



# Accountability in the evaluation of teacher effectiveness: views of teachers and administrators

Xumei Fan<sup>1</sup> 

Received: 2 February 2023 / Accepted: 3 October 2023 / Published online: 12 October 2023  
© The Author(s), under exclusive licence to Springer Nature B.V. 2023, corrected publication 2024

## Abstract

Student learning outcomes have been used to measure a teacher's accountability for student learning in the evaluation of teacher effectiveness. This study employed a survey research design and investigated the views of teachers and administrators regarding using student learning objectives (SLOs) and classroom observations in the evaluation of teacher effectiveness in one southeastern state in the USA. The results revealed that administrators had significantly more positive views than teachers regarding the impact of SLOs and the classroom observations. Educators had less positive views of the impact of SLOs and the classroom observations after the full implementation of the evaluation system. Educators reported various concerns regarding the issues of using students' assessment results, assessment methods, subjectivity in goal setting, and lack of feedback. The findings can inform policy making, development and implementation of teacher evaluation system, and teachers' professional learning and development.

**Keywords** Teacher effectiveness · Teacher evaluation · Student learning objectives · Classroom observations

## 1 Introduction

A priority in education is to improve student learning, and teachers play an important role and have significant impact on student learning (Hattie, 2009; Odden et al., 2004). To be qualified for the teaching profession, teachers generally need to obtain certain certification and licensure, receive an academic degree, have educational experience, and have sufficient knowledge about the subject, students, and pedagogy. Studies found an association between student learning outcomes and teacher

---

✉ Xumei Fan  
fan9@unc.edu

<sup>1</sup> School of Education, University of North Carolina-Chapel Hill, 100 E Cameron Ave, Chapel Hill, NC 27514, USA

characteristics including having a teaching certificate or license (Darling-Hammond & Young, 2002), years of teaching experience (Goldhaber & Brewer, 1997), and academic degrees obtained (Rowan et al., 1996).

However, obtaining these teacher characteristics and/or qualifications does not guarantee effective teaching. Teacher effectiveness focuses on specific teaching practices that involve the interactions between teachers and students in classrooms, teachers' course design and lesson planning, teachers' management of classroom learning environment, classroom assessment, and teachers' other professional activities (Danielson, 2013; Goe et al., 2008). Goe et al. (2008) summarized five aspects of effective teaching that emphasize on having high expectations for students, contribution to social outcomes, monitoring and evaluating student learning, encouraging diversity and civic-mindedness, and collaboration with other stakeholders to ensure student success. In addition to these teaching practices, teacher effectiveness is often linked to student learning outcomes.

The evaluation of teacher effectiveness generally serves two major purposes. One purpose of teacher evaluation is to help make decisions on teacher promotion, compensation, and employment, and the other purpose is to use the evaluation as a mode of professional development to improve teaching practices (Donaldson & Papay, 2015; Hanushek, 2009; Papay, 2012). A system of evaluating teacher effectiveness should be developed and implemented in accordance with the purposes of teacher evaluation. Marzano (2012) indicated that "measuring teacher effectiveness and developing teachers are different purposes with different implications. An evaluation system designed primarily for measurement will look quite different from a system designed primarily for development" (p. 15). The evaluation of teacher effectiveness has evolved overtime. Researchers in China indicated that the main purpose of teacher evaluation had shifted from rewarding or punishing teachers based on teachers' evaluation results to improving teaching practices through evaluation as a mode of professional development (Liu & Zhao, 2013). Researchers in Belgium found that the teacher evaluation system had both formative and summative purposes, and school leadership is key to effective teacher evaluation and professional learning (Delvaux et al., 2013; Tuytens & Devos, 2011).

Commonly used methods in the evaluation of teacher effectiveness include classroom observations, teacher self-evaluation, student evaluation of teachers, and student learning outcomes. Among these evaluation methods, classroom observation is broadly used as a component of teacher evaluation. Researchers (e.g., Campbell et al., 2010; Goe et al., 2008) emphasized that effective teaching should be evaluated through the classroom experiences that teachers create. Classroom observation rubrics are guided by different frameworks, among which the Framework for Teaching (FFT) developed by Danielson (1996) has been widely used in the USA. The FFT Evaluation Instrument was released in 2013 and has been widely used in teacher evaluation, school coaching and mentoring, and teacher professional development (Danielson, 2013). The FFT consists of 22 components of effective instruction that are clustered in four major domains: planning and preparation, classroom environment, instruction, and professional responsibilities. Studies found a small to moderate positive association between teachers' FFT scores and student learning, with some variation by grade level and subject matter (Gallagher, 2004; Milanowski,

2004). Students taught by a teacher in the top quartile scored 0.10 standard deviations higher in math and 0.13 standard deviations higher in reading than students taught by a teacher in the bottom quartile (Kane et al., 2013). However, a study conducted in Canada revealed that the school administrators believed that teacher evaluation based on classroom observations did not result in substantial improvement of teaching (Maharaj, 2014).

Another comparatively new component in the teacher evaluation system is using students' learning outcomes, which is used to measure a teacher's accountability for student learning. Romzek and Dubnick (1987) described accountability as managing expectations, which plays a great role in the process of public administration. In the evaluation of teacher effectiveness, teachers are expected to be accountable for student learning outcomes; thus, student learning outcomes are used as a component of teacher evaluation. Accountability has a great impact on policy making and the development of teacher evaluation systems, and student learning outcomes are linked to the evaluation of school effectiveness and teacher effectiveness in some countries. In the UK, schools are required to publicize students' test score information as they were held accountable for student learning (Machin & Vignoles, 2006). In the USA, student outcome measures are used as components of teacher evaluation systems in response to the federal program Race to the Top (RTT) which emphasized school and teacher accountability for student learning (Lachlan-Haché et al., 2012).

Students' learning outcomes are used in systems of evaluating teacher effectiveness. The value-added models (VAMs) that examine students' academic growth based on standardized tests are used to evaluate a teacher's contribution to student learning. Some studies found that using VAMs has advantages of being objective and cost-efficient (Goe et al., 2008) and can accurately predict students' long-term outcomes (Chetty et al., 2014a, 2014b; Sanders, 2000). However, other studies revealed many limitations of using VAMs. For example, Morganstein and Wasserstein (2014) indicated that it is problematic to use VAMs to analyze students' standardized test scores while they are not randomly assigned to classrooms, which might further cause unintended consequences. Morgan et al. (2014), Papay (2011), and Lockwood et al. (2007) found that teachers' evaluation ratings based on VAMs vary substantially depending on the types of standardized tests used and the groups of students involved. Regarding whether VAMs should be used in the teacher evaluation system, the American Educational Research Association (AERA) suggested that states and districts should acknowledge the considerable risks of misclassification and misinterpretation based on VAMs (AERA, 2015). Rockoff and Speroni (2010) suggested that other information should be used to gain more accurate and stable teacher evaluation results.

As a measure of student growth, the student learning objectives (SLOs) have become popular in the evaluation of teacher effectiveness in the USA. SLOs are defined as a set of goals that measure teachers' progress in achieving student growth targets that focus on students' expected learning at the end of the instructional period (Lachlan-Haché et al., 2012). In developing SLOs, teachers often consider six major components: student groups to be included, time span, curriculum standards, growth targets, instructional strategies, and assessment methods.

The process of using SLOs to measure student growth consists of (1) developing SLOs, usually constructed by an individual teacher or a team of teachers; (2) submitting SLOs for the approval from trained evaluators; (3) checking in through midcourse conversations between teachers and evaluators; (4) reviewing SLO attainment and scoring by both teachers and evaluators to determine if the student growth targets are achieved; and (5) completing the summative rating of the teachers and reflecting on the lessons learned from the process (Lachlan-Haché et al., 2012). In teacher evaluation, teachers are evaluated and given an overall score based on how well their students have achieved the learning objectives.

Various studies found mixed views of SLOs. One major advantage of using SLOs in the evaluation of teacher effectiveness lies in its adaptability to all subject areas and grade levels. Although teachers understood the importance of being held accountable, they seemed to be not very familiar with the evaluation score computing details (Moran, 2017). SLOs provided teachers with opportunities to use data, and they were more actively engaged in the evaluation after the SLOs implementation (Donaldson, 2012). Fan (2022) found that early career teachers had more positive views of SLOs, less knowledge about SLOs, and more support needed in using SLOs. A 5-year study showed that students at the schools where teachers used SLOs had higher growth rates in reading and math than the students at the schools where teachers did not use SLOs (Slotnik et al., 2013). However, studies showed that the majority of teachers did not consider the evaluation feedback as effective in improving their instruction, though they did value the specific, frequent, evidence-based feedback (Liu et al., 2019). Marshall et al. (2016) indicated the challenges of measuring teacher effectiveness due to different content areas, grade levels, and groups of students. In addition, the validity, reliability, and accuracy of teachers' SLO scores due to various factors including the quality of the assessment and the quality of evaluators were a big concern for the educators (Crouse et al., 2016). Furthermore, researchers found that holistic principal judgments of teacher effectiveness were much more strongly influenced by classroom observations of teacher practices rather than the evidence of student growth collected (Briggs & Dadey, 2017).

The implementation of the teacher evaluation system is a crucial part and has been a focus of many studies (e.g., Derrington (2014). Tuytens and Devos (2014) investigated school leadership actions and teacher evaluation development characteristics in the implementation of teacher evaluation policy in Belgium. They found that there were differences in leadership actions and teacher evaluation development characteristics between schools where teachers perceived the policy positively and the schools where teachers perceived the policy negatively. Cosner et al. (2015) suggested that teacher evaluation innovations had the benefits of providing opportunities for the improvement of instructional supervision and teacher quality. However, they also caused challenges including work and time demands and cognitive challenges for the principals in the implementation of the system. Similarly, Derrington (2014) investigated the views of K-12 superintendents and principals regarding the impact of a 2-year implementation of a teacher evaluation system, and they found that the implementation of the evaluation system had an impact on instructional

leadership, time and training tensions, supportive superintendent strategies, and unintended consequences.

In the implementation of teacher evaluation systems, teachers and administrators play different roles. District and school administrators often work as evaluators and make a judgment of teacher effectiveness either through classroom observations or using student learning outcomes. Huber and Skedsmo (2016b) indicated that role and authority have a crucial influence on defining accountability relationships. Administrators have authority over teachers, and they are involved in decision making based on teacher evaluation results. In the evaluation of teacher effectiveness, teachers and administrators hold different views due to their different roles. Bradford and Braaten (2018) described an incongruity between teachers' vision of high-quality teaching and administrators' vision of great teaching, and school leaders and teachers had different interpretations of scores. Reddy et al. (2018) found that school administrators reported more favorable experiences than teachers regarding teacher evaluation. Jones et al. (2022) indicated that principals hold multiple goals while evaluating teacher effectiveness through observations, and they may inflate teachers' evaluation scores to achieve multiple goals. In addition, principals had concerns regarding the evaluation systems' negative impact on morale, their lack of autonomy in decision making on evaluating teachers and staffing, and their perceived lack of value as professionals (Pauffer, 2018). Similarly, Slotnik et al. (2014) found that about half of the teachers and principals reported needing support to have access to data and analyze student data.

Teacher evaluation is complicated, and it involves multiple disciplines of education, psychology, sociology, and economics, and the consequences of teacher evaluation should be considered in the implementation (Donaldson, 2021). Due to the different roles that teachers and administrators play in the evaluation of teacher effectiveness, it is important to understand the perceptions of teachers and administrators and to identify the elements that contribute to the successes and failure of the implementation of a system or policy. Tuytens and Devos (2009) investigated Flemish teachers' perception of a new teacher evaluation policy and found that teachers were fairly positive toward the new policy reform, but they had questions about the policy's implementation.

Implementation of teacher evaluation system has been challenging, and teachers do not hold a positive view of teacher evaluation systems in some countries. The implementation of a new teacher evaluation system was described as a "hot potato" in Italy, which raised intense conversations among teacher unions, policymakers, and academics (Barzanò & Grimaldi, 2013). Flores (2012) explored teachers' perceptions of the implementation process of a teacher appraisal policy in Portugal and found that teachers considered the lack of recognition of the appraisers and bureaucratic and summative dimension as the most critical issues, and they felt uncertain and skeptical about the implementation of the policy. Similarly, teachers in Australia are skeptical and concerned about using the Australian Professional Standards for Teachers (APST) in the evaluation of their teaching effectiveness (Clinton et al., 2015; Clinton et al., 2016; Clinton & Dawson, 2018).

SLOs are a comparatively new method in the evaluation of teacher effectiveness, and teachers and administrators have comparatively less experience with

implementing SLOs. Previous research found that it is challenging to accurately measure teacher effectiveness using student learning outcomes due to different content areas, grade levels, and groups of students, which has been a big concern for the educators (Crouse et al., 2016; Marshall et al., 2016). Therefore, it is extremely important and much needed to explore how teachers and administrators view the use of SLOs in teacher evaluation. This study was intended to explore the views of teachers and administrators in terms of using classroom observations and student learning outcomes (SLOs) in teacher evaluation, their knowledge about SLOs, and support needed to implement SLOs. In addition, this study was also aimed at examining the impact of the implementation of the teacher evaluation system on the views of teachers and administrators. This study was expected to address the following research questions:

- How do teachers and administrators differ in their views of using SLOs and the classroom observations in the evaluation of teacher effectiveness?
- How do teachers and administrators differ in their knowledge about SLOs and support needed to implement SLOs?
- How does the implementation of the teacher evaluation system shape the views of teachers and administrators?

## 2 Theoretical framework

This study is guided by the accountability theory (Lerner & Tetlock, 1999) and the model for effective professional development (Guskey, 2002a, 2002b, 2016). Accountability suggests that a person has an obligation or is expected to explain and justify their beliefs, feelings, and actions to others (Bovens, 2010; Lerner & Tetlock, 1999). Lerner and Tetlock (1999) further explained that the expectation of evaluation is a person's belief that their performance will be assessed by others based on normative ground rules and with some consequences. In the context of teacher evaluation, teachers are expected to be accountable for student learning, and the evaluation of teacher effectiveness includes a component of student learning outcomes.

Teachers are evaluated for two major purposes. One purpose is summative with an attempt to identify effective teachers and ineffective teachers for making high-stakes decisions on recruitment, retention, promotion, and compensation (Hanushek, 2009). The other purpose is formative with a goal to provide teachers with professional development (Donaldson & Papay, 2015). Darling-Hammond (2013) described the teacher evaluation system as a “teaching and learning system” that supports improvement for teachers throughout their career. This study of teacher evaluation is informed by models of professional development (Guskey, 2002a, 2002b, 2016), which emphasize collecting data about participants' reactions to the professional learning experience; participants' learning of new knowledge, skills, and attitudes or dispositions; organizational support

and change; participants' use of new knowledge and skills; and student learning outcomes.

### 3 Methods

A survey research design was employed to examine the views of teachers and administrators regarding the evaluation of teacher effectiveness in one southeastern state in the USA. Two surveys were conducted to understand educators' views of the impact of SLOs and classroom observations, their knowledge about SLOs, and support needed to implement SLOs before the implementation (BI) and after the implementation (AI) of the teacher evaluation system in the state. In the survey after the implementation, I added one open-ended question attempting to gain some information about educators' experience with using SLOs in the implementation as well as their views of the evaluation of teacher effectiveness.

#### 3.1 Study context

This study involved one southeast state in the USA. The state has been dedicated to the development, implementation, and improvement of the teacher evaluation system since the establishment of its first evaluation system in the 1990s. Classroom observation is the primary evaluation method for classroom-based teachers, and SLOs data are collected as an artifact that supports the ratings of teachers. The classroom observation includes four domains (instruction, planning, environment, and professionalism) with a total of 22 indicators on a 4-point scale (1-Unsatisfactory; 2-Needs Improvement; 3-Proficient; 4-Exemplary). The SLOs focus on measuring teachers' contribution to student learning, and key components include setting learning targets; assessing and analyzing student growth; planning, implementing, and adjusting instructions; and ensuring student growth. A holistic rubric is used to evaluate teacher effectiveness, and there are four performance levels ranging from 1 (Unsatisfactory) to 4 (Exemplary). For example, if a teacher sets up rigorous goals for students, uses appropriate assessments to monitor student progress, strategically revises instruction, and between 90 and 100% of his/her students meet their growth targets, the teacher obtains 4 points (Exemplary). If a teacher inconsistently uses assessments, fails to monitor progress or adjust instruction based on progress monitoring data, and 0–50% of students meet their growth targets, this teacher obtains 1 point (Unsatisfactory).

Teachers' overall evaluation score is based on a 4-point composite score scale. A composite score of 1.24 points or below indicates a performance level of Unsatisfactory, a composite score ranging between 1.25 and 2.25 points indicates a performance level of Needs Improvement, a composite score ranging between 2.26 and 3.75 points indicates a performance level of Proficient, and a composite score of 3.76 or above indicates a performance level of Exemplary. The final evaluation results have two categories: Not Met (Unsatisfactory or Needs Improvement) and

Met (Proficient or Exemplary). All districts were required to implement the system in 2018–2019 school year, and school districts report evaluation data to the state board of education annually.

### 3.2 Participants

The first survey was conducted before the full implementation of the teacher evaluation system in the state, and a convenience sampling method was used. Thirteen school districts were recruited through the Teacher Advancement Program (TAP) and a Partnership Program in the state department of education. The TAP is a performance-based compensation system that encourages schools to recruit, evaluate, and compensate teachers based on their performance. The Partnership Program was a teacher professional learning initiative. These school districts have various levels of poverty, located in rural, urban, and suburban areas, and are representative of all schools in the state. Participants consisted of 438 teachers and administrators from 13 school districts (Table 1). The majority (95%) were teachers, and 5% were school administrators. Less than two-thirds (63%) had a master's degree or above, and more than one third (37%) had a bachelor's degree or below. Most (86.9%) of the participants were career educators with more than 3 years of experience in education, and about 13% were early career educators with 3 or fewer years of experience in education.

The second survey was conducted after the full implementation of the teacher evaluation system in the state. I used a stratified random sampling method, and eight school districts were selected based on their poverty index and enrollment. I invited the district administrators to participate in the study, and three school districts participated in the study. One school district has a low poverty and a medium enrollment, one school district has a medium poverty and a large enrollment, and one school district has a high poverty and a small enrollment. These school districts are considered representative of schools in the state. Participants consisted of 260 teachers and administrators in the state (Table 1). The majority (93.8%) were teachers, and 6.2% were school administrators. Slightly more than two-thirds (67.6%) had a master's degree or above, and less than one third (32.4%) had a bachelor's degree or below. Most (88.3%) of the participants were

**Table 1** Participants' information

| Variable   | Level              | Before implementation |      | After implementation |      |
|------------|--------------------|-----------------------|------|----------------------|------|
|            |                    | <i>N</i>              | %    | <i>N</i>             | %    |
| Position   | Teacher            | 416                   | 95.0 | 244                  | 93.8 |
|            | Administrator      | 22                    | 5.0  | 16                   | 6.2  |
| Degree     | Bachelor or below  | 162                   | 37.0 | 82                   | 32.4 |
|            | Master or above    | 276                   | 63.0 | 171                  | 67.6 |
| Experience | Early career (0–3) | 57                    | 13.1 | 30                   | 11.7 |
|            | Career (3+)        | 379                   | 86.9 | 227                  | 88.3 |



career educators with more than 3 years of experience in education, and about 12% were early career educators with 3 or fewer years of experience in education.

### 3.3 Instrument

Both surveys used the same instrument with four sections. The first section included four questions regarding the impact of SLOs on teacher effectiveness, instruction, student learning, and professional development. The questions were on a 4-point scale, with 1 (strongly disagree) to 4 (strongly agree). The second section included nine questions about teachers' knowledge about SLOs, and the questions were on a 4-point scale, with 1 (no knowledge) to 4 (substantial knowledge). The third section had six questions about teachers' need for support to successfully implement SLOs, and the questions were on a 3-point scale (1-need no support; 2-need some support; 3-need a lot of support). The fourth section included four questions regarding the impact of classroom observations on teacher effectiveness, instruction, student learning, and professional development. The questions were on a 4-point scale, with 1 (strongly disagree) to 4 (strongly agree). In addition to the content area questions, educators' educational and professional background information (e.g., degree, years of experience in education) was collected.

The two survey instruments are a revised version of the survey instrument that was previously used in a project that was focused on evaluating teacher effectiveness. Based on the responses in the previous project, Cronbach's alphas are 0.83 for the first subscale and 0.76 for the second subscale, which are acceptable (Nunnally & Bernstein, 1994). In this study, I made revisions, so the survey items are applicable for both teachers and administrators. I also developed additional questions related to using SLOs in the evaluation of teacher effectiveness. To examine the reliability of the instrument, I calculated Cronbach's alpha coefficients which ranged from .88 to .96 based on the BI survey and from .92 to .93 based on the AI survey. These coefficients are acceptable according to Nunnally and Bernstein (1994). To examine the content validity of the instrument, I invited five professionals who have expertise in the fields of educational assessment, evaluation, and survey design to review the instrument. I made revisions based on suggestions from the reviewers.

To gain more information about educators' perceptions of the evaluation of teacher effectiveness, I added one open-ended question at the end of the survey after implementation. Educators were asked to share their thoughts about using SLOs in teacher evaluation or teacher evaluation in general. This open-ended question provided educators an opportunity to express their opinions that might not be covered in the quantitative questions, which helps gain more in-depth information about teacher evaluation.

### 3.4 Data collection and data analysis

The first survey was conducted about 1 year before the full implementation of the teacher evaluation system in 2017, and the second survey was conducted about 1 year and a half after the full implementation of the teacher evaluation system in 2020. District and school administrators were contacted to distribute the survey link to the teachers and administrators. I used SurveyMonkey to collect data, and the data collection took about 6 weeks for each survey.

Data were analyzed based on the research questions. Descriptive statistics including percentages and means were calculated regarding teachers' and administrators' views of the impact of SLOs and classroom observations, their knowledge about SLOs, and support needed to implement SLOs. To understand whether the differences of these domains between teachers and administrators were statistically significant, I used independent *t*-tests. The overall alpha was set to be .05. With 4 items for comparison in the views of impact, I applied Bonferroni correction to adjust the familywise alpha of .05 to reduce the Type I error to .013 (i.e., .05/4). For knowledge, with 9 items for comparison, I applied Bonferroni correction to adjust the familywise alpha of .05 to reduce the Type I error to .006 (i.e., .05/9). For support, with 6 items for comparison, I applied Bonferroni correction to adjust the familywise alpha of .05 to reduce the Type I error to .008 (i.e., .05/6). In addition, confidence intervals were constructed to better understand the specific ranges of differences between teachers and administrators regarding their views of the impact of SLOs and classroom observations, knowledge about SLOs, and support needed in implementing SLOs. Further, considering the small sample sizes of administrators in both the BI survey and the AI survey, I also used a non-parametric method Mann-Whitney test to check the results based on the parametric method independent *t*-test.

Finally, the open-ended question in the AI survey was coded qualitatively to explore in-depth the possible reasons and explanations for the differences of educators' views before and after the full implementation of the evaluation system. In vivo coding method was used to emphasize the actual language used by the educators. In data analysis, I first read all responses carefully and identified responses that were relevant to the impact of SLOs and classroom observation. Second, I used R for Qualitative Analysis (RQDA) and coded the relevant responses for patterns and themes. Finally, I checked the themes with the findings based on the quantitative questions in the AI survey to identify associations and consistencies.

## 4 Results

### 4.1 Impact of SLOs

To understand how teachers and administrators view the impact of SLOs, I calculated both percentages and means for the two surveys before and after the implementation of the teacher evaluation system in the state (Table 2). Overall, educators in the AI survey reported much less positive views of SLOs than those in the BI survey. Between 66 and 79% of the educators in the BI survey and between 28 and

**Table 2** Views of the impact of SLOs

| SLOs impact                                | % agreement |         |               |               | Mean    |         |               |               | P    | 95% CI        |
|--|-------------|---------|---------------|---------------|---------|---------|---------------|---------------|------|---------------|
|  | Teacher     |         | Administrator |               | Teacher |         | Administrator |               |      |               |
|  | All         | Teacher | All           | Administrator | All     | Teacher | All           | Administrator |      |               |
| BI   |             |         |               |               |         |         |               |               |      |               |
| Evaluating teacher performance effectively | 66.0        | 64.5    | 95.0          | 95.0          | 2.68    | 2.66    | 3.05          | 3.05          | .000 | [0.19, 0.59]  |
| Improving teachers' instructional practice | 75.1        | 74.0    | 90.5          | 90.5          | 2.86    | 2.84    | 3.25          | 3.25          | .024 | [0.06, 0.76]  |
| Promoting student learning                 | 79.1        | 78.3    | 95.0          | 95.0          | 2.94    | 2.92    | 3.35          | 3.35          | .014 | [0.09, 0.77]  |
| Informing teacher PD                       | 74.8        | 73.9    | 90.0          | 90.0          | 2.86    | 2.85    | 3.10          | 3.10          | .142 | [-0.09, 0.59] |
| Overall                                    | 73.8        | 72.7    | 92.6          | 92.6          | 2.83    | 2.82    | 3.19          | 3.19          | .001 | [0.17, 0.57]  |
| AI   |             |         |               |               |         |         |               |               |      |               |
| Evaluating teacher performance effectively | 27.6        | 25.1    | 43.8          | 43.8          | 2.07    | 2.03    | 2.25          | 2.25          | .296 | [-0.20, 0.65] |
| Improving teachers' instructional practice | 38.9        | 36.4    | 56.3          | 56.3          | 2.24    | 2.21    | 2.44          | 2.44          | .306 | [-0.21, 0.68] |
| Promoting student learning                 | 38.4        | 36.8    | 46.7          | 46.7          | 2.26    | 2.23    | 2.33          | 2.33          | .651 | [-0.36, 0.57] |
| Informing teacher PD                       | 38.7        | 36.7    | 50.0          | 50.0          | 2.21    | 2.17    | 2.38          | 2.38          | .354 | [-0.23, 0.63] |
| Overall                                    | 35.9        | 33.8    | 49.2          | 49.2          | 2.19    | 2.16    | 2.35          | 2.35          | .323 | [-0.19, 0.58] |

BI before implementation, AI after implementation

39% of the educators in the AI survey agreed or strongly agreed that using SLOs evaluates teacher performance effectively, improves teachers' instructional practice, promotes student learning, and informs teachers' professional development. Within each survey, administrators reported more positive views of SLOs than teachers. Between 90 and 95% of the administrators and between 65 and 78% of the teachers agreed or strongly agreed on the impact of SLOs in the BI survey, and between 44 and 56% of the administrators and between 25 and 37% of the teachers agreed or strongly agreed on the impact of SLOs in the AI survey.

Independent *t*-tests were conducted based on the mean differences of teachers' and administrators' views of SLOs. In the BI survey, there were statistically significant differences between teachers' and administrators' views of the overall impact of SLOs ( $p = .001$ ,  $d = 0.65$ ) and SLOs' evaluating teacher performance effectively with medium to large effects ( $p < .001$ ,  $d = 0.62$ ) (Cohen, 1988). It suggests that in comparison with teachers, administrators had statistically significantly higher agreement with the statement that using SLOs evaluates teacher performance effectively and the overall impact of SLOs. Confidence intervals indicated that the belief in whether SLOs can be used to evaluate teacher performance effectively is likely between 0.19 and 0.59 points higher for administrators than for teachers. The overall impact of SLOs is likely between 0.17 and 0.57 points higher for administrators than for teachers. To check the analysis, I also used Mann-Whitney tests considering the sample size of administrators was small. Based on the analysis results, there was a statistically significant difference between teachers' views and administrators' views of the overall impact of SLOs ( $Z = -2.84$ ,  $p = .005$ ) and SLOs' impact on student learning ( $Z = -2.53$ ,  $p = .012$ ). In the AI survey, administrators appeared to hold slightly more positive views of the impact of SLOs than teachers, but the differences of their views were not statistically significant.

## 4.2 Impact of classroom observations

To understand how teachers and administrators view the impact of classroom observations, I calculated both percentages and means for the two surveys (Table 3). Overall, educators in the AI survey reported less positive views of classroom observations than those in the BI survey. Between 77 and 81% of the educators in the BI survey and between 58 and 71% of the educators in the AI survey agreed or strongly agreed that using classroom observations evaluates teacher performance effectively, improves teachers' instructional practice, promotes student learning, and informs teachers' professional development. Within each survey, administrators reported more positive views of classroom observations than teachers. All of the administrators (100%) and between 76 and 80% of the teachers agreed or strongly agreed on the impact of classroom observations in the BI survey, and between 94 and 100% of the administrators and between 55 and 69% of the teachers agreed or strongly agreed on the impact of classroom observations in the AI survey.

Independent *t*-tests were conducted based on the mean differences of teachers' and administrators' views of classroom observations. In both surveys, administrators reported statistically significantly more positive views of classroom observations

**Table 3** Views of the impact of classroom observations

| Classroom observation impact | % agreement                                |               |               |               | Mean    |               |               |               |       |              |
|------------------------------|--|---------------|---------------|---------------|---------|---------------|---------------|---------------|-------|--------------|
|                              | All  |               | Administrator |               | All     |               | Administrator |               |       |              |
|                              | Teacher                                    | Administrator | Teacher       | Administrator | Teacher | Administrator | Teacher       | Administrator |       |              |
| BI                           | Evaluating teacher performance effectively | 77.4          | 76.0          | 100.0         | 2.85    | 3.44          | 2.81          | 3.44          | <.001 | [0.28, 0.97] |
|                              | Improving teachers' instructional practice | 80.4          | 79.3          | 100.0         | 2.87    | 3.38          | 2.84          | 3.38          | .002  | [0.20, 0.88] |
|                              | Promoting student learning                 | 79.2          | 78.0          | 100.0         | 2.89    | 3.44          | 2.85          | 3.44          | .001  | [0.25, 0.92] |
|                              | Informing teacher PD                       | 81.1          | 79.9          | 100.0         | 2.90    | 3.44          | 2.87          | 3.44          | .001  | [0.25, 0.89] |
|                              | Overall                                    | 79.5          | 78.3          | 100.0         | 2.88    | 3.42          | 2.84          | 3.42          | <.001 | [0.27, 0.88] |
| AI                           | Evaluating teacher performance effectively | 71.3          | 69.3          | 100.0         | 2.81    | 3.31          | 2.77          | 3.31          | .011  | [0.12, 0.96] |
|                              | Improving teachers' instructional practice | 70.1          | 68.8          | 100.0         | 2.79    | 3.31          | 2.76          | 3.31          | .000  | [0.28, 0.83] |
|                              | Promoting student learning                 | 57.6          | 55.4          | 93.8          | 2.59    | 3.25          | 2.55          | 3.25          | .003  | [0.24, 1.16] |
|                              | Informing teacher PD                       | 70.0          | 68.5          | 100.0         | 2.79    | 3.44          | 2.74          | 3.44          | .003  | [0.25, 1.15] |
|                              | Overall                                    | 67.3          | 65.5          | 98.5          | 2.75    | 3.33          | 2.71          | 3.33          | .002  | [0.22, 1.02] |

BI before implementation, AI after implementation

than teachers. Administrators reported significantly higher agreement on the overall impact of classroom observations with large effect sizes (BI survey:  $p < .001$ ,  $d = 1.11$ ; AI survey:  $p = .002$ ,  $d = 0.94$ ), and on the impact of classroom observations on evaluating teacher performance effectively with a large effect size (BI survey:  $p < .001$ ,  $d = 1.04$ ; AI survey:  $p = .011$ ,  $d = 0.79$ ), improving teachers' instructional practice with a large effect size (BI survey:  $p = .002$ ,  $d = 0.91$ ; AI survey:  $p < .001$ ,  $d = 0.77$ ), promoting student learning with a large effect size (BI survey:  $p = .001$ ,  $d = 0.98$ ; AI survey:  $p = .003$ ,  $d = 0.83$ ), and informing teachers' professional development with a large effect size (BI survey:  $p = .001$ ,  $d = 0.98$ ; AI survey:  $p = .003$ ,  $d = 0.95$ ) (Cohen, 1988). The confidence intervals provided projected ranges of differences between administrators and teachers regarding their views of the impact of classroom observations. The Mann-Whitney tests revealed consistent findings, and there were statistically significant differences between teachers' and administrators' views of the overall impact of classroom observations (BI survey:  $Z = -4.29$ ,  $p < .001$ ; AI survey:  $Z = -3.38$ ,  $p = .001$ ). These findings suggest that administrators hold significantly more positive views of using classroom observations in teacher evaluation regardless of the implementation of the evaluation system.

### 4.3 Knowledge about SLOs

To understand how teachers and administrators differ in their reported knowledge about SLOs, I calculated both percentages and means for the two surveys (Table 4). Overall, educators reported similar levels of knowledge about SLOs in both BI and AI surveys. Between 82 and 93% of the educators in the BI survey and between 81 and 92% of the educators in the AI survey reported to have some or substantial knowledge about SLOs. Based on the BI survey, higher percentages of administrators than teachers reported to have some or substantial knowledge on seven out of the nine items. Based on the AI survey, higher percentages of administrators than teachers reported to have some or substantial knowledge on all nine items.

Independent *t*-tests were conducted based on the mean differences of teachers' and administrators' knowledge about SLOs. Based on the BI survey, administrators and teachers did not report statistically significantly different levels of knowledge about SLOs. The confidence intervals provided projected ranges of differences between administrators and teachers regarding their knowledge about SLOs. The Mann-Whitney tests revealed consistent findings. Based on the AI survey, administrators had significantly more knowledge than teachers regarding overall knowledge ( $p = .019$ ,  $d = 0.64$ ), student groups to be included ( $p = .003$ ,  $d = 0.69$ ), content to be included ( $p = .003$ ,  $d = 0.66$ ), and analyzing student assessment data ( $p = .001$ ,  $d = 0.77$ ) with medium to large effects (Cohen, 1988). Confidence intervals were constructed to understand the ranges of the differences between teachers and administrators. The Mann-Whitney tests revealed consistent findings, and administrators reported statistically significantly more overall knowledge than teachers ( $Z = -2.58$ ,  $p = .010$ ).

**Table 4** Knowledge about SLOs

| SLOs Knowledge                                | % agreement |         |               | Mean |         |               | P    | 95% CI        |
|---|-------------|---------|---------------|------|---------|---------------|------|---------------|
|   | All         | Teacher | Administrator | All  | Teacher | Administrator |      |               |
| BI  |             |         |               |      |         |               |      |               |
| Purpose of SLOs                               | 91.6        | 91.4    | 95.2          | 3.33 | 3.33    | 3.43          | .498 | [-0.20, 0.40] |
| Student groups to be included in SLOs         | 88.4        | 88.0    | 95.3          | 3.25 | 3.25    | 3.38          | .404 | [-0.18, 0.45] |
| Content to be included in SLOs                | 92.6        | 92.4    | 95.2          | 3.34 | 3.34    | 3.43          | .526 | [-0.19, 0.38] |
| Implementation of SLOs in the district        | 82.2        | 81.7    | 90.5          | 3.08 | 3.07    | 3.29          | .206 | [-0.12, 0.56] |
| Developing high-quality SLOs                  | 85.2        | 85.6    | 81.0          | 3.17 | 3.18    | 3.10          | .606 | [-0.40, 0.23] |
| Selecting appropriate assessments             | 87.1        | 87.2    | 85.7          | 3.21 | 3.21    | 3.24          | .868 | [-0.28, 0.33] |
| Setting growth targets for SLOs               | 86.9        | 86.4    | 95.3          | 3.19 | 3.19    | 3.24          | .763 | [-0.26, 0.35] |
| Instructional strategies to meet SLOs targets | 89.3        | 89.3    | 90.4          | 3.25 | 3.25    | 3.24          | .951 | [-0.31, 0.29] |
| Analyzing student assessment data             | 89.4        | 89.3    | 90.4          | 3.25 | 3.25    | 3.24          | .934 | [-0.31, 0.28] |
| Overall                                       | 88.1        | 87.9    | 91.0          | 3.23 | 3.23    | 3.29          | .670 | [-0.21, 0.32] |
| AI  |             |         |               |      |         |               |      |               |
| Purpose of SLOs                               | 90.9        | 91.7    | 93.8          | 3.37 | 3.35    | 3.75          | .017 | [0.08, 0.71]  |
| Student groups to be included in SLOs         | 89.3        | 89.7    | 100.0         | 3.35 | 3.34    | 3.75          | .003 | [0.16, 0.66]  |
| Content to be included in SLOs                | 92.1        | 90.9    | 100.0         | 3.46 | 3.44    | 3.81          | .003 | [0.15, 0.61]  |
| Implementation of SLOs in the district        | 84.8        | 84.7    | 93.8          | 3.25 | 3.23    | 3.69          | .018 | [0.08, 0.84]  |
| Developing high quality SLOs                  | 81.6        | 82.3    | 93.8          | 3.16 | 3.16    | 3.50          | .081 | [-0.04, 0.71] |
| Selecting appropriate assessments             | 85.8        | 86.5    | 93.8          | 3.26 | 3.25    | 3.50          | .201 | [-0.13, 0.62] |
| Setting growth targets for SLOs               | 80.9        | 81.4    | 93.8          | 3.20 | 3.20    | 3.50          | .140 | [-0.10, 0.69] |
| Instructional strategies to meet SLOs targets | 88.5        | 89.3    | 93.8          | 3.33 | 3.33    | 3.50          | .353 | [-0.19, 0.52] |
| Analyzing student assessment data             | 87.9        | 89.0    | 100.0         | 3.38 | 3.37    | 3.81          | .001 | [0.22, 0.67]  |
| Overall                                       | 86.9        | 87.3    | 95.9          | 3.31 | 3.30    | 3.65          | .019 | [0.06, 0.64]  |

BI before implementation, AI after implementation

**Table 5** Support needed in implementing SLOs

| Support needed                              | % agreement |         |               | Mean |         |               | P    | 95% CI        |
|---|-------------|---------|---------------|------|---------|---------------|------|---------------|
|   | All         | Teacher | Administrator | All  | Teacher | Administrator |      |               |
|   |             |         |               |      |         |               |      |               |
| BI  |             |         |               |      |         |               |      |               |
| Understanding standards in SLOs             | 39.8        | 39.9    | 38.1          | 1.43 | 1.43    | 1.43          | .983 | [-0.25, 0.24] |
| Implementing standards in SLOs              | 42.9        | 42.7    | 47.7          | 1.48 | 1.47    | 1.52          | .710 | [-0.21, 0.31] |
| Understanding cognitive levels of standards | 59.3        | 58.7    | 70.0          | 1.66 | 1.66    | 1.75          | .511 | [-0.18, 0.37] |
| Developing assessments for SLOs             | 61.3        | 60.3    | 80.9          | 1.69 | 1.68    | 1.90          | .107 | [-0.05, 0.49] |
| Setting growth targets for SLOs             | 65.6        | 64.5    | 85.0          | 1.76 | 1.75    | 1.95          | .162 | [-0.08, 0.48] |
| Analyzing assessment data in SLOs           | 62.9        | 62.4    | 71.4          | 1.71 | 1.71    | 1.71          | .971 | [-0.26, 0.27] |
| Overall                                     | 55.3        | 54.8    | 65.5          | 1.62 | 1.62    | 1.71          | .369 | [-0.11, 0.30] |
| AI  |             |         |               |      |         |               |      |               |
| Understanding standards in SLOs             | 34.7        | 33.7    | 25.0          | 1.39 | 1.38    | 1.25          | .274 | [-0.38, 0.11] |
| Implementing standards in SLOs              | 35.5        | 35.0    | 37.5          | 1.39 | 1.39    | 1.38          | .934 | [-0.29, 0.27] |
| Understanding cognitive levels of standards | 46.7        | 43.8    | 50.0          | 1.53 | 1.50    | 1.50          | .979 | [-0.32, 0.31] |
| Developing assessments for SLOs             | 38.5        | 37.0    | 43.8          | 1.45 | 1.42    | 1.56          | .373 | [-0.17, 0.45] |
| Setting growth targets for SLOs             | 47.1        | 45.1    | 50.0          | 1.55 | 1.52    | 1.63          | .521 | [-0.22, 0.43] |
| Analyzing assessment data in SLOs           | 37.9        | 36.0    | 25.1          | 1.44 | 1.41    | 1.31          | .508 | [-0.40, 0.20] |
| Overall                                     | 40.1        | 38.4    | 38.6          | 1.46 | 1.44    | 1.44          | .992 | [-0.26, 0.25] |

BI before implementation, AI after implementation



#### 4.4 Support needed in implementing SLOs

To understand how teachers and administrators differ in their support needed in implementing SLOs, I calculated both percentages and means for the two surveys (Table 5). Overall, smaller percentages of educators in the AI survey than in the BI survey reported needing some or a lot of support in implementing SLOs. Between 40 and 66% of the educators in the BI survey and between 35 and 47% of the educators in the AI survey reported needing some or a lot of support. Based on the BI survey, higher percentage of administrators (65.5%) than teachers (54.8%) reported needing some or a lot of overall support, and higher percentages of administrators than teachers reported needing some or a lot of support on five out of the six items. Based on the AI survey, similar percentages of administrators (38.6%) and teachers (38.4%) reported needing some or a lot of overall support, and higher percentages of administrators than teachers reported to need some or a lot of support on four out of the six items.

Independent *t*-tests were conducted based on the mean differences of teachers' and administrators' reported support needed in implementing SLOs. Based on both the BI survey and the AI survey, administrators and teachers did not report statistically significantly different levels of support needed in implementing SLOs. The confidence intervals provided projected ranges of differences between administrators and teachers regarding their support needed. The Mann-Whitney tests revealed consistent findings, there were no statistically significant differences between teachers' and administrators' reported support needed based on both the BI survey and AI survey.

#### 4.5 Educators' thoughts about SLOs and teacher evaluation

The quantitative questions appeared to reveal that educators reported less positive views of SLOs and classroom observations after the implementation of the system. To understand the reasons that could possibly help explain this finding, I coded the open-ended question in the AI survey. At the end of the survey, participants were asked to share their thoughts about using SLOs in teacher evaluation or teacher evaluation in general. Among 260 participants, 114 educators (106 teachers and eight administrators) shared additional thoughts. No differences were identified between administrators and teachers regarding their views. Overall, educators expressed various concerns about using SLOs in the evaluation of teacher effectiveness. These concerns included issues of using students' test performance to evaluate teacher effectiveness; issues of the assessment methods in measuring students' growth; time and timeline in implementing SLOs; paperwork; applicability to special education teachers, arts teachers, ESL teachers, and media specialists; lack of supervision in the process; subjectivity in goal-setting and assessment; missing some teaching standards; teaching to the test; and lack of feedback.

Educators indicated that SLOs were neither effective nor accurate in the evaluation of teacher effectiveness, and they considered that their SLOs results did not accurately reflect teachers' performance or effectiveness. They shared that

students' test performance and academic growth were affected by various factors, and it was not fair, nor reliable or valid, to judge teacher performance or effectiveness based on student test performance/results. Educators explained that teachers should be evaluated based on teacher performance rather than student performance. Some supported the use of classroom observations in the evaluation of teacher effectiveness and considered classroom observations as a better way to capture teacher performance. However, a few teachers shared that the observational instrument was too cumbersome, and limited times of observations might not capture some indicators. In addition, a few teachers shared that some evaluators might not have the required qualifications and might be biased in evaluating teacher effectiveness fairly and accurately.

Setting growth targets and assessment method was a major concern. Educators shared that they did not receive guidance in setting reasonable learning goals for students. There were mainly two types of assessment methods: standardized tests and teacher developed assessment. There were issues of using standardized tests to measure student growth. Some standardized tests (e.g., Measures of Academic Progress) did not align with the grade-level standards. The pre/post assessment was invalid for some subject areas (e.g., science). Regarding the teacher developed assessment methods, there were flaws as well. A few educators indicated that some of their colleagues purposefully set low learning objectives/goals/targets and manipulated/alterd their students' growth data to meet the goals. In addition, some teachers set growth targets based on one standard, which led to the neglect of other teaching standards.

Time, timeline, and paperwork were a major concern. Educators considered the SLOs process to be time-consuming, required a lot of paperwork, added a lot of unnecessary extra work to their schedule, and became a burden or stress. In particular, many educators shared that timeline was inappropriate. They were required to submit their SLOs student assessment results a few months before the final assessment of student learning. The assessment data would not accurately capture students' growth for the year.

Some educators indicated that SLOs were not applicable or less effective for special education teachers, arts teachers, gifted teachers, ESL teachers, librarians, speech pathologists, media specialists, and guidance counselors. Educators believed that SLOs should not be one size fits all. They indicated that it was challenging for special education students to show growth required by the SLOs, and some special education teachers were already using Individualized Education Program (IEP) goals and objectives, and using SLOs was redundant to them.

Although educators indicated various concerns about using SLOs in the evaluation of teacher effectiveness, some did indicate that they supported the idea of using SLOs to measure student growth. Educators understood the importance of teacher evaluation, and they believed that teachers should be accountable for student learning. It appears that early career teachers need some training or instruction in implementing SLOs. A few early career educators shared that the SLOs information was overwhelming, and they did not have enough knowledge about SLOs, and they needed better orientation and training about the implementation

of SLOs. Educators expected the evaluation to help them grow and improve rather than judging their teaching. Educators hoped that the evaluators should provide them with constructive feedback and help them grow/improve instead of assigning them with a number/rating, checking the box, or being judgmental and punitive.

## 5 Discussion and conclusion

This study found that administrators and teachers held different views of classroom observations and SLOs in the evaluation of teacher effectiveness. Administrators had a significantly stronger belief that using classroom observations evaluates teacher performance effectively, improves teachers' instructional practice, promotes student learning, and informs teachers' professional development both before and after the implementation of the evaluation system. In comparison with teachers, administrators had a stronger belief that using SLOs evaluates teacher performance effectively before the implementation of the evaluation system. These findings echo those by Reddy et al. (2018) who found that school administrators reported more favorable experiences than teachers regarding teacher evaluation. The differences of the views between teachers and administrators might be attributed to the different roles that they play in the process of teacher evaluation. Administrators have authority, and they are often the evaluators of teacher effectiveness and decision makers in terms of teachers' employment and promotion. The findings also echo Huber and Skedsmo (2016b) who indicated that role and authority have a crucial influence on defining accountability relationships. Teachers are classroom instructors who work closely with students and are often evaluated through various methods. While being evaluated, teachers often reported a lot of stress and anxiety (Ryan et al., 2017) and face dilemmas (Shaw, 2019). These reasons might be able to help explain the different views of teachers and administrators regarding the evaluation of teacher effectiveness.

Different views between teachers and administrators were also documented by Bradford and Braaten (2018) who indicated that school leaders and teachers had different interpretations of scores. The impact of school administrators on teacher development and retention was investigated by various studies. Studies found that lack of administrative support was one of the key factors contributed to high teacher turnover rates (Carver-Thomas & Darling-Hammond, 2019). Teachers' views of school administration had the greatest influence on their decisions on stay or leave (Boyd et al., 2010). Therefore, in teacher evaluation, school administrators and evaluators should be aware of and acknowledge these differences between teachers and administrators, build mutual understanding and trust, and collaboratively implement the evaluation system for the ultimate goal of improving teaching practice and supporting student learning. While teachers are skeptical about the evaluation of teacher effectiveness, teacher voices should be heard, and they should actively participate in the design and implementation of the teacher evaluation systems (Skedsmo & Huber, 2018). Teachers should be

empowered and become part of the decision-making team in the evaluation of teacher effectiveness.

The differences between teachers and administrators were also identified in their reported knowledge about SLOs and support needed to implement SLOs. Although higher percentages of administrators reported to have some or substantial knowledge on seven out of the nine items, the differences were not statistically significant before the implementation. Higher percentages of administrators than teachers reported to have some or substantial knowledge on all nine items after the implementation, and administrators reported significantly more overall knowledge about SLOs and knowledge about student groups and content to be included and analyzing student assessment data. This is probably due to the statewide training among school administrators regarding the SLOs implementation. It further suggests that teachers should receive trainings and obtain advanced knowledge about SLOs implementation, and schools should build learning communities to help teachers develop knowledge about SLOs implementation. Administrators and teachers reported very similar levels of support needed after the implementation, although notably higher percentages of administrators than teachers reported needing some or a lot of support before the implementation. This could possibly be attributed to their experience of using SLOs after the implementation of the evaluation system. Even after the implementation of the evaluation system, teachers and administrators seemed to need more support regarding setting growth targets for students, understanding the cognitive levels of standards, and developing assessments. Therefore, professional development for teachers and administrators should focus on these areas.

This study revealed that teachers and administrators had less positive views of SLOs and classroom observations after the full implementation of the teacher evaluation system in the state. Regarding the impact of SLOs, much smaller percentages of educators in the AI survey than in the BI survey reported agreement on the overall impact of SLOs and the impact of SLOs on evaluating teacher performance, improving teachers' instructional practice, promoting student learning, and informing teachers' professional development. Regarding the impact of classroom observations, notably smaller percentages of educators in the AI survey than in the BI survey reported agreement on the impact of classroom observations. These changes of educators' views after the implementation of the evaluation systems could be explained through the themes based on the coding of the open-ended question in the AI survey. It appears that educators had various concerns related to SLOs implementation and teacher evaluation in general. These concerns were related to linking students' test performance to teacher effectiveness, issues of the assessment methods, time and paperwork in implementing SLOs, applicability to some subject areas, lack of supervision in the process, subjectivity in goal-setting, missing some teaching standards, teaching to the test, and lack of feedback. These challenges that educators faced in evaluation might help explain their less positive views of SLOs and classroom observation after the implementation of the system. This further indicates the inconsistency between teachers' views of using classroom observations and their views of using student learning outcomes in the evaluation of teacher effectiveness. The findings are consistent with those by Skedsmo and Huber (2017) who

highlighted the concerns related to implementation in different contexts of teacher evaluation models that were designed to measure teaching quality.

One major concern was the issue of linking students' test performance to teacher effectiveness, and educators appeared to be skeptical about accountability in the evaluation of teacher effectiveness. As Bovens (2010) and Lerner and Tetlock (1999) described, accountability means that a person has an obligation or is expected to explain and justify their beliefs, feelings, and actions to others. While teachers mostly agreed before the full implementation of the teacher evaluation system that using student learning outcomes can evaluate teacher performance effectively, there are other factors that contribute to student learning, which include classroom learning environment, strong relationship with students, and engaging students. Meng and Muñoz (2016) indicated that teachers in both the USA and China considered engaging students as one of the most important characteristics of effective teaching. However, these factors are not included in the teacher evaluation frameworks. The teacher evaluation system should be reformed. If student learning outcomes are used as a measure of teachers' accountability for student learning in the evaluation of teacher effectiveness, there must be multiple accurate measures of student outcomes. In addition to academic learning, student's other areas of learning and development (e.g., student engagement, social emotional learning) should be considered in future teacher evaluation system.

Educators' negative views of using student learning outcomes in the evaluation of teacher effectiveness appeared to be related to the assessment methods, and students' assessment results may not be valid due to various issues. Paufler and Clark (2019) indicated that administrators had concerns about the validity and reliability issues in measuring teacher effectiveness, and they discouraged external, high-stakes uses of teacher evaluation results. They valued the evaluation process as a support teacher growth. There is a need for valid measure of teacher effectiveness (Skedsmo & Huber, 2018). Student learning outcomes should be used as a formative assessment process to help understand student learning in regard to the learning objectives rather than a template to be completed by the teachers (Schneider & Johnson, 2019). The development and implementation of the teacher evaluation system should be aligned with appropriate evaluation purposes. As Huber and Skedsmo (2016a) suggested that "combining too many purposes in one evaluation model might diminish the likelihood of achieving them" (p. 109, 2016a). The design, development, and implementation of a teacher evaluation system should take the purposes of assessment including summarizing student learning, informing instruction, and evaluating teacher effectiveness into consideration (Briggs et al., 2019; Ford & Hewitt, 2020).

This study revealed that some teachers left their profession due to the stress originated from teacher evaluation. These are probably the unintended consequences of evaluation models (Huber & Skedsmo, 2016a). Considering the current serious issues of nationwide teacher shortage and teacher attrition (Sutcher et al., 2019), teacher evaluation should aim at improving teacher quality that focuses on craft, care, and creativity rather than technical effectiveness (Anagnostopoulos et al., 2021). Schools should build a positive school climate and supportive administration to motivate and support teachers, thus hopefully retaining effective teachers (Skaalvik & Skaalvik, 2011). Considering that early career teachers are more likely

to leave their profession, they should be provided with more support because they reported less knowledge about SLOs and more support needed in implementing SLOs (Fan, 2022).

There are several limitations that should be noted. First, the two surveys collected educators' self-reported data, and some of the responses might be subjective. Therefore, future studies should employ multiple methods (e.g., interviews, focus groups) to gain a full picture of educators' views of using SLOs in teacher evaluation. Second, educators' perceptions of SLOs were collected before and after the implementation of the evaluation system. I believe that their perspectives might change along with longer time of implementation and more experience of using SLOs. Therefore, longitudinal data about educators' views of SLOs should be collected to identify the changes over time. Third, this study was based on the data collected from one southeastern state in the USA. The characteristics of teacher workforce are different due to the different geographic, social, and demographic characteristics among states. Therefore, we should be cautious about generalizing the findings of this study to the other states in the nation. In addition, only 22 administrators completed the survey before the implementation, and 16 administrators completed the survey after the implementation. Future studies should use effective strategies to improve the response rate of administrative participants. Furthermore, this study focused on comparing educators' views based on their position, and I believe that other factors including school district, school location, and school type might also impact educators' views. Future studies should consider these factors to better understand the views of educators. Finally, it would be important to involve students and parents in understanding the use of SLOs in the evaluation of teacher effectiveness (Salazar & Lerner, 2019).

The findings of this study could be used to inform decision making in teacher evaluation and teachers' professional development. This study revealed that teachers anticipated to receive constructive feedback from the evaluators to help improve their teaching practices. The findings echo those by Reddy et al. (2018) who found that teachers considered collaborative communication and evaluation feedback as the most helpful aspects in the process of evaluation. I believe that teacher level learning and development are crucial in the evaluation of teacher effectiveness. Tuytens et al. (2020) emphasized the importance of stimulating the ability and motivation of teachers through teacher evaluation, which may ultimately contribute to student outcomes. Similarly, Donaldson (2016) indicated that "the key to getting the most out of teacher evaluation is figuring out how to implement it in a way that challenges, supports, and motivates teachers" (p. 76). Teacher evaluation should be used to motivate teachers rather than adding stress and anxiety to them. I recommend reforming the teacher evaluation system and using a comprehensive evaluation framework that incorporates student learning and development, teacher learning and development, and school effectiveness and improvement as the goals of the evaluation. Most importantly, policy making in the evaluation of teacher effectiveness should be based on evidence from scientific and empirical research. Hallinger et al. (2014) studied teacher evaluation and school improvement and indicated that in the reform of teacher evaluation system, policy logic played a considerably stronger role than the empirical

evidence. There was a lack of support for the belief that teacher evaluation represents a high impact school improvement strategy. A successful teacher evaluation system should focus on teacher learning and development. Through the process of teacher evaluation, teachers become more motivated, confident, qualified, and effective in supporting student learning and development.

**Data availability** Teachers and administrators who participated in the surveys were informed that the data collected would be kept confidential, and only researchers in this project have access to the data.

## Declarations

**Conflict of interest** The author declares no competing interests.

## References

- AERA. (2015). Statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448–452. <https://doi.org/10.3102/0013189X15618385>
- Anagnostopoulos, D., Wilson, S., & Charles-Harris, S. (2021). Contesting quality teaching: Teachers' pragmatic agency and the debate about teacher evaluation. *Teaching and Teacher Education*, 98, 1–15.
- Barzanò, G., & Grimaldi, E. (2013). Discourses of merit: The hot potato of teacher evaluation in Italy. *Journal of Education Policy*, 28(6), 767–791. <https://doi.org/10.1080/02680939.2013.774439>
- Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics*, 33(5), 946–967. <https://doi.org/10.1080/01402382.2010.486119>
- Boyd, D., Grossman, P., Loeb, S., & Wyckoff, J. (2010). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, 48(2), 303–333.
- Bradford, C., & Braaten, M. (2018). Teacher evaluation and the demoralization of teachers. *Teaching and Teacher Education*, 75, 49–59.
- Briggs, D. C., Chattergoon, R., & Burkhardt, A. (2019). Examining the dual purpose use of student learning objectives for classroom assessment and teacher evaluation. *Journal of Educational Measurement*, 56(4), 686–714.
- Briggs, D. C., & Dadey, N. (2017). Principal holistic judgments and high-stakes evaluations of teachers. *Educational assessment, Evaluation and Accountability*, 29, 155–178.
- Campbell, R. J., Kyriakides, L., Mujs, R. D., & Robinson, W. (2010). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education*, 29(3), 347–362.
- Carver-Thomas, D., & Darling-Hammond, L. (2019). The trouble with teacher turnover: How teacher attrition affects students and schools. *Education Policy Analysis Archives*, 27(36). <https://doi.org/10.14507/epaa.27.3699>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632. <https://doi.org/10.1257/aer.104.9.2593>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.1257/aer.104.9.2633>
- Clinton, J., & Dawson, G. (2018). Enfranchising the profession through evaluation: A story from Australia. *Teachers and Teaching*, 24(3), 312–327. <https://doi.org/10.1080/13540602.2017.1421162>
- Clinton, J., Dinham, S., Savage, G., Aston, R., Dabrowski, A., Gullickson, A., Calnin, G., & Arbour, G. (2015). *Evaluation of the implementation of the Australian professional standards for teachers: Final report*. Centre for Program Evaluation, The University of Melbourne. Retrieved December

- 2015 <https://rest.neptuneprod.its.unimelb.edu.au/server/api/core/bitstreams/ac67f530-368e-5581-b1b5-4abcf71866ee/content>
- Clinton, J. M., Anderson, M., Dawson, G., Dawson, A., Bolton, S., & Mason, R. (2016). *Teacher effectiveness systems, frameworks and measures: A review*. Centre for Program Evaluation <https://www.dese.gov.au/teaching-and-learning/resources/teacher-effectiveness-systems-frameworks-and-measures-review>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge Academic.
- Cosner, S., Kimball, S. M., Barkowski, E., Carl, B., & Jones, C. (2015). Principal roles, work demands, and supports needed to implement new teacher evaluation. *Mid-Western Educational Researcher*, 27(1), 76–95.
- Crouse, K., Gitomer, D. H., & Joyce, J. (2016). An analysis of the meaning and use of student learning objectives. In A. Amrein-Beardsley & K. Kappler Hewitt (Eds.), *Student growth measures: Where policy meets practice*. Palgrave Macmillan.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Association for Supervision and Curriculum Development.
- Danielson, C. (2013). *The framework for teaching evaluation instrument*. The Danielson Group.
- Darling-Hammond, L. (2013). *Getting teacher evaluations right: What really matters for effectiveness and improvement*. Teachers College Press.
- Darling-Hammond, L., & Young, P. (2002). Defining “highly qualified teachers”: What does “scientifically based research” actually tell us? *Educational Researcher*, 31(9), 13–25.
- Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Petegem, P. V. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education*, 36, 1–11. <https://doi.org/10.1016/j.tate.2013.06.011>
- Derrington, M. L. (2014). Teacher evaluation initial policy implementation: Superintendent and principal perceptions. *Planning and Changing*, 45(1/2), 120–137.
- Donaldson, M. L. (2012). *Teachers’ perspectives on evaluation reform*. Center for American Progress <http://www.americanprogress.org/wp-content/uploads/2012/12/TeacherPerspectives.pdf>
- Donaldson, M. L. (2016). Teacher evaluation reform: Focus, feedback, and fear. *Educational Leadership*, 73(8), 72–76.
- Donaldson, M. L. (2021). *Multidisciplinary perspectives on teacher evaluation: Understanding the research and theory*. Routledge.
- Donaldson, M. L., & Papay, J. P. (2015). Teacher evaluation for accountability and development. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of research in education finance and policy* (pp. 174–193). Routledge.
- Fan, X. (2022). Teachers’ perspectives on the evaluation of teacher effectiveness: A focus on student learning objectives. *Teaching and Teacher Education*, 110, 103604.
- Flores, M. A. (2012). The implementation of a new policy on teacher appraisal in Portugal: How do teachers experience it at school? *Educational Assessment Evaluation and Accountability*, 24, 351–368. <https://doi.org/10.1007/s11092-012-9153-7>
- Ford, T. G., & Hewitt, K. (2020). Better integrating summative and formative goals in the design of next generation teacher evaluation systems. *Education Policy Analysis Archives*, 28(63), 1–30.
- Gallagher, H. A. (2004). Vaughn Elementary’s innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79, 79–107.
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. National Comprehensive Center for Teacher Quality <http://www.gtlcenter.org/sites/default/files/docs/EvaluatingTeachEffectiveness.pdf>
- Goldhaber, D. D., & Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. In W. J. Fowler (Ed.), *Developments in school finance, 1996* (pp. 197–210). National Center for Education Statistics, U.S. Department of Education.
- Guskey, T. R. (2002a). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6), 45–51.
- Guskey, T. R. (2002b). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, 8(3), 381–391.
- Guskey, T. R. (2016). Gauge impact with 5 levels of data. *Journal of Staff Development*, 37(1), 32–37.
- Hallinger, P., Heck, R., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Journal of Educational Assessment, Evaluation and Accountability*, 26, 5–28.



- Hanushek, E. A. (2009). Teacher deselection. In D. Goldhaber & J. Hannaway (Eds.), *Creating a new teaching profession* (pp. 165–180). Urban Institute Press.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Huber, S. G., & Skedsmo, G. (2016a). Teacher evaluation – Accountability and improving teaching practices. *Educational Assessment, Evaluation and Accountability*, 28, 105–109. <https://doi.org/10.1007/s11092-016-9241-1>
- Huber, S. G., & Skedsmo, G. (2016b). Educators' responses to policies and accountability expectations. *Educational Assessment, Evaluation and Accountability*, 28, 295–298. <https://doi.org/10.1007/s11092-016-9253-x>
- Jones, E., Bergin, C., & Murphy, B. (2022). Principals may inflate teacher evaluation scores to achieve important goals. *Educational Assessment, Evaluation and Accountability*, 34, 57–88. <https://doi.org/10.1007/s11092-021-09366-8>
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Bill & Melinda Gates Foundation.
- Lachlan-Haché, L., Cushing, L., & Bivona, L. (2012). *Student learning objectives as measures of educator effectiveness: The basics*. American Institutes for Research [http://educatoralent.org/inc/docs/SLOs\\_Measures\\_of\\_Educator\\_Effectiveness.pdf](http://educatoralent.org/inc/docs/SLOs_Measures_of_Educator_Effectiveness.pdf)
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- Liu, S., & Zhao, D. (2013). Teacher evaluation in China: Latest trends and future directions. *Educational Assessment, Evaluation and Accountability*, 25, 231–250. <https://doi.org/10.1007/s11092-013-9168-8>
- Liu, Y., Visone, J., Mongillo, M. B., & Lisi, P. (2019). What matters to teachers if evaluation is meant to help them improve? *Studies in Educational Evaluation*, 61(2019), 41–54.
- Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stetcher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44(1), 47–67.
- Machin, S., & Vignoles, A. (2006). *Education policy in the UK*. Centre for the Economics of Education.
- Maharaj, S. (2014). Administrators' views on teacher evaluation: Examining Ontario's teacher performance appraisal. *Canadian Journal of Educational Administration and Policy*, 152, 1–58.
- Marshall, J. C., Smart, J., & Alston, D. M. (2016). Development and validation of Teacher Intentionality of Practice Scale (TIPS): A measure to evaluate and scaffold teacher effectiveness. *Teaching and Teacher Education*, 59(2016), 159–168.
- Marzano, R. J. (2012). The two purposes of teacher evaluation. *Educational Leadership*, 70(3), 14–19.
- Meng, L., & Muñoz, M. (2016). Teachers' perceptions of effective teaching: A comparative study of elementary school teachers from China and the USA. *Educational Assessment, Evaluation and Accountability*, 28(2), 179–199. <https://doi.org/10.1007/s11092-015-9230-9>
- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33–53.
- Moran, R. M. R. (2017). The impact of a high stakes teacher evaluation system: Educator perspectives on accountability. *Educational Studies*, 53(2), 178–193.
- Morgan, G. B., Hodge, K. J., Trepinski, T. M., & Anderson, L. W. (2014). The stability of teacher performance and effectiveness: Implications for policies concerning teacher evaluation. *Education Policy Analysis Archives*, 22(95), 1–21. <https://doi.org/10.14507/epaa.v22n95.2014>
- Morganstein, D., & Wasserstein, R. (2014). ASA statement on value-added models. *Statistics and Public Policy*, 1(1), 108–110.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Odden, A., Borman, G., & Fermanich, M. (2004). Assessing teacher, classroom, and school effects, including fiscal effects. *Peabody Journal of Education*, 79(4), 4–32.
- Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193. <https://doi.org/10.3102/0002831210362589>
- Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, 82(1), 123–141.

- Pauffer, N. A. (2018). Declining morale, diminishing autonomy, and decreasing value: Principal reflections on a high-stakes teacher evaluation system. *International Journal of Education Policy & Leadership*, 13(8). <https://doi.org/10.22230/ijepl.2018v13n8a813>
- Pauffer, N. A., & Clark, C. (2019). Reframing conversations about teacher quality: School and district administrators' perceptions of the validity, reliability, and justifiability of a new teacher evaluation system. *Educational Assessment, Evaluation and Accountability*, 31, 33–60. <https://doi.org/10.1007/s11092-019-09292-w>
- Reddy, L. A., Dudek, C. M., Peters, S., Alperin, A., Kettler, R. J., & Kurz, A. (2018). Teachers' and school administrators' attitudes and beliefs of teacher evaluation: A preliminary investigation of high poverty school districts. *Educational Assessment, Evaluation and Accountability*, 30(1), 47–70. <https://doi.org/10.1007/s11092-017-9263-3>
- Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *American Economic Review*, 100(2), 261–266.
- Romzek, B. S., & Dubnick, M. J. (1987). Accountability in the public sector: Lessons from the challenger tragedy. *Public Administration Review*, 47(3), 227–238.
- Rowan, B., Chiang, F. S., & Miller, R. J. (1996). Using research on employee performance to study the effects of teachers on students' achievement. *Sociology of Education*, 70, 256–284.
- Ryan, S. V., von der Embse, N. P., Pendergast, L. L., Saeki, E., Segool, N., & Schwing, S. (2017). Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education*, 66, 1–11.
- Salazar, M., & Lerner, J. (2019). *Teacher evaluation as culture: A framework for equitable and excellent teaching*. Routledge.
- Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education*, 14(4), 329–339.
- Schneider, M. C., & Johnson, R. L. (2019). *Using formative assessment to support student learning objectives*. Routledge.
- Shaw, R. (2019). Jumping through hoops: Troubling music teacher dilemmas in the new era of accountability. *Music Educators Journal*, 105(3), 23–29. <https://doi.org/10.1177/0027432118816146>
- Skaalvik, E. M., & Skaalvik, S. (2011). Teacher job satisfaction and motivation to leave the teaching profession: Relations with school context, feeling of belonging, and emotional exhaustion. *Teaching and Teacher Education*, 27(4), 1029–1038.
- Skedsmo, G., & Huber, S. G. (2017). Evaluation of educators' performance—balancing various measures to improve practice. *Educational Assessment, Evaluation and Accountability*, 29, 107–110. <https://doi.org/10.1007/s11092-017-9262-4>
- Skedsmo, G., & Huber, S. G. (2018). Teacher evaluation: The need for valid measures and increased teacher involvement. *Educational Assessment, Evaluation and Accountability*, 30, 1–5. <https://doi.org/10.1007/s11092-018-9273-9>
- Slotnik, W. J., Bugler, D., & Liang, G. (2014). *Real progress in Maryland: Student learning objectives and teacher and principal evaluation*. Mid-Atlantic Comprehensive Center <https://www.wested.org/resources/real-progress-in-maryland-59136/>
- Slotnik, W. J., Smith, M. D., Helms, B. J., & Qiao, Z. (2013). *It's more than money: Teacher Incentive Fund-Leadership for educators' advanced performance, Charlotte-Mecklenburg Schools*. Community Training and Assistance Center (CTAC) <http://ctacusa.com/wp-content/uploads/2013/11/MoreThanMoney.pdf>
- Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2019). Understanding teacher shortages: An analysis of teacher supply and demand in the United States. *Education Policy Analysis Archives*, 27, 35. <https://doi.org/10.14507/epaa.27.3696>
- Tuytens, M., & Devos, G. (2009). Teachers' perception of the new teacher evaluation policy: A validity study of the Policy Characteristics Scale. *Teaching and Teacher Education*, 25(6), 924–930.
- Tuytens, M., & Devos, G. (2011). Stimulating professional learning through teacher evaluation: An impossible task for the school leader? *Teaching and Teacher Education*, 27(5), 891–899. <https://doi.org/10.1016/j.tate.2011.02.004>
- Tuytens, M., & Devos, G. (2014). The problematic implementation of teacher evaluation policy: School failure or governmental pitfall? *Educational Management Administration & Leadership*, 42(4\_suppl), 155–174. <https://doi.org/10.1177/1741143213502188>
- Tuytens, M., Devos, G., & Vanblaere, B. (2020). An integral perspective on teacher evaluation: A review of empirical studies. *Educational Assessment, Evaluation and Accountability*, 32, 153–183. <https://doi.org/10.1007/s11092-020-09321-z>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.