

Oklahoma Teacher and Leader Evaluation Systems

Focus Group Findings

November 2014



TABLE OF CONTENTS

Executive Summary.....	3
Background.....	5
Methodology.....	7
Participants.....	8
Findings & Recommendations.....	9
Conclusion.....	17
Appendix A.....	18
Appendix B.....	20

This report was prepared by Tysza Gandha, research consultant, and Andy Baxter, vice president of Educator Effectiveness, Southern Regional Education Board.

The authors would also like to acknowledge the following individuals for data collection, analysis and editorial support: Danielle Douez, Matt Smith and Lorén Trull.

EXECUTIVE SUMMARY

In 2013, Oklahoma began implementing the Teacher and Leadership Effectiveness (TLE) system to assess educators' strengths and weaknesses using multiple measures. As leaders from many other states can attest, implementing a new-generation educator evaluation system like TLE is resource-intensive. State agencies must figure out a meaningful, workable and effective system in a relatively short time. Teacher evaluations also demand a lot of administrators and teachers who are juggling many other programs and changes (school accountability, learning standards, etc.), while continuing daily practices of leading and teaching. Nevertheless, federal and state education leaders continue to champion teacher evaluation reform based on a growing body of research showing that good teaching has an enduring impact on student success and can level the playing field for students from different backgrounds and circumstances.

In the fall of 2014, the Oklahoma State Department of Education (OSDE) invited the Southern Regional Education Board (SREB) to conduct focus groups with educators across the state. They wanted to know how TLE was working on the ground and gather honest feedback about how TLE could be improved. Focus groups were conducted in 10 cities between September 22 and October 3, 2014. A total of 131 educators participated, including 71 administrators and 60 teachers, representing 58 school districts. Although participants were nominated by district superintendents and did not represent a random sample of Oklahoma educators (newer teachers were not well represented), highly consistent messages received from both administrators and teachers across schools and teaching contexts lend credibility to the following major findings:

1. **The qualitative portion of TLE is a significant improvement over the old teacher evaluation system, although implementation is in its nascent stage.**

Some teachers still receive little to no feedback from administrators about their strengths and areas for improvement. Administrators and teachers alike need more time to fully unpack the teaching quality rubric and align professional development opportunities to growth areas.

2. **Many educators, especially teachers, do not *understand* the quantitative components of TLE and need guidance for developing meaningful and fair local measures.** Administrators who received Value Added Measures (VAM) training

lack a firm grasp of the measure and do not feel qualified to train teachers on it.

Educators are generally unclear about the difference between Other Academic Measures (OAM), Student Learning Outcomes (SLO) and Student Outcome Objectives (SOO), and need guidance on choosing or creating measures of student growth that are meaningful and fair across teachers.

3. **The vast majority of educators do not *believe* the quantitative components are valid, fair or helpful for improving their instruction. They want to see the quantitative portion either delayed or eliminated.** Educators expressed an immense amount of pushback against the quantitative components. Some of the resistance seemed to stem from misunderstanding, but most of it came from disagreement about the theory of action, practicality and helpfulness of quantitative measures. In other words, even those who understood it rejected it.
4. **Lack of buy-in for TLE is significantly rooted in a severe distrust of OSDE.** Many focus group participants described OSDE’s school and educator accountability programs as means of “tearing down public education” and “beating down on teachers,” instead of a system genuinely intended to improve teaching and learning. The distrust was deep and wide, including everything from OSDE’s handling of sensitive data, to organizational capacity, to the motive behind the TLE work.

As OSDE faces the impact of losing the state’s Elementary and Secondary Education Act Waiver and readies for a change in leadership, some educators are naturally wondering, “Is TLE going to go away?” Although focus group participants identified many problems with TLE at this time, nearly all also noted the improvement of the qualitative portion over the old “check-box system of evaluation.” They emphasized the importance of stability so they can really “dig in” and “get confident” about implementing a productive teacher feedback and evaluation system.

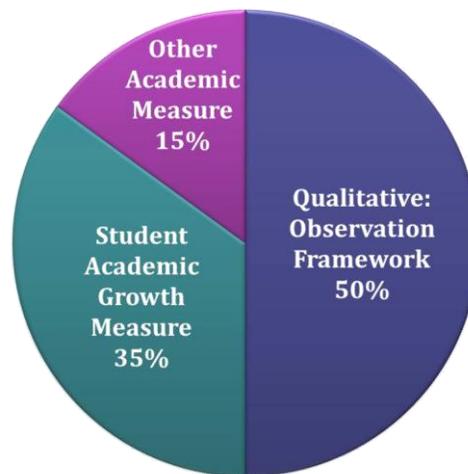
The findings and recommendations in this report are offered to guide necessary changes for TLE’s continuous improvement, while maintaining some system stability. Moving forward, OSDE should establish a mechanism to continuously solicit feedback from more Oklahoma educators (e.g., annual surveys) not only to validate the findings reported here, but also to gather additional insight into how TLE can become a beneficial and trusted system for improving teaching and learning in Oklahoma.

BACKGROUND

In 2013, Oklahoma began implementing TLE to assess educators' strengths and weaknesses using multiple measures. Teachers are now evaluated on qualitative measures of professional practice and quantitative measures of student outcomes (see Figure 1). The system's goal is to ensure that every student has an effective teacher every year, by providing educators with more detailed feedback on their practice and impact on students to guide their continuous learning and professional growth

Figure 1. TLE system for evaluating teachers with multiple measures

TLE and Multiple Measures of Effectiveness

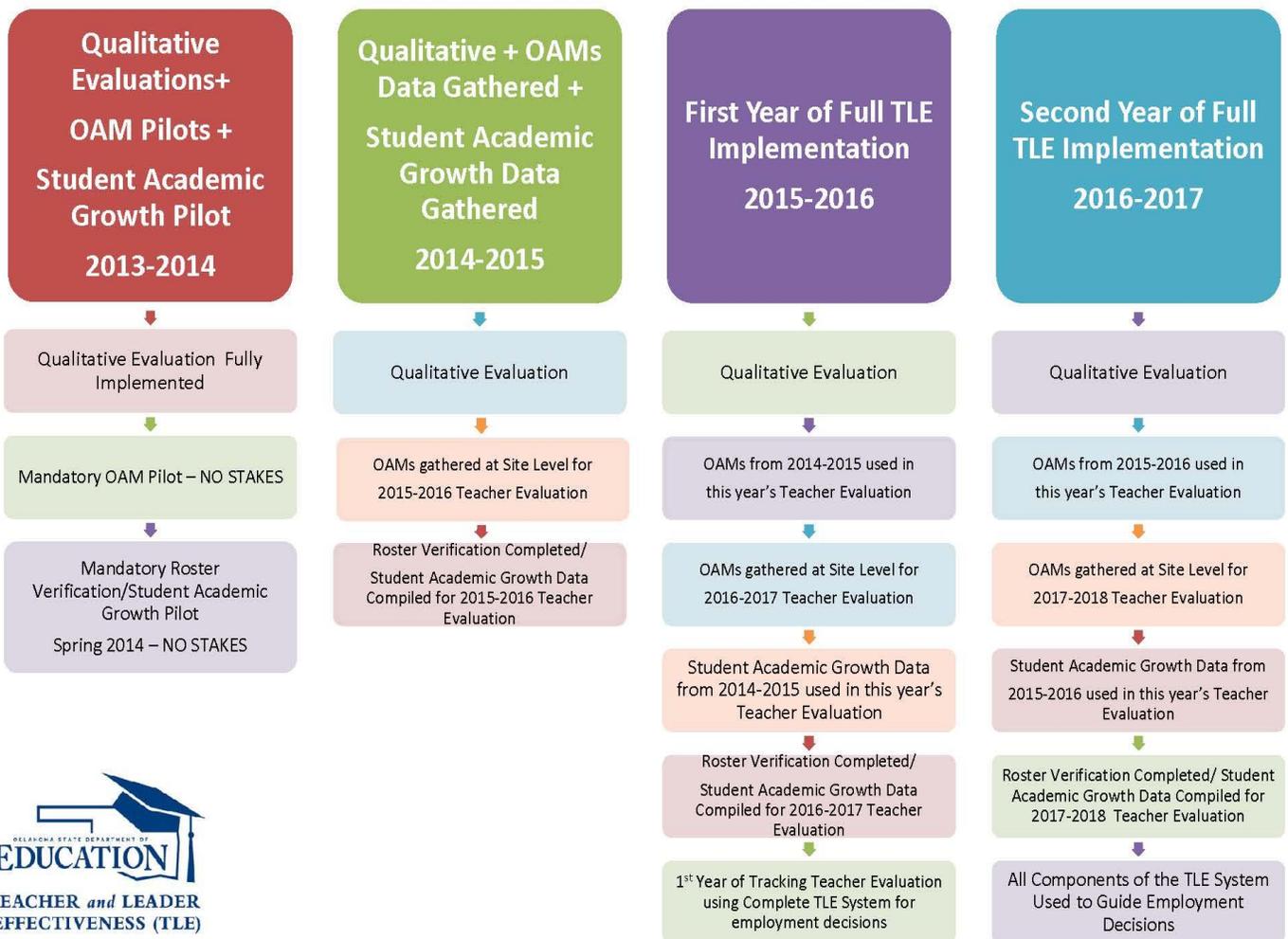


As many other states in the U.S. are experiencing, implementing a new-generation teacher evaluation system like TLE is a massive undertaking that requires substantial effort at all levels of the educational system. State agencies must design meaningful measures and workable processes. Training providers must plan and provide effective training for observers and those who will be observed. Educators must learn a new framework of teaching quality and participate in more in-depth observation, dialogue and reflective planning on an annual basis. Additionally, federal and state legislation intended to energize this work has laid out a tight implementation timeline that has created tremendous pressure for all involved (see Figure 2).

In the fall of 2014, OSDE invited SREB to conduct focus groups with educators across the state. Focus groups were intended to find what was and was not working well; to hear educators' perspectives, questions and concerns about the system; and to elicit their ideas about how to improve TLE. Since educators have the best knowledge about TLE implementation on the ground, their input is critical for informing the continuous development and refinement of the system. The findings and recommendations offered in this report are based on educators' comments in those focus group discussions.

Figure 2. TLE implementation timeline in accordance with Senate Bill 426

TLE Implementation Side-By-Side in Accordance with SB 426



METHODOLOGY

OSDE invited SREB to conduct this focus group project as an impartial third party and gather honest feedback from educators. SREB worked closely with OSDE to prioritize questions for focus group discussions and develop the focus group protocol. As shown in Appendix A, the focus group questions were broadly worded (“Which components or aspects of TLE are meaningful and beneficial/concerning for you?”) to allow participants the greatest degree of freedom in sharing their thoughts. At the same time, facilitators were prepared to probe into specific details to elicit more ideas (“Would you share an example of _____,” “What do you recommend about _____”).

OSDE provided logistical support in setting up 10 meeting locations around the state so all educators had access to at least one focus group site. In addition, OSDE facilitated the process of asking superintendents to nominate participants and encourage attendance. Superintendents were asked to nominate individuals from a variety of roles, although ultimately participants did not represent a random sample. Each superintendent could nominate up to three individuals from their district including one administrator (principal, assistant principal, central office personnel or superintendent) and one teacher (classroom teacher, counselor, nurse or library media and other specialist).

SREB independently conducted data collection, analysis and reporting to ensure the findings and recommendations would be unbiased. The focus group team was comprised of SREB staff and consultants with training and experience in qualitative methods and background knowledge in educator evaluation systems. Data were carefully collected and analyzed in a multi-step process to ensure findings would be grounded in the data and recommendations would be justified.

Every focus group was led by a facilitator with a notetaker present to scribe in real-time. In case of any discrepancies, sessions were also audiotaped so notes could be verified against the audio recording. After every focus group, the facilitator and notetaker wrote separate analytic summaries, then debriefed to corroborate what they had heard and discuss any differences in interpretation to improve validity. Facilitators’ and notetakers’ summaries were then used as the basis for cross-case analysis. Summaries were coded line-by-line and codes were tabulated to identify the major findings and supporting points reported here.

PARTICIPANTS

SREB conducted a total of 26 focus groups in 10 cities between September 22 and October 3, 2014. Each day, one or more focus groups were held in the morning for administrators and one or more were held in the afternoon for teachers. A total of 131 participants attended, including 71 administrators (principal, superintendent, dean, director, facilitator/coordinator) and 60 teachers (classroom teacher, counselor, media specialist, literacy coach). Participants represented elementary, middle/junior high and high school grades in fairly even numbers (47, 30, and 31 respectively; central office staff not counted). Participating teachers taught a wide variety of subject areas: tested (ELA, math, reading) and untested (e.g., art, music, speech, PE), regular education, special education and gifted. Across positions, grade levels and teaching subjects, participants' experiences and perspectives about TLE were far more similar than not.

Participants represented 58 school districts across Oklahoma (see Figure 3). As the list shows, they included large, small, urban and rural districts with a wide range of student population demographics. Participants described varying numbers of resources available in their districts in general, and for supporting TLE implementation. However, the findings reported here were highly consistent across district and school contexts. Many used the Tulsa framework for teacher observation and evaluation, while some used Marzano's. Again, educators' experiences and comments generally cut across observation frameworks so the findings are not specific to the Tulsa or Marzano model except where noted.

Figure 3. Districts represented by TLE focus group participants

Ada, Adair, Antlers, Bartlesville, Beggs, Bethel, Braggs, Burns Flat-Dill City, Byng, Catoosa, Checotah, Choctaw-Nicomma Park, Claremore, Coalgate, Cordell, Durant, Elk City, Enid, Fort Gibson, Henryetta, Hulbert, Indianahoma, Indianola, Jenks, Justice, Justus-Tiawah, Latta, Lomega, Madill, McAlester, McCord, Miami, Midwest City-Del City, Moore, Norman, Nowata, Oklahoma City, Owasso, Perry, Ponca City, Porter Consolidated, Porum, Pryor, Putnam City, Roff, Ryan, Sallisaw, Sand Springs, Sapulpa, Sayre, Stigler, Stillwater, Tecumseh, Tulsa, Union, Vanoss, Watonga, Woodward

It must be acknowledged that participants ultimately constituted a small percentage of Oklahoma educators. Notably, few teachers with less than five years of experience were represented, which limits the generalizability of findings to the newest members of Oklahoma’s teaching force. While participants may not be representative of all educators, the findings reported are based on consistent comments from teachers and administrators across a variety of classroom and school contexts, lending confidence to their significance. These findings are valid for informing OSDE’s decision making about TLE, especially when correlated to findings from other sources of information (e.g., town hall meetings).

FINDINGS & RECOMMENDATIONS

1. The qualitative component of TLE is a significant improvement over the old teacher evaluation system, although implementation is in its nascent stage.

- Participants using both the Tulsa and Marzano frameworks were familiar with the qualitative component of TLE. All discussion groups described many benefits of associated tools (rubric, online platform) and processes (self-reflection, classroom observation, dialogue, collection of artifacts) for improving teachers’ professional practice.
- For many educators, having a concrete and objective rubric has improved instructional conversations between teachers and administrators, but the extent of dialogue varied across sites partly depending on administrators’ time constraints and instructional knowledge. For some, the rubric continued to be used as a checklist with teachers receiving little to no feedback from administrators about strengths or areas for improvement. Administrators using the Tulsa model explained that they are not required to produce additional documentation when they rate teachers a “3”, which could lead them to unfairly give 3s when faced with unmanageable deadlines.
- Educators widely agreed that qualitative measures of professional practice will contribute to teacher growth, but most districts and schools have yet to align TLE with systemic professional development (in-service trainings, professional learning communities) and individualized professional growth planning. Administrators and teachers alike wish the

TLE timeline would slow down so they can fully unpack the teaching quality rubric (which seems more difficult for some teaching grades/subjects) and better align professional development opportunities to growth areas.

Recommendations for improving the qualitative component of TLE:

- 1.1. Study the time required to continue implementing TLE with fidelity. Consider some flexibility during the initial years (e.g., in number or duration of observations) to allow evaluators and teachers time to adequately learn the system. Similarly, consider some flexibility to accommodate differences in school contexts and situations. For example, give more time to evaluators in districts with the greatest numbers of teachers.
- 1.2. Provide and support ongoing training for evaluators beyond one-time summer workshops. Training should include guidance for managing time and resources to support the qualitative component of TLE that educators agree is important.
- 1.3. Develop a program to educate teachers about TLE directly; rely less on the train-the-trainer model. Recordings of high-quality training sessions and reader-friendly “quick guides” made easily accessible online are low-cost ways to ensure everyone throughout the state has access to critical, trustworthy and up-to-date information.
- 1.4. Create and support means for districts and educators to showcase successful TLE practices and learn from one another. For example, encourage and celebrate the development of regional networks and locally-initiated workgroups around TLE implementation. Spotlight districts or schools with an effective TLE practice in the TLE newsletter as a way of sharing good ideas and celebrating successes.

2. Many educators, especially teachers, do not *understand* the quantitative components of TLE and need guidance for developing meaningful and fair local measures.

- Educators' key criticism of VAM is that it would not measure individual student growth, which indicates they still lack understanding of this measure. Many administrators who received VAM training reported that they still lack a firm grasp of the measure and are not at all qualified to train teachers on VAM. Teachers corroborated administrators' reports.
- Furthermore, some administrators who have a better understanding of VAM as a calculation of individual student growth adjusted for background variables lack trust in any score based on Oklahoma's standardized tests. Participants described past experiences with invalid standardized tests (writing assessments that had to be rescored, unmodified tests for Special Education students for which they were not prepared) to explain their mistrust.
- The well-intended plan to slowly introduce VAM to educators in the spring of 2014, perhaps before the system was ready, might have created more confusion and negative perceptions. Educators who received VAM samples reported they had no support in interpreting the numbers and some found data that were incorrect (negative scores).
- Educators are more interested in the concept of using locally-selected measures of student growth/achievement, such as data from benchmark assessments they are already using at their school. However, all discussion groups questioned how comparable different measures of OAM, SLO and SOOs would be.
- Participants spoke very openly about the likelihood of OAMs, SLOs and SOOs being "manipulated" when used for high-stakes teacher appraisal because of a lack of guidance in developing meaningful and fair local measures. Speaking from their experiences with OAMs so far, many groups reported they had intentionally chosen measures and scoring rubrics that would "look the best." They also shared stories they heard of other educators who used "watered down" pre- and post-tests such as "how many continents or occupations students know." While participants did not have deep knowledge about SLOs

and SOOs, they expressed concern that these measures will be subject to similar manipulations.

Recommendations for improving the quantitative components of TLE:

- 2.1. Whatever role VAM may play in TLE (whether it directly or indirectly contributes to appraisal ratings), more effective training is needed to educate administrators and teachers on this complex measure. Successful training will need to clearly explain how VAM scores account for “where each student started,” how student background characteristics are factored in the calculation (and why some characteristics are not), and the mechanisms ensuring that scores will not be unfairly affected by situations beyond teachers’ control (students who were largely absent, significant events affecting some students or the whole school community). Administrators who have tried explaining VAM to teachers noted it took ongoing conversations to go over the complex calculations and address teachers’ natural concerns. High-quality resources (illustrative examples and interactive tools) to support conversations and learning about VAM could be worthwhile investments. This will eliminate some confusion, although much resistance will likely remain.

- 2.2. Provide clearer guidance and expert assistance to support the development of meaningful and equitable OAMs, SLOs and SOOs. Possibilities should include nationally-benchmarked assessments, high-quality district common assessments and validated student surveys. Give examples of appropriate and inappropriate measures in various school contexts. For example, explain why the creation of a skateboard club or increased participation in prom committee (both examples mentioned in discussions) might be valid measures for one school but not another. Consider providing a list of pre-approved measures by grade and subject, and setting some guidelines for combining state-approved and district-chosen measures.

- 2.3. Establish a quality-assurance mechanism to monitor and audit what student growth measures are being used across the state. Educators are more likely to buy into a system involving locally-selected quantitative measures if there is a mechanism to ensure the system is fair.

2.4. With any test-based measures, consider using multi-year data or multiple data points to enhance data validity and trustworthiness. Data collected across the year are also perceived to be more instructionally useful by providing teachers with better feedback about student learning and on what is and is not working in their instruction.

3. The vast majority of educators do not *believe* the quantitative components are valid, fair or helpful for their instruction. They want to see the quantitative components delayed or eliminated.

- Educators repeatedly asserted that they are not against accountability, but argued that current quantitative measures used in TLE do not accurately measure what they do or do not accomplish. All discussion groups expressed doubt about the accuracy of measures based on Oklahoma’s standardized tests. In addition, most had developed OAMs to comply with requirements but do not find the measures meaningful for improving their practice.
- Since the current measures hold no inherent value to educators and given the perceived high stakes (potential loss of livelihood, less-than-perfect appraisal that could influence future employment options), many educators expressed concern that the system as it currently stands will be gamed by those who choose measures to “get the best grade” instead of aspiring to achieve higher standards.
- Furthermore, some warned that high-stakes quantitative measures could detract from the qualitative TLE component that has potentially great value for guiding educators’ continuous growth. If educators are concerned about the validity of quantitative measures, inflating qualitative ratings could become a widespread strategy to protect teachers. This would ultimately undermine TLE’s goal of giving teachers accurate feedback about their professional practice to drive instructional improvement.
- Many teachers across subject areas argued that evaluating teachers of tested subjects on VAM scores and evaluating teachers of non-tested subjects on “whatever they wanted to set for their SLO/SOO” is unfair, divisive and would drive many teachers away from

critical tested subjects. The concern about fairness pertained to both teachers and students and was often educators' main argument for delaying the TLE quantitative portion until fair measures could be developed — or eliminating it altogether if such measures are not possible.

Recommendations for preventing system manipulation and ensuring TLE's positive impact:

- 3.1. Oklahoma legislators should consider amending state statutes stipulating the timeline for TLE implementation to allow further development of valid and fair quantitative measures. Accommodate a pilot or learning period to identify and address issues that could threaten system validity, and make key refinements to the system to ultimately secure greater buy-in from educators.
- 3.2. While quantitative measures are being worked out, focus expectations and support on implementing the qualitative component of TLE with fidelity; allow time for schools and districts to solidify their understanding of the teaching quality rubric and fully implement related processes for guiding, recognizing and celebrating continuous improvement in instruction.
- 3.3. Educators should continue being evaluated for student growth within the qualitative frameworks of effective teaching. Both the Tulsa and Marzano frameworks already include a dimension about teachers' use of data to monitor and improve instruction. As part of their professional practice, teachers should be using student data for self-assessment, goal setting and professional growth planning. Training and professional development can strengthen this aspect of professional practice across the state. Random audits can be conducted to ensure teachers are continuing to improve to best serve students in this way.
- 3.4. Even while VAM scores are not being used for teacher appraisals, they could continue to be calculated and used to inform school-, district- or state-level decisions. For example, to focus professional development resources where most needed.

4. Lack of buy-in for TLE is significantly rooted in a severe distrust of OSDE.

- In most organizational hierarchies, people at one level are a bit wary of people at the next. Teachers may be frustrated by their principals who are frustrated by their superintendents who are frustrated, in turn, with the state department. However, educators in the focus groups expressed a far more intense distrust of OSDE than would seem normal. Participants cited numerous examples, from the TLE working groups to setting cut-scores on state exams, where they felt ignored by OSDE. Participants also pointed to a “revolving door” of staff at OSDE and last-minute changes in policies and deadlines as evidence for their distrust in OSDE’s ability to implement TLE well.
- Although TLE is separate from the A-F system of grading districts and schools, they were frequently discussed in connection to one another, suggesting that they are intertwined in educators’ experience and perception of educational accountability. Trailing the A-F system, which is widely perceived as “a way to call us failing schools,” TLE is seen as another means of “beating down on teachers.” Despite OSDE’s balanced description of the system in official documents, educators experience TLE as a system primarily designed to ask teachers to “prove their worth,” with not enough focus on acknowledging and celebrating educators’ efforts and achievements.
- Nearly all groups reported low morale among educators: new teachers are leaving the profession quickly and career teachers are considering changing careers or retiring early. Many administrators described increased hiring difficulties, not only because Oklahoma pays teachers less than neighboring states, but also due to poor public perception of Oklahoma education. Whether or not these were intended effects, such problems were repeatedly attributed to the A-F school accountability system and TLE.
- In many groups, participants expressed disappointment with not having “real input” in TLE design thus far; opportunities that were available (TLE workgroups) seemed like superficial attempts for OSDE to “look good” and educators who had participated did not feel respected or heard. In fact, many are pessimistic that their focus group contribution will result in any change (see Appendix B). Nevertheless, many committed and passionate educators continue to express interest in wanting to help shape and guide TLE because

they want an evaluation and accountability system that will benefit them, their career and students.

Recommendations for increasing trust and buy-in for TLE:

- 4.1. Signal to Oklahoma educators that their concerns are heard and valued by allowing greater flexibility for TLE's quantitative component and more time for the state department and educators to work out meaningful and fair student growth measures.
- 4.2. Continue to seek input from educators in designing, implementation monitoring and improving TLE. Provide a variety of ways for educators to engage in the decision-making process (participate in workgroups and serve as reviewers) and voice responses and suggestions (town halls, focus groups and surveys).
- 4.3. Continue to improve communication and transparency about TLE and decision-making behind it; send the message "we hear you" when educator feedback has, in fact, led to some action or decision. Establish key sources of information that are easily accessible to all education stakeholders such as an up-to-date webpage, and proactively communicate important messages. Explain key components and decisions clearly and address educators' concerns head-on to avoid fueling misunderstandings and misinformation.

CONCLUSION

Although critiques of TLE have been reported plainly in this report, nothing is intended to be a summative statement on the potential of TLE to help teachers and students in Oklahoma. Findings must be understood in their proper context, which is that TLE is in a relatively early phase of implementation. SREB, along with many Oklahoma educators, remains hopeful that the system will improve over time. This study was conducted precisely to elicit important feedback from educators who have a clear view of TLE implementation on the ground and collect their insights to make TLE better in the future. OSDE's willingness to invite SREB to conduct the focus groups is a good-faith sign that it wants to hear the concerns of educators.

The legislative and regulatory processes involved in carrying out the recommendations offered in this report are not to be underestimated, nor should the substantial number of resources necessary to design and implement a more meaningful, valid and effective educator evaluation system. There are simply no easy solutions for addressing the key concerns of Oklahoma educators whose buy-in and trust in the TLE system are critical for its success.

APPENDIX A: Focus group protocol

Facilitators: Lead with the questions in bold. Use probes to follow up as a topic comes up/as needed.

We know that Oklahoma’s TLE system has several components — the qualitative, the VAM, SLO, SOO, and OAM — and we welcome your thoughts about all of them.

a. To start, which components or aspects of TLE are meaningful and beneficial to you? Anyone feel free to start and others can add.

- Would you share an example? What does that look like at your district/school?
- How is it different from what happened before you had TLE?
- Would you share an example of a change you made or experienced as a result?

b. Now, which components or aspects of TLE are concerning to you, or maybe just where you need more clarification and support?

- What advice do you have for the state department to address the problem of ___ (e.g., time demand)?
- What do you recommend the state department do to clarify ___ (e.g., quant component)?
- Specifically, what kind of information or support would be most helpful to you? (e.g., training format/content)
- Just so I am clear, would you explain ___ (e.g., consequences) as you understand it? Is that everyone’s understanding or does anybody here have a different understanding?

c. TLE involves big changes, especially when it comes to the quantitative portion that is unfamiliar territory for most educators. What do you think would make the transition to this new system easier for schools?

- One idea to simplify the quantitative portion is to give districts more flexibility with how they incorporate student growth measures. What do you think about that possibility?
- One idea is to gradually increase the percentage of the quantitative portion. What do you think about that?

d. TLE measures teaching quality using multiple measures which should be more valid, at least in theory. How do you think the multiple-measures evaluation system is better or not better than what you had before?

e. TLE is also designed to give educators useful information about their strengths and areas for improvement. From your experience so far, how is TLE helping or hurting schools?

- In your district now, how is professional development informed by or aligned with TLE?
- What do you need to improve instruction and student learning in your districts/schools?

APPENDIX B: Exit survey summary

- According to participant exit surveys, **98 percent enjoyed participating in the focus groups.** Over half appreciated hearing that other educators had similar perceptions and experiences (*“I enjoyed hearing others from across the state voice the same concerns and together try to come up with a solution”*), while some also appreciated hearing different perspectives (*“It was interesting to hear the perspectives of colleagues in other roles”*) and learning from one another (*“I heard some good ideas from others about how to improve TLE”*).
- **Nearly all participants (98 percent) agreed that their voices were heard and that the focus group was worth their time, despite 23 percent expressing concern that their input will not result in any change.** Some are cautiously hopeful (*“I hope the state board and legislators will listen! Are we doing this only so they can say “We heard from teachers and administrators?”*), while others are less optimistic (*“I don't think this will have any influence. Everything about education has become very political”*).