroom observation ratings contribute directly to 38.5 percent of the professional practice measure and as much as 40 percent of the student growth measure.

[2] In Virginia, observations are one piece of data used to provide a comprehensive picture of a teacher's level of performance related to Virginia's performance standards.

[3] In West Virginia, evaluators are to use evidence from multiple sources (anecdotal records, assessments, classroom observation, communications, portfolio, student feedback, student work samples, etc.) to judge teaching quality.

Framing Observations

In summer 2014, all SREB states were either piloting or implementing a new research-based classroom observation framework. All new frameworks were based on research identifying teaching practices that were most important for improving student learning. The majority of state models are built on frameworks developed by researchers such as Charlotte Danielson, James Stronge, Robert Marzano and the National Institute for Excellence in Teaching (see Table 2). Locally-developed models were similarly informed by research about classroom practices that raise student achievement.

TABLE 2: Observation frameworks in SREB states

Danielson-based	Marzano-based	Stronge-based	NIET-based	Local model
Arkansas	Florida [1]	Georgia	Oklahoma [4]	Alabama
Delaware	Oklahoma [4]	Virginia	South Carolina [5]	Florida [1]
Florida [1]			Tennessee	Louisiana [2]
Kentucky			Texas [6]	Maryland [3]
Louisiana [2]				North Carolina
Maryland [3]				Oklahoma [4]
Mississippi				South Carolina [5]
Texas [6]				Texas [6]
				West Virginia

Sources: Review of state policy documents and interviews with state education agencies in summer 2014

[1] Florida districts can choose to use the state Marzano-based model or any other locally-determined model. Many local models are Danielson-based.

[2] Most Louisiana districts utilize the state framework and rubric but are also allowed to submit an alternative rubric for use.

[3] Maryland districts can choose to use the state Danielson-based model or any other locally-determined model.

[4] The majority of school districts in Oklahoma currently use Tulsa's model. Districts can choose other models that have been state approved, which currently include the Marzano model and NIET's TAP framework.

[5] South Carolina is currently piloting two systems: the locally developed Enhanced ADEPT and the South Carolina Teaching Standards developed by NIET.

[6] Texas's state observation process is based on the NIET cycle of pre-conference, observation and post-conference, while the rubric is locally developed by a steering committee to reflect state teaching standards. Districts can also use local systems, with state approval.

Engaging stakeholders

Many states have made commendable efforts to engage stakeholders in developing a teacher evaluation and classroom observation system that they find credible. Incorporating input and feedback, especially from administrators and teacher groups, improves the likelihood that observations (and other components of the teacher evaluation systems) not only reflect research-based recommendations but also educators' understanding of what constitutes effective teaching.

Almost every state has a teacher evaluation advisory committee or working group that contributes significantly to system selection, design and refinement. In addition, many states gather wider input and feedback during system development through routine meetings, listening sessions, emails, phone calls and in-person communications. Engaging multiple stakeholder groups takes a great deal of time and energy, but state staff tend to agree that it is critical to creating goodwill and securing buy-in for the work.

Shared understanding

The ongoing challenge for many states is developing an accurate understanding of different levels of teaching quality that is shared by all educators. What is considered distinguished, proficient or unsatisfactory? For example, what is effective questioning, and what does that look like in a kindergarten class and in a high school class? Many states have developed a rubric that aims to clearly describe each performance level for each dimension of their state framework, while other states expect or recommend that rubrics be developed by local agencies.

> *The process of rubric development may be a meaningful way to engage educators and increase their ownership of the framework so they don't feel as if "it was something handed down to them from the state."*

The process of rubric development may be a meaningful way to engage educators and increase their ownership of the framework so they don't feel as if "it was something handed down to them from the state." District-created rubrics might be more valid if they are tightly aligned with existing district curriculum, policies and programs. At the same time, states and districts recognize vast differences in local resources and know that some districts lack adequate capacity to develop a high-quality rubric. State support is therefore important for ensuring all educators have classroom observation tools such as rubrics of acceptable quality.

Tools of support

Many states still need to develop or purchase more tools to support educators in forming a clear and detailed understanding of observation frameworks. States expressed a common need for richer resources to support educators in the foundational task of "becoming truly familiar with the observation rubric — what do you look for when observing a classroom, and what will it take to be exemplary?" In particular, many states recognize that educators want tools that provide more specific details and examples of what effective teaching looks like for various grades and subjects.

Several states have developed enhanced rubrics with added details (for example, specific look-fors). Louisiana and Tennessee have a growing online library of videos illustrating instructional expectations, something that most other states hope to have. A majority of states would eventually like to provide a comprehensive video library to illustrate authentic lessons at every performance level in various teaching contexts (for example, school levels, subject areas, school student populations). Other types of rubric resources that states currently provide are noted in Table 3. SREB will conduct a more in-depth study to catalog member states' efforts in the future.

TABLE 3: Sampling of rubric resources currently available in SREB states

Rubric resources	Examples
Video library	Louisiana's video library has examples of instruction at different levels to show instructional expectations.
	Tennessee provides educators access to the NIET best practices center portal and video library.
Enhanced rubrics	Kentucky's rubric has details of what observers should look for, and examples.
	Virginia has subject-specific frameworks (e.g., for English, Science, fine arts) with differentiated standards and elements.
	Louisiana has observation guides with "look fors" for ELA and math, and plans to produce more .
Professional learning materials	Alabama has e-learning courses aligned with each indicator of the observation framework.
	Arkansas has prepared presentations and activities for every indicator of the observation framework.
	Kentucky provides materials for local professional development, including a framework workbook, activities and vignettes for discussion.

Training Observers

Training those responsible for conducting observations (mostly, but not exclusively, administrators) is an essential undertaking across the SREB region and can be offered by state department staff, regional centers, district personnel or vendors. States with a statewide observation system are challenged to ensure that observers across districts and schools receive and understand the same information to implement the process with consistency. Decentralized states with many local systems cannot rely on standardized training or materials, but similarly need to ensure that observations are of uniform rigor to be fair to all teachers.

Delivering training for statewide systems

Many states provide some direct training to ensure that every observer receives the same information, although this is a resource-intensive method. Direct training occurs when a state department or vendor provides the same training to all trainees, often in face-to-face sessions. In theory, evaluators who receive the same training could be expected to implement observations with greater consistency. Such trainings are costly, not always possible to execute statewide, and tend to be infrequent. Still, the majority of SREB states devote resources for some direct training, arguing it is critical to communicate key messages to all system users and to give state staff opportunities to address major misconceptions that hinder system implementation.

The "train the trainer" model is popular but could inadvertently contribute to implementation inconsistency. Given the challenges of providing direct training as discussed above, many states provide direct training to select district personnel (and sometimes principals and teacher leaders), who are then responsible for providing local training. While this training approach helps minimize training costs and sounds reasonable in theory, a common concern is that secondary training "is where the trains go off the rails." Key strategies employed by SREB states to improve the consistency and quality of local training while managing cost are as follows:

- **Standardized training materials:** Several states using the "train the trainer" model provide secondary trainers with standardized training materials that include PowerPoints, talking points and handouts for trainees. For example, Oklahoma created a presentation assistance kit (PAK) with everything secondary trainers need to provide the same information they received to others in their district.

  Virginia has also created video modules and PowerPoints to be used for local trainings. Furthermore, Virginia provides explicit guidance for trainer selection to increase the likelihood that those receiving the primary training have suitable knowledge and skills, and are in ideal positions to offer training and support for their district.

- **Taped trainings and webinars:** Several states have created taped trianings and webinars that are accessible online as a viable means of providing consistent information to evaluators who do not or cannot attend in-person training for a variety of reasons (for example, distance or mid-year hiring). Mississippi archives all webinar trainings for this purpose. North Carolina has created training modules that are also accessible to evaluators anytime.

- **Supplemental web-based resources:** Several states have created supplemental resources to support evaluators beyond training sessions. Such resources include a video library (see earlier discussion), "quick guides" and other guiding documents, sample products and vignettes/scenarios that can be useful materials for ongoing professional learning at the local level (such as in professional learning communities). See Georgia and North Carolina for examples.

- **Ongoing communication:** Several states have also established a formal channel of communication with districts and schools — through which they can provide the most up-to-date and official information to educators statewide. Maryland publishes monthly updates that are archived on a state webpage. Oklahoma publishes a monthly newsletter on the state website, which educators can also choose to receive by email.

Supporting local system training

States with local decision making and those with several state models want to ensure that all observers receive adequate training. Providing standardized training and learning resources may be impractical or not feasible. In some cases, state involvement in locallycontrolled teacher evaluation systems is unwelcome.

Still, several state department leaders expressed the importance of addressing substantial differences in system implementation from one district to another and would like to provide support to those with fewer resources and less capacity (such as resource-poor districts in rural areas). In attempts to ensure consistent and fair classroom observations, some decentralized states have employed the following strategies to enhance local implementation of teacher evaluation systems, and target state support where it is needed most:

- Although standardized materials for trainers are not as practical for states with many local systems, explicit guidelines could be applicable. Florida provides training guidelines detailing what all local evaluator trainings should cover.

- State departments in partnership with other state educational organizations can facilitate knowledge and resource sharing among districts. Several Florida districts were recently convened by the Northeast Florida Educational Consortium to learn with and from one another. While districts accustomed to local decision-making might not welcome state involvement, some district leaders acknowledged the importance of gaining "perspective on how we do things and learn fresh ideas … maybe there is a better way."

- The best way to meet the needs of local agencies might entail gathering feedback from districts and educators about the key resources currently missing. As some states have already heard, a video library illustrating major instructional strategies (such as questioning and classroom assessment) could be critical for some districts and beneficial for many. A vetted list of consulting services who can support local implementation efforts could also be helpful according to feedback from the field.

Ensuring Observation Quality

Training observers does not guarantee they will be qualified to conduct accurate and valid observations. As states report, educators are realizing the practice of observing and judging teaching is as complex as teaching itself. Noticing all the important nuances during a lesson is not easy. Observers are challenged to correctly connect observations to particular dimensions of the observation framework. Furthermore, researchers are finding that observers' judgments of teaching quality are easily influenced by prior knowledge and beliefs about teachers and students, and evaluations could change with time (drift) as observers consciously or subconsciously adjust their expectations.

As of 2014, many SREB states are still weighing the costs and benefits of an observer assessment and certification program. Some researchers have argued that rigorous assessment and certification are critical for ensuring that observers meet at least a minimum level of competency. But some states contend that an annual test provides limited quality control at best; passing one or two tests each year does not necessarily mean observers will make accurate and valid ratings across teachers and time. As described in Table 4, there are currently more states that require observers to be certified than those that do not.

Whether they currently assess observers or not, states uniformly agree that assessment and certification is not a substitute for continuous observer training and calibration. In particular, many states agree that pre-scored videos, which observers can use to practice rating, are useful for assessment and simultaneous training.

Ideally, these videos have high visual and audio quality, match the observer's context (such as type of school and length of observation), and are replaced regularly to ensure assessment validity. Even if currently available videos are imperfect, states generally agree that they would still be useful to practice scribing and scoring. Videos could also be the basis for productive discussions among observers; by comparing where they agree and disagree, observers are engaging in a form of calibration that could improve their inter-rater reliability (see later section).

TABLE 4: Observer assessment and certification in SREB states

State	Are observers assessed by state?	Assessment frequency	Is observer certification required by state?	Certification criteria
Alabama	No, determined by districts	N/A	No, determined by districts	N/A
Arkansas	Yes, paper and video tests	Not decided	Yes	Not decided
Delaware	Yes, online assessment module	Every 5 years	Yes	Pass online assessment module
Florida	Not determined by state	Determined by districts	No, determined by districts	Determined by districts
Georgia	Yes, paper and video tests	Once	Yes	Pass tests and complete required annual updates
Kentucky	Yes, video test	Every year	Yes	Pass test
Louisiana	No	N/A	Yes	Complete training
Maryland	No, determined by districts	Determined by districts	No, determined by districts	No, determined by districts
Mississippi	No	N/A	No	N/A
North Carolina	No	N/A	No	N/A
Oklahoma	Yes, varies by model	Every 2 years	Yes	Pass test
South Carolina	Yes, video test	Not decided	Yes	Not decided
Tennessee	Yes, video test	Every year	Yes	Pass test
Texas	Yes, video test for districts using the state system	Every year for districts using the state system	Yes for districts using the state system	Complete training and pass test
Virginia	No, determined by districts	Determined by districts	No, determined by districts	Determined by districts
West Virginia	No	N/A	Yes, part of earning administrator license	Complete training

Sources: Review of state policy documents, and interviews with state education agencies in summer 2014

States are starting to use other calibration and monitoring strategies beyond initial observer certification to enhance the trustworthiness of classroom observations. Even states that have rigorous assessment and annual certification requirements for observers find they are "necessary but insufficient." Tennessee, for example, has started a process of peer calibration in which they bring together groups of principals to observe the same lesson and then discuss their observations and ratings.

Several states are also considering viable means of monitoring observation data and feedback from educators. This may be through systematic data collection or through selective audits. Data collection and analysis will help states understand the challenges or threats to observation accuracy, identify observers that might be less reliable and determine what trainings and supports are effective in improving observation quality.

Building a System and Culture of Educator Support

As states work to improve the accuracy of observations and quality of feedback for educators, it is important to keep the end goal in mind: better teaching for students. Many SREB states are beginning to see that supporting continuous improvement in teaching extends beyond the scope of initial observer training. In particular, building educators' capacity to give and receive feedback is critical to the success of classroom observations. In most cases, state agencies will need to collaborate with other educational agencies toward building a system and culture of educator support.

Going beyond initial observation training

Understanding the observation framework should be a continued focus of statewide, regional and local trainings, for example, in principal association and school team meetings. States are increasingly aware that evaluators likely need more time and engagement with the observation framework to understand it well. No matter how clear the rubric document is, educators need to "unpack" it for themselves. The initial training currently provided to observers generally does not go into sufficient depth, even if trainings include active-learning components.

Across various observation frameworks, observers now need to base evaluations on specific evidence of teaching practice from classroom observations and other sources instead of general opinion or intuition. Observers in many states are now expected to scribe, that is write down the details of what they see and hear (for example, "22 of 27 students worked on solving the problem"). This is a new practice for most.

Trainings in many states now include opportunities for trainees to practice scribing, critique scribe examples, practice matching observations to framework dimensions and practice scoring. But states are learning, in part from calibration results, that accurate and reliable scoring continues to be a challenge. Evaluators need more practice and targeted feedback on gathering and using evidence for evaluation.

> *"Observers need ongoing opportunities to practice gathering and using evidence for evaluation beyond their initial training."*

Observers could also improve their rating quality through ongoing calibration opportunities. Calibration could involve watching videos that have been pre-scored by expert observers and observers, for example, as part of the observer assessment system created by vendors such as Teachscape, NIET and Empirical Education. Calibration could also occur in a live context, for example, when two or more observers observe the same teacher and compare notes and ratings afterwards. The point is to give observers multiple and ongoing opportunities to reflect on their rating accuracy and the basis on which they evaluate teaching, and to increase awareness of potential systematic biases influencing their judgments.

Focusing on feedback for educators

Ultimately, accurate classroom observations that educators trust create a means for generating useful feedback that informs teaching improvements and teacher growth. Yet giving feedback tends to be least addressed in observer training. Many states agree that giving effective feedback is difficult to teach. Even when states have an observation system that includes teacher self-reflection and a protocol to facilitate dialogue, there are no guarantees that teachers will receive constructive feedback that could ultimately drive instructional improvement.

One challenge is the emotional dimension of giving and receiving feedback. Dialogues around observations can be uncomfortable and highly emotional for both parties; even if most of the comments are positive and suggestions are only constructive, many principals and teachers have never or rarely experienced these types of "courageous conversations." Training in many states aims to prepare educators for the feedback process by introducing a research-based model (such as cognitive coaching) and providing practical tips including do's and don'ts, sentence starters and suggested protocols for post-observation conferencing (see North Carolina for an example). But no amount of preparation can make the practice of professional critique easy.

Another challenge is that many principals lack the instructional capacity to provide teachers with useful feedback, especially for subject areas in which they are not trained. In order to be able to give constructive feedback to teachers, principals must learn new ideas about good teaching, develop more instructional knowledge across multiple content areas and become competent coaches and mentors.

> *"Principals must increase their capacity to provide useful, instructional feedback for teachers, especially for subject areas in which they are not trained."*

For many who traditionally play the role of administrative managers, principals who are now required to be instructional leaders at their schools are facing a massive change in their work. Many will need sustained professional development that extends beyond the current scope of observer training.

Teachers also need the knowledge and skills to improve teaching in response to evaluation feedback. Effective professional learning opportunities and support must be available so that teachers can learn new instructional strategies as needed. Teachers may also need more technical support for formative assessment practices so they can monitor the impact of their instructional changes on student learning, and make informed decisions about appropriate instructional moves (what to teach or reteach to whom and how) in a cycle of continuous improvement.

There is an accumulation of research on effective professional development to suggest that one-time training sessions have limited impact on improving teaching practice. Teachers need time and opportunities to integrate new knowledge into their classroom practices with support from other practitioners with expertise. Although teacher development is typically viewed as a separate domain from teacher evaluation, systems to evaluate teachers ultimately need to be linked toward a development system and culture of educator support.

Collaborating to support teacher growth

Providing support to observers beyond initial training and ensuring an effective feedback system for teachers are massive efforts that no state agency should be expected to accomplish alone. Many SREB states have taken notable strides toward building a system and culture of educator support in collaboration with regional and local district agencies and educators themselves. For example:

- Arkansas, North Carolina, West Virginia and Texas partner with regional centers/coops to provide expanded observer training and on-the-ground support. In close partnership with the state's Center for Professional Development, West Virginia requires all new administrators to participate in the Evaluation Leadership Institute program before they are certified. In Texas, one of the regional education centers is responsible for providing ongoing training to cover materials in more depth than is possible in one-time training events.

- Kentucky and Virginia work closely with principal associations planning training days together to make sure principals have sufficient training and information on teacher evaluation policies and processes.

- Georgia provides training for district-level staff who will receive district- or state-level credentialing to build greater local capacity. Individuals with district-level credentialing have a verified level of qualification to support the implementation of Georgia's teacher evaluation system in their district, while those with state-level credentialing can train and credential other districts throughout the state.

- Louisiana leads collaborations among district leaders around planning for assessment, curriculum, and teacher support systems. In addition, teacher leaders representing every district were convened by the state to create tools such as sample units and lessons in various subject areas, to be used by other teachers.

Monitoring Observations for Continuous Improvement

Reasons for monitoring

General studies such as this are limited in pinpointing priority issues and high-potential solutions for individual states. Ongoing monitoring of observation systems is important to ensure that feedback systems intended to drive teaching improvement have the desired positive impact.

Accessing data can be tricky. Gathering quantitative and qualitative data takes time — even if cases are selectively sampled for audit. Analyzing data takes substantial amounts of human and technical resources that may not be available to every state department. But some states are experiencing the benefits of having a monitoring mechanism that far outweigh the costs.

- Many states have been able to improve the design and implementation of teacher evaluation systems based on stakeholder input. Mississippi has simplified and clarified their observation rubric and streamlined conferencing requirements based on formal and informal feedback from educators.

Many other states are learning that educators struggle with the time observation systems require and are considering possible adjustments. Gathering information to inform decisions about observation frequency is essential for ensuring the sustainability and integrity of the system.

- Some states have used stakeholder input and analysis of pilot data to determine key implementation barriers and identify areas with the greatest need. Maryland has gathered feedback to assess stakeholder confidence in the new educator evaluation system, and commissioned a field study to examine implementation successes and challenges. Tennessee routinely analyzes observation ratings to identify "outlier" districts and schools that would benefit from working with coaches.

- A few states are also beginning to use data to inform stakeholders and increase their engagement. Some evidence suggests providing educators with data from their own school, district or state (such as observation ratings for one school compared to district average) motivates them to change their behavior. North Carolina has conducted some analyses with their state data and plan to use them in training. Similarly, Louisiana provides district-level reports so that each district can self-reflect about what they are doing well and how they can improve teacher evaluation implementation.

Recommendations for monitoring

1) **SREB states should continue using feedback to inform the initial development and refinement of new teacher evaluation systems.**

Stakeholder input should be gathered continuously, through formal and informal communication channels to ensure that state systems improve and evolve. Various data collection methods could be useful to this end, including surveys, focus groups and town hall meetings or webinars. Feedback from external partners (such as vendors and consultants) who are trusted as "critical friends" could also help states identify potential risks so they can be addressed or considered at key decision-making points.

2) **States should start or continue monitoring fidelity of implementation and observation quality.**

Even the best theory of evaluation is of little use if it is not implemented well. States should first ask the most essential questions: Are observations actually happening? Who are and are not completing them? The feedback from stakeholders will help states understand why observations are occurring or not.

In addition to monitoring completion rates, some states are starting to analyze observation quality. Many are examining the distribution of observation ratings to check for appropriate differentiation in scores and potential score inflation. Those who have calibration data are also examining observer accuracy to identify groups and individuals who may need additional training.

3) **System monitoring should examine biases that could undermine the validity and fairness of classroom observations.**

A Brookings Institution study recently found that teachers' observation scores were systematically related to the demographics of their students. After taking other factors into account, teachers of higher-performing students tend to receive higher observation scores than those who teach academically struggling students.

A Measures of Effective Teaching (MET) Project study found that observation ratings vary systematically by rater characteristics; administrators rated their own teachers higher than did administrators from other schools and peer teacher raters. Another MET Project study pointed out other potential observer effects such as halo and fatal-flaw effects. These refer to instances when one salient aspect of an observation or teacher influences all other aspects of the observation. Observers may also routinely give middle-of-the-scale scores even when not warranted, known as a central tendency effect.

4) **The intermediate and long-term impacts of classroom observations should be monitored.**

All states want to know if changes to classroom observations positively impact teaching and student learning. Already, many states are gathering self-report data about the changes in classrooms and schools as a result of new teacher evaluation systems. In addition, many states are interested in conducting more objective, systematic analysis of the quality of feedback teachers receive.

What kind of guidance and support are teachers receiving from instructional leaders? What follows from observations and summative evaluations? Understanding the intermediate effects of classroom observations is critical for monitoring the system's potential impact on instructional improvement, but studying such changes is arguably even more difficult than monitoring biases in observation scores.

5) **Monitoring at both state and local levels is likely to be most effective.**

The responsibility for monitoring efforts should not fall only on states, although states could have unique analytic capabilities. Specifically, those with large statewide datasets can examine important questions using powerful statistical methods, for example: What is the relationship between classroom observation ratings and student growth measures? Are there systematic differences in classroom observations based on student, teacher, or observer characteristics?

Local monitoring is equally important because districts and educators can make direct implementation adjustments and improvements. To this end, many states have developed or licensed a statewide data management system that facilitates local data monitoring. Evaluators in Arkansas are able to use BloomBoard to track the classroom observations they have completed and to review scores to uncover trends. States and other educational partners could examine statewide data trends to offer guidance and support that further enhance classroom observation implementation.

The following table contains a list of potential research questions for monitoring classroom observation implementation and outcomes.

TABLE 5: Critical questions for monitoring

Monitoring focus	Key questions
System design	How do educators and other educational stakeholders perceive system components and the system as a whole?
	What design changes would improve their perception of the system?
System implementation	To what extent are observations being conducted and completed on time?
	Are observation ratings differentiated to reflect the range seen in teaching quality?
	Are observation ratings inflated compared to other measures of teaching quality?
	Do observation ratings seem fair, with no systematic differences between teacher groups and observation contexts where differences should not be expected (across subject areas/courses/observers)?
	What are educators' experiences of the observation process?
System outcomes	What is the nature and quality of observation feedback?
	What changes to their practice do educators attribute to the observation system, positive and/or negative?
System impact	In what ways, and to what extent, is teaching practice improving as a result of better observations?
	In what ways, and to what extent, are student outcomes improving as a result of better observations?

Conclusions

Classroom observation is as significant a component as student growth in most state systems and has the potential to be a powerful mechanism for improving teaching and student learning. Initial efforts to frame new observation systems and train observers is proving to be challenging, but there are successes worth celebrating.

Still, the work is far from over. States are realizing that observation quality is something that cannot be addressed by annual assessments or a certification process alone. Establishing routine calibrations and ongoing support around feedback to improve teaching are goals that cannot be easily met by state agencies working alone. Collaboration among educational agencies and groups is critical for continuing and sustaining successful classroom observation and feedback systems for teachers.

In addition to the lessons learned from SREB states and guidance offered in this report, state decisions must still be made with information and insights from local implementation evaluation. Although the gathering and analysis of teacher evaluation data have costs and risks, continuous monitoring is an essential mechanism for tracking progress and identifying barriers of successful implementation of classroom observation policies.

How SREB can help

SREB's mission is to support state and local agencies in their efforts to improve educator effectiveness and ultimately improve learning experiences and outcomes for students. Future SREB efforts in the area of educator effectiveness include the following:

- A self-assessment tool that will stimulate and guide reflection and conversation among state stakeholders about their educator evaluation policies and practices. SREB continues to serve as a thought partner in any and all conversations with SREB state agency staff.

- SREB will routinely research resources to address the common needs of member states (for example, vendors that can support the development of video libraries). Vetted tools and ideas will be shared through our bi-weekly newsletter and other channels.

- SREB can help states gather honest feedback from the field about implementation successes and challenges (for example, conducting statewide focus groups). We will also conduct in-depth studies of successful implementation efforts to gain insight about how to improve educator effectiveness in and across SREB states.

- SREB will create ongoing opportunities for member states to collaborate. For example, SREB will continue to host region-wide convenings to address common issues and challenges (for example, how to monitor implementation and impact). SREB will also organize smaller work groups differentiated by states' needs (e.g., to discuss options for observer assessment, to work out ways of monitoring local systems in decentralized states).

References

Aspen Institute. (2011, March). *Building teacher evaluation systems: Learning from leading efforts.* Retrieved from http://www.aspeninstitute.org/publications/building-teacher-evaluation-systems-learning-leading-efforts

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, 17, 62-87.

Brown Center on Education Policy at Brookings. (2014, May). *Evaluating teachers with classroom observations: Lessons learned in four districts.* Retrieved from http://www.brookings.edu/~/media/research/files/reports/2014/05/13%20teacher%20evaluation/evaluating%20teachers%20with%20classroom%20observations.pdf

Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.

EducationCounsel. (2014, April). *The quality framework: A tool for building evaluation systems that improve instruction.* Retrieved from http://www.educationcounsel.com/docudepot/QualityFramework2014.pdf

Gitomer, D. H. & Bell, C. A. (2013). Evaluating teachers and teaching. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA Handbook of Testing and Assessment in Psychology (Volume 3: Testing and Assessment in School Psychology and Education)*, 415-444. Washington, DC: American Psychological Association.

Joe, J. N., McClellan, C.A., & Holtzman, S. L. (2014). Scoring design decisions: Reliability and the length and focus of classroom observations. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project*, 415-443. San Francisco, CA: Jossey-Bass.

National Council on Teacher Quality. (2011, October). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies.* Retrieved from http://www.nctq.org/dmsView/State_of_the_States_Teacher_Evaluation_and_Effectiveness_Policies_NCTQ_Report

Park, Y. S., Chen, J., & Holtzman, S. L. (2014). Evaluating efforts to minimize rater bias in scoring classroom observations. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project*, 383-414. San Francisco, CA: Jossey-Bass.

Southern Regional Education Board. (2015, January). *State implementation of college- and career-readiness standards: Evaluation of teachers and leaders.* Retrieved from http://publications.sreb.org/2015/Evaluation_SREB.pdf

Southern Regional Education Board. (2013, June). *Toward better teaching: A view of evaluation policies, practices and lessons in SREB states* (SREB's educator effectiveness series). Retrieved from publications.sreb.org/2013/13E08_Teacher_Eff.pdf

The Bill & Melinda Gates Foundation. (2014, June). *Building trust in observations: A blueprint for improving systems to support great teaching* (MET Project policy and practice brief). Retrieved from http://metproject.org/downloads/MET_Observation_Blueprint.pdf

The Bill & Melinda Gates Foundation. (2013, June). *Foundations of observation: Considerations for developing an observation system that helps districts achieve consistent and accurate scores* (MET Project policy and practice brief). Retrieved from http://www.metproject.org/downloads/MET-ETS_Foundations_of_Observation.pdf

The Bill & Melinda Gates Foundation. (2013, January). *The reliability of classroom observations by school personnel* (MET Project Research Paper). Retrieved from http://www.metproject.org/downloads/MET_Reliability_of_Classroom_Observations_Research_Paper.pdf

TNTP. (2013, November). *Fixing classroom observations: How common core will change the way we look at teaching* (TNTP Issue analysis report). Retrieved from http://tntp.org/assets/documents/TNTP_FixingClassroomObservations_2013.pdf

SREB

Educator
Effectiveness