2021

Teacher Perceptions of Using Standards-Based Rubrics for Monitoring Student Growth in Teacher Evaluation

Jennette Susan Winters
Walden University

Walden University

College of Education

This is to certify that the doctoral dissertation by

Jennette S. Winters

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. John Harrison, Committee Chairperson, Education Faculty
Dr. Sherry Lowrance, Committee Member, Education Faculty
Dr. Marilyn Robb, University Reviewer, Education Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2021

Abstract

Teachers' Perceptions of Using Standards-Based Rubrics to Monitor Student Growth in Teacher Evaluation

by

Jennette S. Winters

MAEL, Aurora University, 2000

MAT, Aurora University, 1998

BS, University of Illinois, 1992

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

November 2021

Abstract

In the United States, national and state legislative mandates have forced school districts to include student growth measures in teacher evaluation systems. However, statistical models for monitoring student growth on standardized tests have not been found to foster teachers' reflective practice or pedagogical content knowledge and goal-based models have been found to lack adequate structure for supporting implementation. This basic qualitative inquiry explored how teachers perceive using standards-based rubrics to monitor student growth for teacher evaluation influences their reflective practice and pedagogical content knowledge in mathematics. Nine teachers who have used standards-based rubrics to monitor student growth were recruited through snowball sampling. Through semi structured interviews and inductive and deductive coding, six themes were identified to understand teacher perceptions of the experience monitoring growth with standards-based rubrics: (a) fosters collaborative dialogue and descriptive feedback, (b) promotes standards-based focus, (c) supports evidence-based assessment, (d) supports student-centered instruction, (e) encourages students' reflective practice, and (f) cultivates a positive teacher evaluation experience. This study may inform standards-based growth monitoring practices for formative and summative teacher evaluation in K–8 education systems. Formative teacher evaluation has been found to promote positive social change by improving both teacher practice and student achievement, thereby supporting teachers and students to continuously grow in knowledge, skill, and understanding. These findings indicate that monitoring student growth on standards-based rubrics may provide the necessary structure other models have been lacking.

Teachers' Perceptions of Using Standards-Based Rubrics to Monitor Student Growth in Teacher Evaluation

by

Jennette S. Winters

MAEL, [Aurora University], 2000

MAT, [Aurora University], 1998

BS, [University of Illinois], 1992

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

November 2021

Dedication

I dedicate this work to the memory of my amazing parents, Eugene and Celeste Shedroff, and aunt, Ruth Cohen, whose love, and encouragement are forever in my heart. My parents always typified hard work, dedication, selflessness, and an appetite for learning that serve as a model for my own endeavors. Their support fostered my inclination for life-long learning, which I strive to emulate with my own children. This work is also dedicated to my beautiful children, Blake, Trevor, and Gabriella, and to my husband, Chuck, who inspire me to be a better teacher, learner, and human being.

Acknowledgments

As I reflect on this journey, I find myself blessed to have encountered so many individuals who have inspired and nurtured me along the way. First, I must express my gratitude to my husband, Chuck, and children, Blake, Trevor, and Gabriella, for their patience and understanding while I engaged in this personal learning quest. I am also appreciative of both my educator and non-educator friends who have shown me care and support with their encouraging words and hugs. Most specifically I want to acknowledge, Jill Farrell, Kathleen LoPriore, and Sandy Simonis, for serving as sounding boards, reading my work, and cheering me on throughout this endeavor. I want to thank my extended family, the ISBE Model Math Curriculum team, for urging me to begin this mission and motivating me along the way. I want to thank all of my Walden residency colleagues with whom I have had the pleasure of reflecting and growing. Thank you to Dr. Cheryl Keen for your feedback and support in this dissertation writing experience. Lastly, I would like to thank my chair, Dr. John Harrison, committee member, Dr. Sherry Lowrance, and URR, Dr. Marilyn Robb, for the guidance and feedback they provided me throughout this arduous process.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Introduction

If a goal of teacher evaluation is to improve teacher practice, then it is necessary to develop tools and strategies that support educators to engage in reflective practice during the teacher evaluation process. According to Dewey (1910), examining the foundation for beliefs and practices is called "reflective thought" and "it alone is truly educative in value" (p.2). Shulman (1986) argued that teachers' reflective awareness to strategically apply content knowledge, pedagogical knowledge, and curricular knowledge influences their effectiveness. Shulman (1986) noted that "the ultimate test of understanding rests on the ability to transform one's knowledge into teaching" (p. 13), which was referred to as pedagogical content knowledge (PCK). Schön (1983) believed that professionals may engage in reflective practice during or after experiences, providing opportunities for learning that can influence future efforts. Meierdirk (2016) stated "It is the cognitive processes of the teachers themselves which leads to professional development; this is achieved through reflective practice" (p. 375). Therefore, reflective practice is a critical element for teachers' continuous improvement efforts and teacher evaluation systems should be designed to foster reflective practice and support such improvement.

Many states require school districts to incorporate student growth data in the teacher evaluation system. Districts typically comply with this requirement by choosing to use a statistical model, such as value-added measurement (VAM), or a goal-setting model, such as student learning objectives (SLOs). Numerous reserachers have examined

the strengths and weaknesses of these models (Amrein-Beardsley & Holloway, 2017; Pivovarova & Amrein-Beardsley, 2018; Plecki et al., 2016). Researchers have found that using statistical models to evaluate student growth does not promote improvement in educator practices (Amrein-Beardsley & Holloway, 2017; Garet et al., 2017). In studies of SLO implementation, researchers have shown diverse interpretations of the SLO process and noted a need for structures to support implementation (Crouse et al., 2016; Plecki et al., 2016).

A major challenge to incorporating student growth into teacher evaluation systems is the lack of consistency in the interpretation of student growth. Although some educators may look at student responses with a dichotomous view of right or wrong answers, assessment experts encourage an examination of the level of thinking students exhibit (Alonzo, 2018; Gotwals, 2018; Pellegrino & Chudowsky, 2003). Some researchers advised the use of learning progressions for monitoring student growth (Alonzo, 2018; Black et al., 2011; Briggs & Peck, 2015; Fonger et al., 2018; Popham, 2008). Researchers have noted strong connections between monitoring student learning on learning progressions and formative assessment practices (Alonzo, 2018; Furtak et al., 2018; Gotwals, 2018). Some researchers have recommended that educators use a cognitive framework to guide development and interpretation of learning progressions (Black et al., 2011; Gagani & Misa, 2017; Gotwals, 2018).

Because statistical models based on standardized tests have not been found to support improvement in teacher practice and goal-based models based on classroom assessments have not been found to provide structure and consistency in monitoring

growth, I explored how the introduction of standards-based rubrics that represent learning progressions influence teachers' reflective practice and PCK as an element of the teacher evaluation system. I considered whether and how educators perceived the standards-based rubrics to provide the structure and consistency lacking in goal-based models for monitoring student growth. Therefore, I explored the efficacy of a teacher evaluation system that uses standards-based rubrics as learning progressions to monitor student growth in an SLO process and how the standards-based rubric fosters teachers' reflective practice and PCK. The cognitive model used for the learning progressions represented in standards-based rubrics for this study was the structure of the observed learning outcome (SOLO) taxonomy.

Although legislative policies require the incorporation of student growth in many teacher evaluation systems, in most states school districts may choose their method for compliance with these legislative mandates. Due to the preponderance of evidence against test-based models and the lack of research regarding structures for goal-based models, a need exists for research exploring whether teacher evaluation systems that enact standards-based rubrics as the structure for monitoring student growth support teachers' reflective practice and PCK. Findings from this study may inform development and monitoring of teacher evaluation systems at local and state levels.

In this chapter, I provide the background for this study. Following the background, I describe the problem statement, purpose, research questions, and the conceptual framework of the study. This chapter also includes the nature of the study,

definitions, assumptions, scope and delimitations, limitations, and significance of the study.

Background

The U.S. Department of Education (2012a, 2012b) encouraged states to pass legislation regarding the incorporation of student performance as an element of teacher evaluation systems by developing a waiver program for the No Child Left Behind Act of 2001 (NCLB). To be granted flexibility from requirements of NCLB, every state education association needed to "incorporate student growth into its performance-level definitions" (U.S. Department of Education, 2012b, p. 20). Also, due to the implementation of the Race to the Top initiative, many states felt incentivized to pass legislation requiring revision of teacher evaluation systems to incorporate student growth data (Munroe, 2017). The National Council on Teacher Quality reported that 39 states in the United States require school districts to include student growth data as an element of teacher evaluation systems (Walsh et al., 2017). Of those states, only one state required student growth to be the determinative factor in a teacher's overall rating. In 2015, the federal government passed the Every Student Succeeds Act (ESSA). Under the reauthorization of the Elementary and Secondary Education Act of 1965, states and local school districts were granted greater flexibility in determining processes for teacher evaluation (U.S. Department of Education, 2017).

Two methods are typically used to incorporate student data in the teacher evaluation process: statistical models, such as VAM and student growth percentiles

(SGP), or goal-based models such as SLOs (McCullough, et al., 2015; Measured Progress, 2014). Statistical models are based on standardized test scores, whereas goal-based models can be based on classroom assessments. Marion et al. (2012) advised that the use of SLOs for monitoring student performance in the context of teacher evaluation systems has promise for promoting both student learning and educational improvement. Marion et al. emphasized the importance of embedding assessment within the system as a status-based focus rather than looking at evaluation of student success as gain-based. Using an SLO process requires setting learning targets based on baseline data and monitoring and reporting student progress toward those performance targets. The researchers noted that the professional development needed to implement SLOs is consistent with the professional development needed to implement the newest generation of standards (Marion et al., 2012).

Most districts that use an SLO process have developed an SLO template for teachers to use to document the experience. This document typically delineates a select group of standards that serve as the focus of assessment and instruction for an agreed on time. The document also details assessment tools that teachers have selected or developed for use in establishing baseline data (preassessments), monitoring students' progress during instruction (formative or interim assessments), and measuring students' performance levels at the end of the chosen time frame (summative assessments). Once teachers collect baseline data, they use the data to set learning targets for their students (Center for Assessment, 2017). Thus, the SLOs provide the student growth framework in the goal-based system.

The status levels of an SLO should represent learning progressions inherent in the content being monitored (Briggs et al., 2015). The concept of growth in this study incorporated the SOLO taxonomy (Biggs & Collis, 1982) as the cognitive model for the learning progression framework (Black et al., 2011) that is structured in a standards-based rubric. Alonzo (2018) and Black et al. (2011) asserted that formative and summative assessment practices should be used to monitor student learning and that learning progressions are a critical feature of formative assessment.

Rubrics provide a framework for monitoring growth along a learning progression. Brown et al. (2014) asserted that rubrics are the most promising method for monitoring the critical-analytic thinking called for in the standards. Bowen (2017), İlhan and Çetin (2016), and Rembach and Dison (2016) all supported the use of the SOLO taxonomy as a framework for rubric design to measure complex thinking. Both Popham (2013) and Darling-Hammond et al. (2012) advocated for the use of classroom assessments to monitor student growth for teacher evaluation. In rubric based SLO, educators align classroom assessments with the standards-based rubric.

In many states, changes to teacher evaluation systems occurred during a time of transition to updated standards for most content areas. Many states adopted new standards for mathematics and English language arts in 2010, based on the Common Core State Standards (CCSS; Council of Chief State School Officers [CCSSO] & National Governor's Association [NGA], 2010). Some adopted new standards for science in 2013 based on the Next Generation Science Standards (National Association of State Boards of Education, 2016). Some adopted new social science standards that stemmed from the

College, Career, and Civic Life framework, released in 2013 (National Council for the Social Studies, 2018). Standards for dance, media arts, music, theater, and visual arts have been revised in some states based on the National Core Arts Standards (American Alliance for Theater & Education, 2018; National Art Education Association, 2018; National Association for Music Education, 2018). Some states also updated physical education expectations following the release of the Adaptive Physical Education Standards (Shape America, 2018).

According to Earl and Ussher (2016), "reflective practice and inquiry are aspects of teacher professional practice that characterize teachers as learners" (p. 47). Russell (2018) noted that reflective practice involves learning from professional experience rather than in professional classes or written assignments for courses. Russell argued that teachers alter practice as a result of reflective thinking. Zwozdiak-Myers (2018) argued that reflective thinking is necessary for teachers to transform knowledge into meaningful learning experiences for students. I examined how the use of standards-based rubrics to structure the monitoring of student growth in an SLO process for teacher evaluation fosters teachers' reflective practice and PCK in mathematics. If teacher evaluation systems are meant to improve teachers' understanding and practices, then an exploration of whether and how SLOs align to a standards-based rubric in the context of teacher evaluations was necessary to ascertain if it supported teachers' professional growth.

In recent studies involving reflective practice in the context of teacher evaluation, researchers examined professional practice as opposed to PCK. Most researchers who examined reflective practice and PCK involved pre-service teachers or educators in

higher education settings (Coon-Kitt et al., 2015; Gabriel, 2017; Olteanu, 2017; Reilly, 2018). In two studies, researchers addressed teachers' reflective practice regarding PCK. Estaji and Dezfoolian (2018) found a significant relationship between PCK and reflectivity. Park and Oliver (2008) found that "PCK development occurred as a result of reflection related to both knowledge-in-action and knowledge-on-action" and "teachers understanding of students' misconceptions was a major factor that shaped PCK in planning, conducting instruction, and assessment" (p. 268).

Papay (2012) argued that teacher evaluation systems can be both summative and formative. In the summative sense, the measurement instruments are used to assess teacher effectiveness. Formatively, however, evaluations "provide valuable information to drive professional growth and, as such, can raise teacher effectiveness" (Papay, 2012, p. 124). Both Kraft and Gilmour (2016) and Malunda et al. (2016) found that the evaluation process can promote teacher development. Teachers have reported that reflecting on student work can enrich their own capacities for assessment and instruction (Darling-Hammond, 2016). In contrast, Garet et al., (2017) found that the use of VAM for feedback on student growth had no impact on student achievement in English language arts (ELA) and minimal impact in mathematics. Garet et al. also found that VAM feedback did not influence teachers' interest in improving practice. Firestone and Donaldson (2019) found that teachers and evaluators struggled to analyze assessment data and use it for improving instruction and student learning.

Many researchers have examined the impact of statistical measures on teacher practice and desire to improve instruction, but limited research has been done regarding

the impact of SLOs on professional learning. Although SLOs use has increased, multiple researchers have noted that the interpretation of the SLO process varies from state-to-state and district-to-district (Crouse et al., 2016; Joyce et al, 2016; Longo-Schmid, 2016; Makkonen et al., 2015; Marion, 2015; McCullough et al., 2015; Plecki et al., 2016; Slotnik, et al., 2015). Although researchers have recommended that student performance be measured using rubrics and aligned to learning progressions, researchers have not examined whether and how teacher evaluation systems that use standards-based rubrics or learning progressions promote reflective practice and PCK. If a purpose for teacher evaluation is to encourage improvement in PCK, and researchers have purported that SLO have promise in promoting such improvement, then examination of teacher experiences using SLO structured around standards-based rubrics merits investigation.

Problem Statement

The problem addressed in this study was that little is known about how using a standards-based rubric to structure SLOs in the teacher evaluation process supports teachers' reflective practice and PCK. Haertel (2013) advised against using statistical measures in teacher evaluation systems. Instead, Marion et al. (2012) recommended that school districts use an SLO process as a method for monitoring student growth. The researchers also argued that SLOs showed promise in improving teacher practice. The Center for Assessment (2017) recommended the incorporation of rubrics for monitoring student growth. Brown et al. (2014) asserted that rubrics are the most appropriate structure for assessment of higher-level thinking required for the newest generation of standards, including CCSS and Next Generation Science Standards. Popham (2013)

asserted that classroom assessment can and should be used to monitor student growth in teacher evaluation systems.

Researchers examining teacher evaluation systems that include student growth measures found no evidence that statistical models, such as VAM or student growth percentiles, promote reflective practice or PCK (Amrein-Beardsley & Holloway, 2017) and goal-based models, such as SLO, do not provide sufficient structure without extensive training to promote PCK (Crouse et al., 2016). No researchers examined teachers' reflective practice regarding their content knowledge and PCK when standards-based rubrics are used as the structure for the SLOs to guide monitoring of student growth in teacher evaluation. Briggs et al. (2015) recommended that states and districts use a learning progression framework in the design of their SLO systems; however, I found no studies in which researchers examined how teachers perceive the experience of implementing SLOs structured by standards-based rubrics as learning progressions.

As educators grapple with the simultaneous implementation of new standards and accountability systems, the findings from this study may inform state and district practices in designing systems for monitoring student learning on standards and guiding teachers to continuously improve implementation of curriculum, assessment, and instruction aligned to standards. The results of this study may also inform state and district design of teacher evaluation systems that include evidence of student growth. Because researchers studying SLOs have identified a lack of structure as a major challenge to their successful implementation, I explored whether and how educators

perceive standards-based rubrics to provide structure for using classroom assessments in an SLO process to promote educators' reflective practice and PCK.

Purpose of the Study

The purpose of this basic qualitative inquiry was to explore how teachers perceived experiences using a standards-based rubric to structure SLOs in the teacher evaluation process supported their reflective practice and PCK. The teacher evaluation systems in this study required educators to monitor student growth using an SLO process structured by standards-based rubrics and incorporating classroom assessments. Most states that require districts to include student growth as a component of their teacher evaluation systems allow districts to use SLOs as a method for monitoring student growth. Student growth represents "changes in student performance across at least two points in time" (Hewitt & Amrein-Beardsley, 2016, p. 10). According to the Illinois State Board of Education (ISBE) Student Learning Objective Guidebook (2015), "when implemented with fidelity, the SLO process benefits students and teachers by supporting collaboration and reflective teaching practices" (p. 4).

Research Questions

The research questions and sub-questions guiding this study were:

- RQ1: How do teachers perceive the experience of implementing SLOs structured by standards-based rubrics to support reflective practice?
- RQ2: How do teachers perceive the experience of implementing SLOs structured by standards-based rubrics to support PCK?

- SQ1: In what ways do teachers reflect on and adapt their mathematical content knowledge as they implement SLOs structured by standards-based rubrics?
- SQ2: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLOs structured by standards-based rubrics?
- SQ3: In what ways do teachers reflect on and adapt their instructional tools and practices as they implement SLOs structured by standards-based rubrics?

Conceptual Framework for the Study

The conceptual framework for this study incorporated two lenses: reflective practice (Schön, 1983) and PCK (Shulman, 1986), both of which stem from the work of John Dewey. As a promoter of constructivism, Dewey (1938) believed that learning can come from experience, but not all learning is intended or desirable. Dewey (1938) noted that experiences do not necessarily lead to intellectual growth, as misconceptions may form based on experiences and routine learning can be done without thinking:

> There is no intellectual growth without some reconstruction, some remaking, of impulses and desires in the form in which they first show themselves. This remaking involves inhibition of impulse in its first estate. The alternative to externally imposed inhibition is inhibition through an individual's own reflection and judgment. The old phrase "stop and think" is sound psychology. For thinking is stoppage of the immediate manifestation of impulse until that impulse has been

> brought into connection with other possible tendencies to action so that a more comprehensive and coherent plan of activity is formed (p. 64).

Schön (1983) believed that professionals might engage in reflective practice during or after experiences, providing opportunities for learning that can influence future efforts. Therefore, reflective practice is a necessary element for teachers' continuous improvement efforts. If a goal of teacher evaluation is to improve teacher practice, then it is necessary to understand what tools and practices support teachers to engage in reflective practice during the teacher evaluation process. In particular, I examined the effectiveness of structuring SLOs around a standards-based rubric and how this rubric supports reflective practice and PCK in the context of teacher evaluation. Schön (1983) noted that professionals practice reflection-in-action as a continuous effort to improve their craft. The act of reflection provides the opportunity to examine both effective and ineffective strategies. A professional considers why a strategy is effective in one context, but ineffective in another. If one can identify factors that contribute to the successful implementation of an assessment or instructional method, this information can be used to replicate the success in other contexts. In this context, participants were asked to reflect on their practices for design; administration, and interpretation of assessments; their planning and implementation of instruction practices, and their understanding and implementation of standards based on their use of standards-based rubrics to monitor student growth. Participants were also asked to reflect on their interpretation of student growth in relation to the learning progressions articulated in the standards. The research

questions in this study explore participants' reflections on their knowledge and experiences.

The second framework, PCK, is derived from the work of Lee Shulman (1986), who argued that teachers' reflective awareness to strategically apply content knowledge, pedagogical knowledge, and curricular knowledge influences their effectiveness. Shulman (1986) noted, "the ultimate test of understanding rests on the ability to transform one's knowledge into teaching" (p. 13), referred to as PCK. Shulman contended that educators develop knowledge of content and knowledge of pedagogy and blending these two aspects into knowledge of pedagogy for content into PCK exemplifies the professionalism of teachers. Teachers use their reflective awareness to make judgments about professional practice. Shulman (1987) also noted that "critical features of teaching, such as the subject matter being taught, the classroom context, the physical and psychological characteristics of the students, or the accomplishment of purposes not readily assessed on standardized tests, are typically ignored in the quest for general principles of effective teaching" (p. 6). Knowledge of learning progressions articulated in the standards (RQ 1, SQ 1), understanding of various instructional models and strategies (SQ 3) for targeted content, and assessment literacy (SQ 2) are examples of PCK. A more thorough description of the conceptual framework can be found in Chapter 2.

Nature of the Study

I used a basic qualitative approach for this study. According to Merriam and Tisdell (2016), qualitative researchers examine "how people interpret their experiences, how they construct their worlds, and what meaning they attribute to their experiences" (p.

6). A basic qualitative approach was consistent with this study because I was addressing participant perceptions (Patton, 2015). I sought to understand how teachers perceived using a standards-based rubric to structure SLOs in the teacher evaluation process supported their reflective practice and PCK.

To collect data for this study, I conducted interviews with teachers who have used standards-based rubrics to plan and implement mathematics SLOs for teacher evaluation. These interviews occurred via the internet using Zoom. I conducted a survey with open-ended questions to gather information about teachers' SLO goal choices, baseline data collection strategies, target-setting methods, and demographics. I then conducted semi structured, individual interviews regarding how teachers perceived the rubric to influence their reflective practice and PCK. The interview transcripts were then coded using qualitative content analysis (Elo et al., 2014) to identify trends and patterns. I used the program NVivo to manage the coding process. These data collection and analysis methods were an appropriate way to document and explore educators' goals and reflections on experiences (Rubin & Rubin, 2012).

Definitions

The following are concise definitions for constructs used in this study.

Assessment literacy: Competence and knowledge of fundamental assessment concepts and procedures (Popham, 2018).

Baseline data: Information regarding students' prior knowledge that is prerequisite to the chosen learning targets. Baseline data may also include demographic and perceptual information (Center for Assessment, 2017).

Common Core State Standards for Mathematics (CCSS-M): The standards developed by a team of researchers and educators assembled by the CCSSO and NGA (2010).

Complexity: The progression of thinking that learners go through to deepen their understanding of ideas. The model for complexity in this study is the SOLO taxonomy, which represents the progression in five levels: (a) pre-structural, (b) unistructural, (c) multi-structural, (d) relational, and (e) extended abstract (Biggs & Collis, 1982).

Evaluator: An administrator or other individual assigned to guide teachers to continuously improve their practice and provide a summative rating for the quality of teacher effectiveness in the context of a teacher evaluation system. In some states, evaluators must attend and pass a certification training (Performance Evaluation Reform Act [PERA], 2010).

Formative assessment: The process of gathering and interpreting information as feedback to adjust teaching and learning (Black & Wiliam, 2010, Heritage, 2010; Popham, 2008).

Growth target: The performance level to which one aspires to perform. Growth targets may be set at group or individual levels (Center for Assessment, 2017).

Learning progression: A common “road map” for students to learn concepts, procedures, skills, and applications of learning goals (Black et al.,

2011). The progression represents a developmental sequence that addresses increases in difficulty and complexity (Briggs & Peck, 2015; Popham, 2007).

Learning target: The specific learning intention for a lesson, that represents the next level of development, as students advance through a learning progression. Learning targets are presented from the learners' point of view to guide them in monitoring their own learning (Andrade, 2013).

Pedagogical content knowledge (PCK): Teachers understanding of content and enactment of how to support students in developing understanding of content that is particular to the subject area (Park & Oliver, 2008; Shulman, 1986). PCK includes knowledge of learning progressions inherent in the standards and the instructional and assessment processes associated with guiding student progress along learning progressions.

Preassessment: Measurement of students' knowledge, skills, understandings and/or dispositions prior to instruction on the topic of interest. (Hockett & Doubet, 2014).

Reflection-in-action: A practitioner considers circumstances and makes decisions while engaged in the task at hand (Schön, 1983).

Reflection-on-action: A practitioner considers circumstances after a task and makes decisions for future practice (Schön, 1983). "To reflect is to look back over what has been done so as to extract the net meanings, which are the capital stock for intelligent dealing with further experiences" (Dewey, 1938, p. 87).

Reflective practice: The act of considering events and circumstances to inform decisions for future actions (Dewey, 1910; Schön, 1983).

Standards-based rubric: A rubric designed to represent the learning progression of a standard or standards by which student performance on classroom assessments can be evaluated.

Student growth: Demonstration of a change in students' understanding between 2 or more points in time (Hewitt & Amrein-Beardsley, 2016). Growth is shown by comparing assessment evidence representing a lower level of a learning progression to evidence of performance at a higher level of the learning progression.

Student learning objectives or Student learning outcomes: The structure in which teachers gather, organize, and analyze evidence of student growth using multiple assessments over a specified time period in the context of teacher evaluation systems (ISBE, 2015, p. 4).

Summative assessment: Assessment evidence collected at the end of an instructional timeframe that is used to make judgments regarding students' mastery of learning targets (Andrade, 2013).

Teacher evaluation: The process of gathering evidence of teacher effectiveness through observation, data analysis, and dialogue. Formative teacher evaluation is intended to improve teacher practice, whereas the purpose of summative evaluation is to render a judgment, typically for employment determinations (Papay, 2012).

Assumptions

Because data were collected through qualitative interviews, one assumption for this study was that participants were truthful in their responses. Another assumption was that participants have used standards-based rubrics to monitor student learning on the chosen standards of the SLOs. I also assumed that evaluators met with participants to discuss the chosen standards, performance level rubric, assessments, and student data at the beginning, middle, and end of the SLO process. Finally, I assumed that participants had some control over their classroom assessment and instructional planning practices. These assumptions were necessary to support validity of the data, as teachers who have not been truthful or have not engaged in the SLO process as expected would be unable to share experiences regarding their reflections on these experiences. Also, if the teachers were forced to follow a scripted program and have no control over classroom assessment and instruction, their perceptions would not represent the process of using data to guide planning decisions.

Scope and Delimitations

In this research study, I addressed teacher evaluation systems for teachers in one midwestern state of the United States who have used standards-based rubrics in an SLO process to monitor student growth. This study was deliberately limited to elementary teachers (Grades K–8) of mathematics. The data only address teacher reflections regarding their mathematical PCK. I explored educator reflections on their content knowledge and pedagogical practices for assessing and instructing mathematics. I used purposeful sampling to select 10 participating teachers who teach mathematics to

students in kindergarten through eighth grade. Recruitment of participants included consideration of district demographics and setting, as every attempt was made to include participants from more than one setting. I collected demographic data regarding the setting (rural, suburban, and urban) and student demographics (English language learner, special education, low-income, etc.) to address the transferability of findings. Attention was also given to the teacher experience levels to support transferability of findings to apply to novice and experienced teachers.

Limitations

Because I only addressed PCK for mathematics, findings for this study may not be generalizable to the implementation of SLOs for content areas other than mathematics. Purposeful sampling was also a limitation of this study, as participants must have experienced using standards-based rubrics to monitor student learning in the context of SLOs as the student growth component within constraints of the district teacher evaluation system. Participating school districts must have allowed SLOs as a structure for monitoring student growth for teacher evaluation purposes. Participants joined this study voluntarily.

As noted in the assumptions, teachers involved in this study had some control over classroom instruction and assessment decisions. If teachers were required to exclusively follow a scripted or structured mathematics program, they were not considered viable participants for this study. Teachers who are unable to alter practices after reflecting in action or reflecting on action would be less able to share how reflections supported growth in PCK and could potentially bias the outcome of the study.

Because the study was purposefully limited to teachers who fit the profile, nothing additional was done to address these limitations.

Significance

Marion et al. (2012) recommended that school districts use an SLO process as a method for monitoring student growth. The researchers asserted that this process shows promise in improving teacher practice (Marion et al., 2012). The Center for Assessment (2017) recommended the incorporation of rubrics for monitoring student growth. Brown et al. (2014) asserted that rubrics are the most appropriate structure for assessment of higher-level thinking required to meet the newest generation of standards, including CCSS and Next Generation Science Standards. Although studies have been conducted regarding student growth scores and the implementation of SLOs (Gill et al., 2013; Makkonen et al., 2015; McCullough et al., 2015; Measured Progress, 2014; Schmitt & Hutchins, 2015; Slotnik et al., 2014; Slotnik et al., 2015), I found no studies in which researchers examined how teachers reflect on the application of standards-based rubrics for monitoring student growth with regard to their own implementation of standards.

Although Briggs et al. (2015) recommended that states and districts use a learning progression framework in the design of their SLO systems, I found no studies in which researchers examined how teachers reflect on the use of standards-based rubrics as learning progressions for CCSS-M. This study has potential implications for positive social change because the results may inform district and state policy makers regarding the influence of formative teacher evaluation systems on teachers' PCK. The results of this study may provide information concerning teacher evaluation systems for

improvement of teacher practice at the individual, team, or school levels. As educators grapple with the simultaneous implementation of standards and accountability systems, findings from this study can inform state and district practices for both the monitoring of student learning on standards and the alignment of assessment and instruction to standards. In this context, PCK includes teachers' knowledge of standards-based assessment and instructional practices.

The results of this study can also inform state and district design of teacher evaluation systems that require the inclusion of evidence of student growth. Because Hill et al. (2005) found that teachers' PCK relates to student achievement gains, the examination of a teacher evaluation structure that supports teachers' PCK has the potential to support gains in student achievement. Thus, the study has the potential to influence how teacher evaluation systems can be structured to improve teacher practices that can lead to improvements in student learning.

One of the primary responsibilities of teachers is to monitor student learning. With the newest generation of standards calling for critical analytic thinking, teachers and their evaluators must design systems for monitoring standards-based growth. However, research has shown that statistical measures of student growth do not enhance educator practices and are not necessarily standards-based (Amrein-Beardsley & Holloway, 2017; Garet et al., 2017). Goal-based systems have been found lacking in structure to support implementation (Crouse et al., 2016; Plecki et al., 2016). Although Popham (2013) argued for the use of classroom assessments to monitor student learning for teacher evaluation, many districts struggle with implementation of this recommendation in a

goal-based system and, therefore, use test-based systems as an alternative. The findings from this study can influence how educators monitor standards-based growth and can guide school districts to develop effective systems for supporting educators' continuous improvement in reflective practice, PCK, standards-based assessment, standards-based instruction, and teacher evaluation. Improvement in any one of these areas has the potential to improve student learning. Understanding what teacher evaluation practices promote teachers' reflective practice and PCK can lead to positive social change because supportive teacher evaluation systems have been found to lead to positive changes in teacher practices (Ford et al., 2018; Robertson-Kraft & Zhang, 2018), increase teacher knowledge and performance (Darling-Hammond, 2016), and support school improvement initiatives (Coburn, et al., 2016).

Summary

I explored whether and how structuring an SLO process for monitoring student growth with standards-based rubrics supported teachers' reflective practice and PCK. Because teacher evaluation systems can serve both formative and summative purposes, this research was conducted to identify trends in educators' perceptions of the SLO process when introducing standards-based rubrics into the teacher evaluation system. For teachers to improve their practice, they must engage in reflective thinking (Dewey, 1910, 1933/1998). In this chapter, I provided the background and structure for the research study. I included the research questions, a brief introduction to the conceptual framework, assumptions, scope, limitations, delimitations, and significance of this study. Chapter 2 includes more detailed information regarding policies and research relative to educators'

reflections and experiences with the implementation of standards-based rubrics in teacher evaluation systems.

Chapter 2: Literature Review

Introduction

In this research study, I explored how using standards-based rubrics to monitor student growth may influence educators' reflective practice and PCK regarding standards, assessment, and instruction. The requirement to incorporate student growth measures on newly adopted state standards into teacher evaluation systems has left many teachers feeling confused, overwhelmed, frustrated, and anxious. Research on teacher evaluation systems indicates that teacher evaluation can serve both formative and summative purposes (Papay, 2012). Malunda et al. (2016) found that formative evaluation yielded greater increases in quality of pedagogical practices than summative evaluation. Research has indicated that formative evaluation can lead to improvement at the individual (Darling-Hammond, 2016), group (Derrington & Kirk, 2017), and organizational levels (Johnson, 2015). Gotwals (2018) found that monitoring student growth along learning progressions in teacher evaluation systems can enhance teacher practices. Enderson et al. (2018) found that teachers who engage in reflective practice can improve their content knowledge, and Camburn and Han (2017) found that reflective practice can improve teachers' PCK.

Researchers have found that many current growth monitoring practices for teacher evaluation need improvement to support teacher and student growth. Wilson and Downs (2014) and Zhang (2014) identified the need for efficient structures and strategies to monitor student learning that can positively impact educators' PCK. Amrein-Beardsley and Holloway (2017), Garet et al. (2017), and Haertel (2013) all found that statistical

models do not support reflective practice or teachers' PCK development. Crouse et al. (2016), Plecki et al. (2016), and Slotnik et al. (2015) all identified the need for structures to support the SLO process. I examined how teacher evaluation systems in which SLO use standards-based rubrics and classroom assessments to monitor student growth promote teachers' reflective practice and PCK.

In this chapter, I present the findings from a review of articles, legislation, and texts around the topics related to teacher evaluation systems that use standards-based rubrics as the structure for an SLO process to monitor student learning. The review begins with a description of the conceptual model for this study. The chapter continues with a discussion of legislative decisions that have influenced both teacher evaluation and standards implementation and the topic of teacher evaluation with a focus on how the evaluation process relates to accountability and the development of teachers' PCK.

The review also addresses research on SLO to monitor student growth. Briggs et al. (2015) promoted the alignment of SLO to a learning progression framework, which merits an investigation of literature around learning progressions. The review included the application of learning progressions to supporting formative assessment practices. The cognitive framework for the learning progressions and rubrics in this study was the SOLO taxonomy, which is another topic explored in the literature. Because the learning progressions in this study are in the form of standards-based rubrics, I sought research on standards-based rubrics. Finding no studies on this topic, the review was limited to learning progressions. This chapter also contains a description of literature search

strategies, conceptual frameworks (reflective practice and PCK), and the research methodology (basic qualitative inquiry).

Literature Search Strategy

I searched several databases to conduct an extensive review of the literature. These included Academic Search Complete, Education Source, ERIC, Research Starters – Education, SAGE Journals, Science Direct, Taylor and Francis Online, and Teachers Reference Center. Also, I used Google Scholar to locate sources. A search *of rubric based SLO* or *standards-based rubrics* and *SLO* yielded no relevant sources. Therefore, searches of *rubrics and SLO, rubrics and student growth, rubrics, teacher evaluation,* and *formative assessment, rubrics and reflection,* and *rubrics and learning progressions* were necessary to locate relevant sources. Searches for *pedagogical content knowledge and teacher evaluation* and *reflective practice and teacher evaluation* yielded sources that primarily addressed pre-service teachers or professional practice of practicing teachers. Some of the studies addressed measuring teachers' PCK or teachers' experiences with reflective practice, but none examined whether of student growth monitoring structures in teacher evaluation systems promote reflective practice and PCK.

Secondary searches included combinations using key words such as *standards implementation*, *Common Core State Standards*, *teacher evaluation, student growth, student learning objectives, student learning outcomes, SLO, standards-based, rubrics, formative assessment, learning progression(s), pedagogical content knowledge, reflective practice, structure of the observed learning outcome,* and *SOLO taxonomy.* These searches yielded many studies addressing topics related to this study. For example,

searches for *student growth* and *teacher evaluation* yielded multiple sources involving policies, findings, and perceptions regarding statistical test-based models. Some sources also addressed goal-based models, but no sources examined a structure for using classroom assessments to monitor student growth. Relevant trends, findings, and recommendations from this literature search are described in this chapter.

Conceptual Framework

Reflective Practice

Dewey (1933/1998) explained reflective thinking as "the kind of thinking that consists in turning a subject over in the mind and giving it serious and consecutive consideration" (p. 3). Dewey (1933/1998) believed that reflection is founded on a belief in evidence and enables goal-oriented planning. Schön (1983) expanded Dewey's ideas by connecting reflective thinking to professionalism. Schön (1983) noted that "a professional practitioner is a specialist who encounters certain types of situations again and again" (p. 60). Both Dewey (1916/1998) and Schön (1983) cautioned that routine experiences could lead a practitioner to miss opportunities to think about actions. Therefore, Schön argued for professions engaging in reflective practice to prevent habitual behavior overtaking thoughtful action.

> A practitioner's reflection can serve as a corrective to overlearning. Through reflection, he can surface and criticize the tacit understandings that have grown up around the repetitive experiences of a specialized practice and can make new sense of the situations of uncertainty or uniqueness which he may allow himself to experience. (Schön, 1983, p. 61)

Dewey (1938) also argued that reflection could lead to professional growth for teachers. Rogers (2002) agreed, noting that "thinking, particularly reflective thinking or inquiry, is essential to both teachers' and students' learning" (p. 842). Rogers (2002) also identified four criteria that represent Dewey's concept and purposes for reflection:

> 1. Reflection is a meaning-making process that moves a learner from one experience into the next with deeper understanding of its relationships with and connections to other experiences and ideas. It is the thread that makes continuity of learning possible, and ensures the progress of the individual and, ultimately, society. It is a means to essentially moral ends.
> 2. Reflection is a systematic, rigorous, disciplined way of thinking, with its roots in scientific inquiry.
> 3. Reflection needs to happen in community, in interaction with others.
> 4. Reflection requires attitudes that value the personal and intellectual growth of oneself and of others (Rogers, 2002, p. 845).

Darling-Hammond (2006) mentioned reflection as a practice of effective teachers stating that teachers need to "reflect on their practice to learn from and improve it continually" (p. 300). Darling-Hammond stressed the importance of guiding teachers to synthesize different types of knowledge: (a) knowledge of learners and their development in social contexts, (b) knowledge of subject matter and curriculum goals, and (c) knowledge of teaching. Darling-Hammond (2006) acknowledged this challenge, stating "teachers need to know how and when to use a range of practices to accomplish their goals with different students in different contexts" (p. 304).

Researchers have found that reflective practice leads to changes in teacher practice (Bell & Mladenovic, 2015; Camburn & Han, 2015, 2017; Farrell & Vos, 2018; Russell, 2018). Mezirow (1997) argued that "self-reflection can lead to significant personal transformations" (p. 7). Haj Sassi (2016) agreed but noted that teachers identified a need for a system in which to apply self-observation strategies. Griggs et al. (2018) found that educators could learn reflective skills but had difficulty transferring those skills into working practice. Camburn and Han (2015, 2017) found reflective practice most successful in embedded learning activities. The researchers noted that reflective practice about school or district-wide goals was far less impactful than when focused directly on classroom instruction (Camburn & Han, 2015, 2017).

Pedagogical Content Knowledge

Shulman (1987) distinguished between content knowledge and pedagogical knowledge. However, Shulman (1986) noted that "the key to distinguishing the knowledge base of teaching lies at the intersection of content and pedagogy" (p.9). PCK was defined as "subject matter knowledge for teaching" referring to the "particular form of content knowledge that embodies the aspects of content most germane to its teachability" (Shulman, 1986, p. 9).

Both Shulman (1987) and Dewey (1904) agreed that teachers could grow in content knowledge and pedagogical knowledge. Dewey (1904) stated, "even though they go on studying books of pedagogy, reading teachers' journals, attending teachers' institutes, etc., yet the root of the matter is not in them unless they continue to be students of subject-matter, and students of mind-activity" (p. 15). Shulman (1986) concurred:

> Mere content knowledge is likely to be as useless pedagogically as content-free skill. But to blend properly the two aspects of a teacher's capacities requires that we pay as much attention to the content aspects of teaching as we have recently devoted to the elements of teaching process (p. 8).

In addition to content knowledge and PCK, Shulman (1986) added a third category of content knowledge for educators: curricular knowledge, described as knowledge of the progression of topics and the variety of available materials that can be used in instruction. This definition includes the "set of characteristics that serve as both the indications and contraindications for the use of particular curriculum or program materials in particular circumstances" (Shulman, 1986, p. 10). Shulman (1986) and Schön (1983) agreed that reflective practice separates professionalism from mere craft or skill. Shulman (1986) stated:

> The teacher is capable of reflection leading to self-knowledge, the metacognitive awareness that distinguishes draftsman from architect, bookkeeper from auditor. A professional is capable not only of practicing and understanding his or her craft, but of communicating the reasons for professional decisions and actions to others.
>
> This sort of reflective awareness of how and why one performs complicates rather than simplifies action and renders it less predictable and regular. (p. 13)

Some researchers have specifically examined PCK in mathematics. Matthews (2017) described how the concept of PCK influenced research on teacher knowledge in mathematics. Matthews cited Carpenter et al.'s (1996) work with cognitively guided

instruction, which examined teacher knowledge of the development of students' mathematical thinking. Matthews also recognized the work of Ball (1997), who built on Shulman's (1986) work to include teachers' knowledge of students under PCK. In support, Hill et al. (2005) found a significant relationship between teachers' mathematical knowledge for teaching and student achievement gains.

Enderson et al. (2018) found that when educators "possessed inadequate content understanding, they were not well positioned to understand and make accurate interpretations about mathematical understanding of student work" (p. 624). Thus, a lack of content knowledge can impair professional judgment. Conversely, educators possessing sufficient content knowledge were able to "more accurately predict student approaches to understand student thinking, and to plan for promising intervention in the event when the misconception emerged" (Enderson, et al., 2018, p. 624). Participants in Enderson et al. (2018) engaged in reflection on action through written journal entries, noting that analyzing student work helped participants better understand student thinking and informed their instructional planning.

Other researchers have examined PCK concerning assessment literacy. In their pilot study, Chapman and Koh (2017) engaged preservice teachers in using authentic assessment learning activities designed using the SOLO taxonomy as the cognitive framework. Participants enhanced their understanding of authentic assessment in mathematics by "making sense of selecting, unpacking, adapting, and designing authentic tasks" (Chapman & Koh, 2017, p. 959). Researchers noted the potential for supporting teachers' development of content knowledge and PCK through an assessment focus.

Lang et al. (2014) engaged primary mathematics teachers in a formative assessment system initiative aligned with CCSS that involved analysis of student work. Educators also analyzed formative assessment data to differentiate instruction. Participating teachers improved their mathematical knowledge for teaching, and their students made statistically significant gains. In contrast, Deneen and Brown (2016) found that teachers who took an assessment literacy course may have made gains in assessment literacy knowledge but did not change their conceptions regarding the purpose and nature of assessment or their assessment practices.

In this study, I focused on incorporating student growth into teacher evaluation systems, but I found that no studies specifically addressed how incorporation of student growth into teacher evaluation systems related to teachers' PCK. Although studies regarding statistical models for incorporating student growth into teacher evaluation systems have not found the practice to support teachers' PCK development (Amrein-Beardsley & Holloway, 2017; Garet et al. 2017), researchers have noted the promise of SLOs for supporting teachers' growth in PCK (Marion et al., 2012).

Relationship between Reflective Practice and Pedagogical Content Knowledge

Gillies (2016) agreed with Shulman (1986, 1987) and Schön (1983, 1987) that reflective practice has a place within teacher professionalism. Gillies (2016) identified the following strengths of reflective practice within the realm of teacher professionalism:

> It places 'thoughtful action' at the heart of teaching and so elevates the notion and importance of professional judgment; it provides the basis for rejecting the claims of technical rationalism and its twin risks of limiting teachers to a functional role

> and misrepresenting the contexts of teaching as invariable and so susceptible to a scientistic model; it reasserts the moral aspect of teaching in relation to the choice of virtuous ends and means; it enhances, and entrenches, the professionalism of teaching by seeing it as not something for which one can be merely 'trained' but rather as a practice where nuanced judgment is required; and, finally, it lends itself well to the current model of continuing professional learning, where reflection is seen as a crucial ingredient, from the novice to the expert levels, from the unpromoted to the most senior rank. (p. 150)

Although many studies regarding the development of reflective practice in the context of teacher professionalism involved preservice and novice teachers, Gillies (2016) noted that reflective practice is beneficial for educators at all levels of experience and rank. Gillies highlighted the connection between effective reflective practice and sound professional judgment, noting that judgment is a major factor elevating teaching to the level of professionalism.

Estaji and Dezfoolian (2018) found a significant relationship between teachers' pedagogical knowledge base and reflectivity. Park and Oliver (2008) included knowledge of students, curriculum, assessment, and instructional strategies and representations in their definition of PCK. Park and Oliver (2008) found that "PCK development occurred as a result of reflection related to both knowledge-in-action and knowledge-on-action" (p. 268). Park and Oliver (2008) also found that "teachers' understanding of students' misconceptions was a major factor that shaped PCK in planning, conducting instruction, and assessment" (p. 268). In addition, Park and Oliver (2008) found that students affected

teachers' PCK development when questions instigated reflection-in-action and when engagement or lack-of-engagement behaviors prompted reflection-on-action.

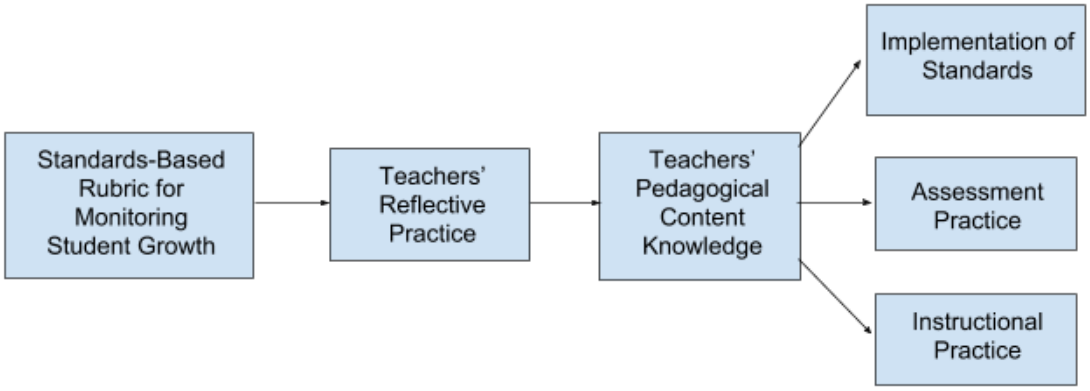
Multiple researchers that examined the impact of teachers analyzing student work found this strategy to be highly effective in promoting reflective practice (Coon-Kitt et al., 2016). Such practice led teachers to deepen their understanding of standards and students and to adapt instructional practices (Lalor et al., 2014). Gabriel (2017) and Kuh (2016) agreed that rubrics promote self-reflection and goal setting. Kuh (2016) identified a "need for rubrics and constructs that help groups define their endeavors, be intentional about focusing on children's work and teaching practices and develop an understanding of the developmental nature of adult learning" (p. 309). Busi and Jacobbe (2018) found that analysis of student work led teachers to increase their mathematical knowledge for teaching.

Thus, the dual conceptual frameworks for this study provide lenses to examine whether and how a teacher evaluation system that uses standards-based rubrics as learning progressions to monitor student growth fosters teachers' reflective practice and PCK. The researcher in this study explored how teachers reflect-in-action and reflect-on-action when using the standards-based rubrics and whether they perceived any impact on their PCK that influenced their implementation of standards, assessment practice, and/or instructional practice. The conceptual model for this study represents an examination of the efficacy of a teacher evaluation system in which standards-based rubrics are introduced as learning progressions to monitor student growth by gathering the

perceptions of teachers who experience SLO implementation based on standards-based rubrics (See Figure 1.).

Figure 1

Conceptual Model



Literature Review Related to Key Variables and/or Concepts

Teacher evaluation that supports teachers' reflective practice and PCK served as the context for this study. Both legislation and research have influenced state and district practices for the refinement of teacher evaluation systems to include student growth as a necessary component. Many researchers examined student growth from a quantitative perspective, identifying patterns of growth, without examining the qualities of systems that promote teacher development. Some researchers used mixed methods approaches to explore perceptions and conditions of teacher evaluation systems. In this study, learning progressions represented in standards-based rubrics provided the structure for each SLO. Therefore, the literature review encompassed legislation and research relating to standards implementation, teacher evaluation, SLO, learning progressions, and rubrics.

Reports and Policies Impacting Teacher Evaluation and Standards Implementation

Federal Legislation

At the federal level, the Elementary and Secondary Education Act of 1965 set the tone for government involvement in the education realm. The legislation represented an effort to promote continuous improvement of the nation's schools through funding for supplies and research in the field of education (Casalaspi, 2017). In 2002, George W. Bush signed the reauthorization of Elementary and Secondary Education Act into law, also known as the No Child Left Behind Act of 2001 (NCLB). The emphasis of NCLB shifted from school support to school accountability (Jacob, 2017). NCLB brought test-based accountability and sanctions to the forefront of educational discourse. To increase the number of effective teachers in hard-to-staff subjects, the U.S. Department of Education developed the Teacher Incentive Fund. The fund promoted the inclusion of student growth as an accountability measure by connecting performance pay to student growth (Humphrey et al., 2012).

Implementation of NCLB illuminated issues such as diversity of standards among states, lack of data regarding teacher effectiveness, and inequalities among low performing schools (Wong & Reilly, 2014). Under the Obama administration, states were offered the opportunity to apply for NCLB waivers to avoid sanctions (Croft et al., 2015). The waiver application process encouraged states to adopt the CCSS and to include student growth as an element of their teacher evaluation requirements (U.S. Department of Education, 2012a, 2012b).

The Elementary and Secondary Education Act was reauthorized again with the signing of the Every Student Succeeds Act (ESSA) of 2015. This legislation replaced NCLB, reducing federal control and involvement in educational oversight (Egalite et al., 2017). ESSA provides states and district leaders the opportunity to redefine teacher quality (Saultz et al., 2017). The new law required that districts use performance-based measures of teacher and principal quality. Fuller et al., (2017) noted that ESSA emphasizes the role of the principal in supporting teacher quality. Although ESSA does not explicitly state a requirement for teacher evaluation, states are required to "disclose the steps they're taking to evaluate and publicly report on the inequitable distribution of teachers and the qualifications of their teachers and school leaders" (Marion, 2016, p. 7). Therefore, federal legislation prompted many states to simultaneously adopt new standards and require the inclusion of student growth in teacher evaluation systems.

State Legislation

The Illinois General Assembly passed the PERA in 2010, requiring districts to evaluate teachers using both professional practice and student growth measures (PERA, 2010). An interim evaluation of the PERA implementation revealed that "communication is weaker on documenting and describing student growth processes compared to professional practice" (Milanowski et al., 2015, p. ix). The authors of the final report continued to emphasize the fact that the implementation of the student growth component presented a greater challenge compared to the professional practice element. They noted challenges regarding development of assessments, assessment literacy, investments in infrastructure and expertise, and data warehouses for teacher evaluation.

Illinois also adopted the CCSS for both English language arts and mathematics in 2010, with the expectation that school districts fully implement the standards by the 2013-2014 school year (ISBE, 2013). A 2014 survey of Illinois teachers regarding standards implementation revealed that only 17.5% felt completely prepared to implement the standards (ISBE, 2014). Teachers self-identified needs included "time to collaborate with colleagues" and "assistance in aligning assessments with Common Core units/lessons" (ISBE, 2014, p. 2). Thus, legislation presented the dual challenge for school districts to implement changes to both teacher evaluation systems and learning standards.

Standards Adoption and Implementation

In *Implementation of the Common Core State Standards: A transition guide for school level leaders*, the Aspen Institute (2013) identified seven indicators of a successful transition and provided descriptors of actions to be taken. Among these recommendations were suggestions to develop or adopt common expectations for what CCSS instruction looks like and design a CCSS-based assessment system. The researchers recommended that change leaders provide teachers training to translate data into CCSS-aligned instruction.

Barrett-Tatum and Smith (2018) structured their investigation of standards implementation around assumptions that stemmed from Loeb et al., (2008) examination of standards-based reform. These five assumptions addressed the need for teachers to develop a deep understanding of standards and the need for administrative support of

teachers' use of standards. They also noted the need for teachers to alter instructional practices and have access to professional development.

Perceptions of Standards Implementation

Since the CCSS were released in 2010, public perceptions have represented both positive and negative points of view. Pense et al. (2015) conducted a frame analysis of newspaper messages regarding the implementation of the CCSS representing the voices of learning experts, journalists, K-12 teachers, community members, politicians, and mixed sources. Researchers found that 47.8% of the messages expressed positive attitudes toward CCSS (p. 169). Supovitz and McGuinn (2017) noted that standards are not controversial, but the CCSS were related to sensitive policy issues, such as accountability testing and federal versus state policy (p. 18). The adoption and implementation of new standards became a highly charged partisan issue (Smith & Their, 2017; Supovitz & McGuinn, 2017). Often reform efforts are challenged by politics (Smith & Their, 2017). However, many researchers have recommended that educators play a large role in reform efforts from the very beginning (Coburn et al., 2016; Matlock et al., 2016). In addition, researchers have found that educator beliefs about CCSS reform efforts often determine their success or failure (Fives & Buehl, 2016; Matlock et al., 2016; VanTassell-Baska & Johnsen, 2016). Thus, researchers identified several challenges educators face in implementation due to public perceptions of the new standards and, therefore, recommended educator involvement in the planning and monitoring of standards implementation.

Multiple researchers examining CCSS implementation agreed with Barret-Tatum and Smith (2018), who revealed that educators generally express positive attitudes toward the standards. Matlock, et al. (2016) noted that teachers having fewer years of experience felt more positively about CCSS than those having taught 21-25 years (p. 298). Endacott et al. (2016) highlighted the importance of distributed leadership in supporting positive attitudes toward CCSS implementation. Teachers in Swars and Chestnut's (2016) study reported that CCSS-M implementation required them to adjust their teaching practices to focus more on visual models and mathematical discourse. Some participants in McDuffie et al. (2015) questioned the instructional pacing required to meet CCSS-M expectations. Barrett-Tatum and Smith (2018) and Smith and Their (2017) both revealed that educators expressed the need for additional supports to have a successful implementation of the standards in their schools and districts. Although researchers have found educators to have positive attitudes toward standards, educators have also identified challenges to be considered for effective implementation.

Standards Implementation Needs and Challenges

Researchers have identified multiple needs of districts for effective standards implementation. Funding was a need commonly identified by teachers and administrators, as professional development and purchase of resources can be quite costly for school districts (Polly 2017; Smith & Their, 2017; Timar & Carter, 2017). Carney et al. (2016) and Floden et al. (2017) found that educators were struggling to adapt or adopt resources to address the new standards. Barrett-Tatum and Smith (2018) noted that one-third of teachers expressed a lack of professional development regarding alignment and

differences between old and new standards and approximately half of respondents expressed the need for curriculum support. Murphy and Torff (2016) found that CCSS implementation has "reduced teachers' perceived teaching effectiveness" (p. 27), noting the challenges teachers face when new standards and increased accountability expectations occur simultaneously. Smith and Their (2017) found that participants identified limited pedagogical knowledge and resource scarcity as challenges to CCSS implementation. Supovitz et al. (2016) found that teachers seek support for standards implementation from their administrators. Polly (2017) surveyed third-fifth grade educators and found a consistent need to supplement a district's or school's primary curricular resource with alternate materials. He found "ambiguity in the quality of such resources that were used" (p. 145). This finding was confirmed by McDuffie et al. (2015), who found that some teachers trusted curriculum writers to align with the CCSS-M, while "many viewed their curriculum materials as not aligned with CCSSM" (p. 18). Educators in Timar and Carter's (2017) study had difficulty locating high-quality instructional materials that were aligned to CCSS-M, noting that many resources designated as "CCSS aligned" were not explicitly written for the CCSS-M (p. 9). Therefore, researchers have identified funding for resources and professional learning as needs for effective standards implementation in school districts.

New standards provided the opportunity for teachers to develop PCK, or "the ways of representing and formulating the subject that make it comprehensible to others" (Shulman, 1986, p. 9). Several authors identified the need for teachers and administrators to study the standards to develop their understanding (Timar & Carter, 2017; Urick et al.,

2018). Barrett-Tatum and Smith (2018) and Lopez and Wise (2015) identified lack of educator preparedness as a major challenge to successful implementation of the CCSS-M. Supovitz et al. (2016) noted that CCSS knowledge was "unequally distributed across schools and within teams inside schools" (p. 12). They found multiple instances where those who were knowledgeable of the standards were not serving as resources for peers, while others who served in the roles to support CCSS implementation lacked standards knowledge (Supovitz et al., 2016, p. 12). Floden et al. (2017) noted the need for teachers to learn how to foster students' abilities to explain reasoning and challenge arguments made by others. These findings indicated a potential for increased PCK among teachers.

Some researchers identified challenges faced by educators in rural or urban settings. Timar and Carter (2017) found that rural districts needed professional development, curriculum guidance, resources, and assessment systems (p. 9). Lopez and Wise (2015) identified collaboration and planning time, knowledge of CCSS-M, and access to appropriate curriculum resources as needs for educators in rural communities (p. 53). Stosich (2016), whose study focused on high-poverty urban schools, found similar needs in an urban setting, noting that teachers "turned to their colleagues for resources, expertise, and partnership in inquiry" (p. 1708). Although teachers may express familiarity with standards, Swars and Chestnut (2016) found that urban teachers felt inhibited by their insufficient content knowledge. For example, although 83% of teachers surveyed agreed or strongly agreed that the CCSS-M would help them improve their classroom teaching practice, they identified "lack of mathematical knowledge for teaching and inadequate curriculum materials" as constraints hindering their

implementation (Swars & Chestnut, 2016, p. 217). Thus, researchers have identified and described unique challenges faced in implementing standards in both rural and urban settings.

Researchers have identified the need for leadership to support standards implementation (Endacott et al., 2016; Filippi & Hackman, 2019; Stosich, 2016; Woulfin & Rigby, 2017). Lopez and Wise (2015) advocated for distributed leadership to build local capacity within schools. Various researchers advocated for use of a coaching model to offer ongoing support to educators (Aspen Institute, 2013; Timar & Carter, 2017; Woulfin & Rigby, 2017). Rigby et al. (2018) examined student performance during a coaching initiative. They noted that students in their study grew best in settings where both the instructional coach and the principal were actively involved in the initiative. In settings where the coach had limited knowledge and skills, but the principal was active, students demonstrated almost as much growth as with the active coach and principal. However, in schools with a strong coach, but an inactive principal, students demonstrated fewer than half as much growth as in the two former contexts. Students in settings with an inactive principal and a limited coach demonstrated negative growth (Rigby et al., 2018, p. 33). Thus, there is a consensus among researchers that standards implementation has been a challenge linked to a need for an increase in both reflective practice and PCK for teachers.

Teacher Evaluation

In *The Widget Effect,* researchers identified a discrepancy between teacher evaluation ratings and student performance data (Weisberg et al., 2009). They found that

99% of teachers received a satisfactory rating even though 57% of teachers and 81% of administrators reported that a tenured teacher on their staff was performing poorly. They also found that the teacher evaluation process did not guide leaders to identify professional development needs. The findings of this study influenced federal and state policy regarding teacher evaluation (McGuinn, 2012).

Most states passed legislation requiring changes to teacher evaluation systems in response to the federal Race to the Top and NCLB waiver initiatives (U.S. Department of Education, 2009). In response, researchers have provided recommendations to states and districts regarding the design of new teacher evaluation systems that incorporate both professional practice and student growth (Marion, 2016; Measured Progress, 2014; Zefran et al., 2015). Hall et al. (2015) noted that teacher evaluation systems could serve multiple purposes: administrative, strategic, and developmental. Administrative purposes address employment matters, such as hiring, retention, or promotion. Strategic purposes deal with the relationship between employees' goals and their functions within the organization. Developmental purposes support employees to improve their performance (Hall et al., 2015). They recommended that states and districts establish clear performance level descriptors as they design teacher evaluation systems to meet the new requirements. Gagnon et al. (2017) considered the degree of local control states afforded districts in the design of teacher evaluation systems. They found considerable variation among states, noting that those who received funding from Race to the Top afforded less local control than the states that did not receive funding. Researchers have identified a

multitude of methods and challenges in the implementation of student growth monitoring practices in teacher evaluation systems.

Formative and Summative Purposes for Evaluation

Teacher evaluation researchers have examined the dual purposes of continuous improvement (formative) and ratings for employment decisions (summative). Studies by Avalos-Bevan (2018) and Liu et al. (2019) agreed with Papay (2012), who argued that teacher evaluation should measure teachers' performance accurately and support teachers' continuous growth. However, Gilles (2017) and Lillejord et al. (2019) found that districts were challenged with balancing both formative evaluation purposes and the accountability nature of summative evaluation. According to Papay (2012), "evaluations can assess how effectively teachers are doing their jobs" or they "can provide valuable information to drive professional growth and, as such, can raise teacher effectiveness" (p. 124). Papay argued that "if teacher evaluation is to improve student learning systematically, it must be used as a tool to promote continued teacher development" (p. 124). Therefore, districts are challenged to design systems that serve both purposes.

Bradley-Levine et al. (2017) presented three categories for teacher evaluations. They concurred that summative evaluation is used for employment decisions such as tenure, assignment, hiring, or dismissal, and formative evaluations support teachers in their growth and development. However, they described emergent evaluations as those that "involve individuals other than the teacher and the principal in the evaluation process" (Bradley-Levine et al., 2017, p. 68) and included merit-pay under this category. Some researchers have found that merit pay systems may impede collaborative cultures

of schools (Bradley-Levine et al., 2017; Kaimal & Jordan, 2016; Mintrop et al., 2017; Munroe, 2017; Sullivan, 2012). Researchers have also found a lack of evidence that merit pay supports consistent increases in student achievement (Kaimal & Jordan, 2016; Manzeske et al., 2016). This additional category of evaluation introduced yet another purpose, which has complicated the implementation of effective teacher evaluation systems in some districts.

Multiple researchers agreed with Berliner (2018) and Ford et al. (2018), who supported the use of teacher evaluation for professional learning purposes. Malunda et al. (2016) found that formative and summative evaluation influenced the quality of pedagogical practices, but formative evaluation yielded greater increases in the quality of pedagogical practices. Darling-Hammond (2016) noted that "teachers reported significant improvements in their knowledge and performance in each area assessed" during teacher evaluation (p. 88). Roussin and Zimmerman (2014) advocated for reflection in formative evaluation stating, "by allowing opportunities for teachers to insert personal learning goals and reflections, these types of conversations shift from episodic to planned, purposeful, and ongoing, creating a job-embedded, collaborative model" (p. 39). Thus, researchers have found formative evaluation to improve teacher knowledge and student learning.

However, researchers in some studies noted misalignment between teacher evaluation and professional development (Delvaux et al., 2013; Golberg, 2018; Ritter & Barnett, 2016). Callahan and Sadeghi (2015) noted that only 5% of respondents indicated that professional development activities were designed to be aligned with the observed

needs (p. 56). In contrast, Derrington and Kirk (2017) found that principals described learner-centered professional development most often and community-centered second most frequently. Assessment-centered job-embedded professional development was the third most frequently mentioned strategy, and knowledge-centered was not mentioned by the principals in these interviews. Therefore, the researchers found that principals did use teacher evaluation results to design professional development for teachers. They also found that principals "used a community-centered job-embedded approach by integrating professional development on teacher evaluation into existing school structures" (Derrington & Kirk, 2017, p 640). Thus, researchers have found misalignment of professional learning and teacher evaluation to impede continuous improvement efforts, while noting that alignment could support such efforts.

Researchers noted that teacher collaboration in the teacher evaluation process supported teacher growth (Darling-Hammond et al., 2012; Derrington, 2016; Derrington & Kirk, 2017). Pham and Heinemann (2014) studied a school district that implemented a teacher evaluation system that included student achievement and teacher reflection. The district implemented a peer assistance model in which all teachers had the opportunity to participate. Some researchers recommended peer assistance and review models for teacher evaluation as an alternative to test-based measures (Darling-Hammond et al., 2012; Katz, 2016). Others noted that teacher evaluation systems could be a key structure to support teachers who are struggling (Berliner, 2018; Goe et al., 2017). Consequently, with a defined structure, teacher evaluation may be an effective vehicle for supporting professional growth.

Several researchers mentioned supporting school improvement efforts as a purpose for teacher evaluation systems (Braun, 2015; Champ, 2015; Darling-Hammond et al., 2012; Holdheide et al., 2012; Huber & Skedsmo, 2016). Champ (2015) noted the emphasis on creating accountability systems that addressed improved student performance. Coburn et al. (2016) noted that the implementation of new teacher evaluation systems following state and national policy "can influence school and classroom instructional practice" (p. 246). Mette et al. (2015) emphasized the importance of collaboration among teachers and principals in the implementation of school reforms. This collaboration is especially vital in light of Bridich's (2015) findings that teachers and administrators have disparate perceptions of teacher evaluation system implementation. Rosen and Parise (2017) noted that school improvement under ESSA could be realized with investments in training for school leaders for connecting professional development with needs identified through teacher evaluation system implementation. Holdheide et al., (2012) found that districts were challenged to measure student growth and attribute that growth to the contributions of individual teachers. Thus, researchers have recommended that teacher evaluation systems be leveraged to support teachers' professional learning and school improvement efforts.

When considering the design of teacher evaluation systems, districts may examine experiences shared by early adopters. Some of these early adopters have focused on developing teacher evaluation systems that emphasize formative evaluation by incorporating coaching models and professional learning structures (Patel, 2012; Pham & Heinemann, 2014; Slotnik et al., 2014). However, Walsh et al. (2017) and Berliner

(2018) noted that many teacher evaluation systems lack alignment between student growth measures and observation measures. Researchers advised that states reevaluate systems and offer districts technical assistance to address misalignment issues (Walsh et al., 2017). Considering these recommendations, designing student growth monitoring systems that foster teachers' reflective practice and PCK would align with the goals of observation measures.

Student Growth in Teacher Evaluation

As noted previously, districts are challenged with designing systems to address multiple purposes for teacher evaluation. According to Bergin (2015), "the purpose of evaluating teacher effectiveness is to increase student learning" (p. 1). This statement is supported by Bolyard (2015) and Tripamer et al. (2014) who argued that the purpose of teacher evaluation policy is to support improved educator practice leading to enhanced student performance. Such arguments justify the inclusion of student growth data in teacher evaluation. The U.S. Department of Education (n.d.) defines student growth as "the change in achievement for an individual student between two or more points in time" (Paragraph 27). Alexander et al. (2017), Taylor and Tyler (2012), and Xu et al. (2016) all found that the incorporation of student growth in teacher evaluation is associated with some increase in student achievement. Teachers are responsible for raising "the knowledge and skill levels of students," and therefore it is reasonable to include direct evidence of student learning in teacher evaluation systems (Measured Progress, 2014, p. 5). However, Lavigne and Chamberlain (2017) found that evaluators felt less confident using student performance data to guide teachers to instructional

improvement compared to providing classroom observation feedback. Therefore, research in effective designs for student growth monitoring in teacher evaluation could support both teachers' PCK and student achievement.

The inclusion of student growth data in teacher evaluation systems can also support teacher growth. According to Darling-Hammond (2016),

> teachers note that the process of analyzing their own and their students' work in light of standards enhances their abilities to assess student learning and evaluate the effects of their own actions while causing them to adopt new practices that are called for in the assessment (p. 88).

This finding is inconsistent with the findings of Garet et al. (2017), who found no impact on teachers' interest in improving practice when the feedback was in the form of test-based data. Muñoz and Dossett (2016) noted the importance of linking teacher evaluation systems and professional development to support teacher and student growth. Therefore, although research supports the inclusion of student growth data in teacher evaluation to promote both teacher PCK and student achievement, educators need support to design teacher evaluation systems that foster these elements.

Given the multiple purposes for teacher evaluation, there are also varied reasons for including student growth in teacher evaluation. Multiple researchers examined the two common approaches to incorporating student growth into teacher evaluations: statistical models, such as VAM and student growth percentiles, and SLOs (Bergin, 2015; Gill et al., 2013; Measured Progress, 2014). Researchers expressed that the purpose of incorporating student growth into teacher evaluation was to assess teacher or school

effectiveness (Gill et al., 2013, p. 1). However, Berliner (2018), purported that educational evaluation is "done primarily to get rid of 'bad' teachers" (p. 4). Many contended that the ultimate purpose of teacher evaluation is to improve student achievement (Callahan & Sadeghi, 2015; Derrington, 2016; Gagnon et al., 2017; Mette et al., 2015; Munroe, 2017; Pham & Heinemann, 2014; Slotnik et al., 2014; Taylor & Tyler, 2012; Tripamer et al., 2014). However, researchers have noted that school districts have struggled to design teacher evaluation systems that both assess effectiveness and support improvement in student achievement.

Bolyard (2015) asserted that there is a difference between accountability and responsibility. She argued that accountability is focused on the relationship between teacher and evaluator, while responsibility focuses on the relationship between teacher and learner. Therefore, she challenged using student growth data to evaluate teachers because teachers and student share responsibility for student growth.

Research regarding the test-based measurement of student growth has yielded controversial findings. Multiple researchers examining VAM and student growth percentiles agreed with Amrein-Beardsley and Holloway (2017), who questioned the validity and reliability of using these models for isolating the effectiveness of individual teachers. Therefore, varied perceptions of purposes for including student growth in teacher evaluation has complicated student growth monitoring practices.

Test-Based Teacher Evaluation

Test-based teacher evaluation is commonly used with an accountability purpose for student growth monitoring. Garrison (2011) argued that the use of test-based teacher

evaluation and compensation stems from the business model of performance pay, which is less likely to be successful in situations that involve higher-level thinking, as it was designed for use in an industrial model. Researchers have also noted problems with manipulation of data when statistical models based on standardized tests are used for teacher evaluation (Ballou & Springer, 2015; Geiger & Amrein-Beardsley, 2017; Haladyna, 2011; Pivovarova & Amrein-Beardsley, 2018). Many researchers agree with the research of Pivovarova and Amrein-Beardsley (2018) who expressed concern with accountability models that represent an over reliance on standardized testing to make high-stakes decisions, such as employment or tenure (Amrein-Beardsley & Holloway, 2017; Backes et al., 2018; Berliner, 2018; Ford et al., 2018). Although Goldhaber (2015) agreed that imprecision makes VAM a questionable model, simulations showed the potential for guiding performance pay and high-stakes decisions.

Critics of test-based teacher evaluation have noted that these statistical measures assume random assignment of students to classrooms, which is not the typical method for schools to use (Everson, 2017; Geiger & Amrein-Beardsley, 2017; Lash et al., 2016). Without random assignment, the results from statistical methods are compromised. Shneyderman and Froman (2015) analyzed three statistical methods for using test-based assessment data for student growth to evaluate teachers: the Florida VAM model, a district covariance adjustment model, and a student growth percentile model. In the discussion of their study they stated, “the fact is, all three of these models are regrettably inadequate when it comes to measuring teacher effectiveness or even its narrower facet of

teacher effect on student assessment results" (Shneyderman & Froman, 2015, p. 9). Thus, alternative methods of growth monitoring merit study.

Challengers to statistical models for teacher evaluation systems also noted numerous factors that influence student achievement other than teachers (AERA, 2015; Amrein-Beardsey & Holloway, 2017; Katz, 2016). School factors may include class size, curricular choices, instructional time, instructional resources, collaboration structures, and peer culture (Darling-Hammond et al., 2012; Everson, 2017). Also, home factors, such as parents' availability and learning backgrounds or students' physical and emotional security, can impact student learning. Students' attendance, health, and summer experiences can lead to gains or losses in achievement (Bolyard, 2015; Darling-Hammond, 2015; Shneyderman & Froman, 2015). Therefore, Haertel (2013) recommended: "teacher VAM scores should emphatically *not* be included as a substantial factor with a fixed weight in consequential teacher personnel decisions" because "the scores may be systematically biased *for* some teachers and *against* others" (p. 23). As these researchers have argued for the exclusion of statistical, test-based models in student growth monitoring, classroom assessment-based models have been presented as an alternative method for monitoring student growth in teacher evaluation.

Classroom Assessments for Student Growth

Classroom assessment models can target accountability or professional growth purposes for student growth monitoring. Gareis and Grant (2015) and McMillan (2016) agreed with Herman et al. (2011), who argued that the most important consideration for classroom assessments used in teacher evaluation is validity. Herman et al., (2011)

proposed that a system that uses student assessment in teacher evaluation should ensure that standards clearly define student learning expectations, that assessment instruments accurately and fairly measure those learning expectations, that scores accurately and fairly measure growth, and that growth can be attributed to the contributions of individual teachers. Wilson (2018) argued that classroom assessment is at least as important as large-scale assessment in the educational process. Popham (2013) agreed that classroom assessment evidence could be used for teacher evaluation depending on whether the instruments assess significant content (versus trivial) and are valid and reliable. He also agreed that scoring must be accurate and noted that data must be collected on two or more occasions to demonstrate growth.

However, Prizovskaya (2018) questioned whether teachers have the necessary assessment literacy skills and understandings to effectively select and identify appropriate classroom assessments. She administered the Assessment Literacy Inventory, developed by Campbell and Mertler, to measure educator competence related to the assessment of students. The participants in her study scored an average of 51% on the inventory. She also found that teachers from high achieving schools performed better compared to teachers from low achieving schools. Prizovskaya (2018) recommended that a system be developed for evaluating teachers' proficiency in educational assessment and that teachers receive support developing assessment measures appropriate for instructional decisions.

These recommendations are supported by Darling-Hammond et al. (2012), who noted that teachers who used classroom assessments as evidence of student growth for

teacher evaluation improved their ability to create tools to assess student learning gains. These teachers also "showed a greater awareness of the importance of sound curriculum development, more alignment of curriculum with district objectives, and increased focus on higher-quality content, skills, and instructional strategies" (p. 14). Teachers in Tripamer et al.'s (2014) study were in favor of using multiple assessments as evidence of student learning. Leo and Coggshall (2013) supported this suggestion, advising that teachers "gather evidence of learning throughout every lesson to monitor student learning and assess the degree to which each student has met the learning goals" (pp. 11-12). Thus, multiple researchers have found classroom assessment-based models to be a viable alternative to test-based models where teachers possess sufficient assessment literacy skills to develop and analyze assessment data.

Perceptions and Impact of Teacher Evaluation

Research on perceptions of teacher evaluation has yielded mixed results. Several researchers found educators held positive perceptions of teacher evaluation systems. Ford et al. (2018), Goe et al. (2017), and Roberson-Kraft and Zhang (2018) found that supportive teacher evaluation structures led to positive changes in teacher practices. Several researchers agreed that frequent, actionable feedback to teachers was indicative of positive perceptions (Delvaux et al., 2013; Ford et al., 2018; Goe et al., 2017; Huber & Skedsmo, 2016; Kraft & Gilmour, 2016; Ritter & Barnett, 2016; Tuma et al., 2018; Tuytens & Devos, 2011). Mette et al. (2015) noted that educators had positive perceptions when teachers had discussions with evaluators about student assessment, while Raudenbush (2015) noted collaboration among teachers and administrators as an

indicator of positive perceptions. Golberg (2018) found positive perceptions when standards-based performance indicators and rubrics were used in the evaluation process. Bradley-Levine et al. (2017) and Tripamer et al. (2014) agreed that educators felt positively about the experience when teacher evaluation was connected to professional development. Slotnik et al. (2014) and Tripamer et al. (2014) agreed that the use of multiple assessment pieces was an indicator of positive perceptions.

Teachers in several studies expressed negative perceptions regarding test-based student growth measures for student growth (Berliner, 2014; Bridich, 2015; Ford et al., 2017; Jiang et al., 2015; Pressley et al., 2018;). Teachers in Callahan and Sadeghi's (2015) study felt that the value of the evaluation had diminished since it changed formats, with 44% of 2012 respondents and 42% of 2014 respondents noting that the evaluation had little effect on the way they teach (p. 53). This finding is echoed in Golberg (2018), who found that teachers did not perceive that they were growing professionally as a result of the new evaluation system (p. 74). However, teachers in Golberg's (2018) study having 1-5 years of experience found the new system helpful.

Ford et al. (2017) compared perceptions of teachers evaluated with statistical models to teachers evaluated with a goal-based approach. Although both groups expressed feelings of stress during the evaluation process, teachers using SLO were stressed to create a system for CCSS due to lack of training. In contrast, VAM teachers expressed that they felt a loss of control, and many questioned the validity of the evaluations (Ford et al., 2017). Teachers in Pressley et al.'s (2018) study also felt the loss of control during value-added model evaluations. In addition, they expressed confusion

over statistical models, which is supported by Prizovskaya's (2018) finding that many teachers lack assessment literacy.

Jiang et al. (2015) found that teachers expressed confusion and concern over the inclusion of student growth and the "narrow representation of student learning that is measured by standardized tests" (p. 112). They noted that teachers using school-wide value-added model scores were significantly more negative compared to those using individual value-added model scores and special education and high school teachers were more negative than general education and elementary teachers. Teachers were also concerned about using the performance tasks for evaluation because the tasks were too challenging at the beginning of the school year as they "generally assessed students on content they had not yet been taught" before the teachers and students had an opportunity to build relationships (p. 113).

Some researchers indicated that job satisfaction and commitment to the profession have been negatively impacted by changes to teacher evaluation systems (Ford et al., 2017; Ford et al., 2018; Lavigne, 2014). Robertson-Kraft and Zhang (2018) found that turnover rates increased in both pilot and nonpilot schools. However, turnover rates grew more in schools piloting the new teacher evaluation systems, which used student growth percentiles for the student growth component. According to Callahan and Sadeghi (2015), "teachers overall exhibited rapidly declining perceptions of self-efficacy, satisfaction, and in some cases, professional commitment" (p. 226). Studies have also shown that implementation of new teacher evaluation systems was related to decreases in teacher motivation and well-being (Berliner, 2014; Cuervas et al., 2018; Firestone, 2014).

In contrast, Ford et al. (2018) found an association between teachers who perceived that the feedback from their evaluation prompted positive changes in their practice and higher job satisfaction on average (p. 18). Therefore, evidence indicates that the models and methods for implementing student growth monitoring in teacher evaluation has influenced teachers' perceptions of evaluation experiences.

Walsh et al. (2017) noted implementation of laws requiring changes to teacher evaluation systems have not impacted the number of teachers rated proficient. The researchers expressed concern that teachers who lacked strong evidence of student growth, could still earn proficient ratings in some states and districts. Xu et al. (2016) found that "principals' ratings could only moderately explain student achievement gains" (p. 218). They hypothesized that the lack of alignment between student achievement and teacher performance could be due to value-added model fallibility or lack of the principal's skill in evaluation.

Derrington (2016) described how the implementation of new teacher evaluation systems positively impacted one school by causing the creation of professional development structures. The school implemented student achievement meetings, where grade level teams met to review data and link it to classroom instruction, and vertical team meetings, where teachers discuss curriculum and share strategies across grade levels (p. 189). Walsh et al. (2017) called for states to reevaluate their systems to offer districts more guidance so that they can establish structures that focus on professional learning, as the school in Derrington's (2016) study has done.

Many researchers noted that principals play a key role in the success or failure of teacher evaluation systems reform (Bradley, 2014; Cannata et al., 2017; Champ, 2015; Delvaux et al., 2013; Derrington, 2016; Mette et al., 2015). Donaldson and Woulfin (2018) and Gill et al. (2014) noted the large role that principals play in SLO implementation. Donaldson and Woulfin (2018) found that principals used discretion to support teacher learning by integrating feedback into improvement efforts. Kraft and Gilmour (2016) found that the quality of feedback teachers received from principals was related to the amount of time principals could spend and the training principals received; and Young et al. (2015) found that principals valued the formative feedback they provided through teacher evaluation systems. However, Goldring et al. (2015) found that principals relied more heavily on data from observations than from value-added model-based student growth. Thus, although researchers have found feedback from evaluators to be beneficial in supporting teacher growth, school districts have struggled to design systems that provide such feedback around student growth data.

Challenges of Simultaneous Initiatives

Various researchers noted the difficulty in implementing new standards and new teacher evaluation systems simultaneously (Backes et al., 2018; Coburn et al., 2016; Doherty & Jacobs, 2015; Leo & Coggshall, 2013). Herman et al. (2011) argued that assessments for teacher evaluation systems should be standards-based. They stated, "assessments that are likely to be sensitive to instruction are composed of items and tasks that reflect the core goals represented in standards and learning progressions and do not include tangentially related content" (p. 10).

Researchers also recommended that training and evaluation for teacher evaluation systems and CCSS implementation be closely aligned (Leo & Coggshall, 2013; Marion et al., 2012). Coburn et al. (2016) predicted the following four possible scenarios for implementation:

> 1. Weak accountability and low alignment to CCSS would likely lead to little change in instructional practice.
> 2. Strong accountability and low alignment to CCSS would likely lead to resistance and superficial change.
> 3. Weak accountability and high alignment to CCSS would likely yield less resistance, but inconsistent implementation.
> 4. Strong accountability and high alignment to CCSS could support teachers to develop deeper understanding of CCSS and result in more substantive implementation of CCSS and teacher evaluation systems. (p. 247)

Leo and Coggshall (2013) advised that implementation should begin with a thorough review of the standards to identify instructional practices that align with the new expectations. Leo and Coggshall (2013) stated, "professional learning focused solely on curriculum implementation of the Common Core standards and disconnected from teachers' individual needs will only add to the confusion about instructional priorities" (p. 5). However, Slotnik et al. (2014) noted that districts were struggling to make connections between teacher evaluation systems and CCSS. Even though participants expressed positive views of using SLOs and classroom assessments to monitor student growth, they desired additional support in using standards and data (Slotnik et al., 2014).

Thus, student growth monitoring that supports teachers' reflective practice and PCK of targeted standards should align to standards implementation efforts.

Researchers have found that many stressors influence ratings in newly implemented teacher evaluation systems. Among the identified stressors were per-pupil spending, enrollment, and student performance (Lenhoff et al., 2018). Researchers noted that, although policymakers expected to see an increase in teachers rated on the low end of the performance scale, this was not the case. Thus, research indicates that educators need support to develop teacher evaluation systems that simultaneously support standards implementation.

Student Learning Objectives (SLOs)

SLOs are student growth measures that use assessment data other than standardized tests and are also referred to as student growth objectives, student growth goals, measures of student learning, analysis of student work, and student learning targets (Hewitt & Amrein-Beardsley, 2016). Cardno et al. (2017) called the evaluation process in their study "teaching as inquiry," likening it to an action research model (p. 17). SLO originated in districts that were implementing incentive pay programs (Crouse et al., 2016; Lacireno-Paquet et al., 2014). However, roughly two-thirds of states discuss the incorporation of an SLO process as a student growth measure either alone or in conjunction with another measure, and most do not connect the use of SLOs to performance pay (Hall et al., 2014).

Some researchers defined SLO as a process in which teachers use baseline data to set measurable goals for students who are monitored for a defined and significant time

period (Joyce et al., 2016; Lacireno-Paquet et al., 2014). Others defined SLO as the classroom- or grade-specific objectives or goals that teachers or teacher teams use to monitor student learning over a set time frame (Gill et al., 2014; Lachlan-Hache et al., 2012; Marion et al., 2012; Reform Support Network, 2010; Reform Support Network, n.d.). Although Kearns et al. (2015) used the latter definition, they noted that SLOs "can constitute an instructional improvement process, driven by teachers in all grades and subjects" (p. 27). States vary in their definitions of SLOs. However, according to the Reform Support Network (2014), most states provide a template with the following common elements: (a) student population (quantity and description); (b) interval of instruction (beginning and end dates); (c) learning content (standards, knowledge, & skills); (d) baseline; (e) assessments; and (f) targets.

Many researchers examined the implementation of SLOs in states and districts and identified both benefits and challenges for using SLOs to monitor student growth in teacher evaluation systems. Several researchers found that SLOs promote collaborative discourse among teachers and administrators (Plecki et al., 2016; Reform Support Network, 2014; Slotnik et al., 2015). Lachlan-Hache (2015), McCullough et al. (2015), Slotnik et al. (2014), and Slotnik et al. (2015) also found that SLOs promote data-driven instruction. Crouse et al. (2016), Plecki et al. (2016), and Slotnik (2015) all noted that SLO promote reflective practice in assessment and instruction. Joyce et al. (2016), Marion et al. (2012), and Marion (2016) agreed that the SLO can be used for monitoring student growth in the evaluation of teachers of nontested subjects and grades. Briggs (2013), Joyce et al. (2016), Marion et al. (2012), and McCullough et al. (2015) found

SLO to actively engage and empower teachers. Joyce et al. (2016) and Lachlan-Hache (2015) found SLO implementation to support teachers in assessment development, which allowed districts alternatives to standardized testing. Marion (2015), Slotnik et al. (2014), and Slotnik et al. (2015) all observed SLO implementation to support instructional improvement. The Reform Support Network (2014) argued that SLO use in teacher evaluation promoted alignment among standards, curricula, assessment, and instruction. Joyce et al. (2016) and Kearns et al. (2015) agreed that implementing SLOs supports the monitoring of students with disabilities using goals and targets that are aligned with both the classroom objectives and the students' individualized education plans.

In contrast, some researchers indicated that SLO implementation was challenging for schools and districts. Lachlan-Hache (2015) and Marion et al. (2012) found that educators had difficulty accessing valid data for their SLOs. Several researchers found that teachers needed much support selecting and developing assessments (Lachlan-Hache, 2015; Marion et al., 2012; Plecki et al., 2016; Slotnik et al., 2015; Thompson, et al., 2016). Lachlan-Hache (2015), McCullough et al. (2015), Plecki et al. (2016), and Slotnik, et al. (2015) all found that both teachers and evaluators needed support analyzing data. McCullough et al. (2015) and Riordan et al. (2015) found that the time needed for SLO implementation drew teachers and administrators away from other responsibilities. Briggs et al. (2015) noted that the lack of clarity regarding SLO expectations, such as "murky definitions of 'growth'" and imbalance between formative and summative use of evaluation, threatens the validity of SLOs. Thus, the lack of clarity around growth and structure has complicated SLO implementation.

In addition to these challenges, some states and districts have implemented SLO processes using questionable practices that may invalidate or bias the processes. For example, Marion et al. (2012) noted that some districts instituted SLOs using gain growth models. These scores are often based on non-equated test scores. Without a scaled score, these judgments about gain may be invalid or unreliable. Simple gain practices may also not consider the context of students' growth, as "students tend to grow at very different rates regardless of the quality of teaching" (p. 4). Many researchers agreed with Balch and Springer (2015) who indicated a need for further research in SLO implementation and interpretation.

Perceptions of SLO Experience

Several studies on SLO implementation have included teacher and evaluator perceptions regarding their experiences. Many researchers have identified positive teacher perceptions of their experience with SLO. Two studies found that teachers perceived the SLO implementation as beneficial to students. Plecki et al. (2016) noted that 47% of teachers perceived a positive impact on student achievement and 44% believed there would be no impact on achievement. Makkonen et al. (2015) found that more Utah teachers agreed than disagreed that the SLO process was beneficial to their students.

Other sources noted that teachers felt the implementation of SLO improved their assessment practice. Lachlan-Hache (2015) cited multiple studies that provided evidence to support this claim. Among those she cited were Slotnik et al. (2013), who noted that "interviewees consistently remark on the SLO baseline data step as one that was

informative, beneficial, and frequently enlightening, in the conduct of their instructional planning," and Lamb et al. (2013), whose participants reported "that using SLOs encouraged teachers, especially new teachers, to analyze student data" (Lachlan-Hache, 2015, p. 4). Similarly, Plecki et al. (2016) found that 62% of teachers felt "that the evaluation system will prompt them to consider alternative forms of assessment" (p. 108).

Several researchers agreed with Kearns et al. (2015), who found that teachers observed SLO participation improved their instructional practice and supported the instructional planning process. McCullough et al. (2015) stated that "teachers and teachers' union officials in districts that used student learning objective reported that the measures informed instructional practice" (p. 11). Plecki et al. (2016) reported that 52% of participants felt implementation of SLO would improve their instruction and 56% felt SLO would support alignment of instructional improvement activities in their school or district (p. 108). Gill et al. (2014) agreed that SLO help teachers to plan instruction (p. ii). Riordan et al. (2015) found that 53.9% of teachers felt the new teacher evaluation systems would improve teaching (pp. B–3).

Teachers have also expressed that SLO implementation was time-consuming and increased their responsibilities (Lachlan-Hache, 2015). Plecki et al. (2016) reported that 94% of teachers felt the new teacher evaluation systems would increase their workload. Riordan et al. (2015) noted that teachers felt it took a lot of time and effort to complete paperwork to prepare for meeting with their evaluators. Collectively, these findings indicate that, although many teachers have increased their workload for monitoring student growth in teacher evaluation, they found the work to improve their practices.

Multiple studies examined evaluator perceptions of SLO implementation. Evaluators in Plecki, et al (2016) reported concerns regarding valid and reliable assessment tools and practices. Woulfin et al. (2016) noted that district leaders promoted both accountability and development purposes of SLO implementation. Riordan et al. (2015) noted that evaluators generally expressed more positive perceptions of SLO implementation than teachers. For example, 83.3% of evaluators compared to 68.6% of teachers perceived the teacher evaluation systems as fair, 66.6% of evaluators compared to 45.1% of teachers felt the system would result in accurate ratings, and 83.3% of evaluators compared to 53.9% of teachers felt the system would improve teaching (p. B-3). Slotnik et al. (2014) agreed, stating "principals are more likely than teachers to agree with statements about positive implications" of the new teacher evaluation systems (p. 1). However, in a follow-up study, Slotnik et al. (2015) noted that "more teachers and principals agree than disagree that they have a common language to describe the SLO process and that expectations are clear" (p. 3). This lack of consistency indicates a need for a clearer structure for monitoring growth with SLOs for teacher evaluation.

McCullough et al., (2015) noted that evaluators reported SLOs were effective for "fostering collaboration, targeting professional development, encouraging data-driven instruction, and building assessment capacity" (p. 10). McCullough et al. (2015 found that "district administrators and principals noted that student learning objectives helped build community and accountability at the school level, galvanizing school staff around similar goals" (p. 13). Plecki et al. (2016) also found that evaluators agreed SLO implementation positively influenced the quality of collaboration and professional growth

(p. 111). Principals reported that the new teacher evaluation systems supported growth for all of their teachers, and a superintendent noted the "level of discourse between the administrator and them [teachers], and the level of discourse in their team around some of this stuff has been significantly deeper and more focused" (p. 111). Thus, researchers have found that evaluators also perceived SLO implementation in teacher evaluation to improve teacher practices.

However, evaluators in many studies also noted that implementation of SLO was challenging (Riordan et al., 2016; Slotnik et al., 2014; Woulfin et al., 2016). Riordan et al. (2015) stated:

> Introducing and designing student learning objectives proved to be more challenging than implementing other features of the new evaluation systems. Evaluators did not feel as prepared to implement SLOs as they did to implement other system features for which they received training. Although 60-70 percent of evaluators participated in training that addressed how to write SLO and determine whether teachers had achieved them, only 53 percent indicated that they felt prepared to write or review SLO. (p. 9)

Slotnik et al. (2015) noted that evaluators continued to feel they needed support in SLO implementation, but the needs had evolved from the beginning of implementation (p.14). Although researchers have found SLO to be a promising model for promoting teacher growth in instructional practice, they found educators were challenged to implement them without a clear structure for monitoring student growth.

Variation in Expectations and Implementation

Several researchers noted the great variation in state and district requirements for student growth implementation that incorporate some form of SLO processes (Crouse et al., 2016; Joyce et al, 2016; Longo-Schmid, 2016). Cushing and Meyer (2014) showed that variability is based on state and district choices regarding the balance between teacher autonomy and SLO comparability. They argued that increases in teacher autonomy led to decreases in comparability among the SLO. Crouse et al. (2016) described the variation in SLO implementation across states as a continuum from more local to more state involvement and control. They classified the variation for four components of the SLO system: focal student population, target comparability, assessment choice, and district quality control and monitoring. Lachlan-Hache (2015), Plecki et al. (2016), and the Reform Support Network (n.d.) also found variation regarding the first three components on Crouse et al.'s list and added variation for the time frame of an SLO. According to Cushing and Meyer (2014), states and districts must determine "whether they value one characteristic more than another and then select an assessment approach that reflects those values" (p. 1).

Consideration is needed for student growth monitoring in non-tested subjects and grades. According to Watson et al. (2009), 69% of teachers are associated with non-tested subjects and grades. Hall et al. (2014) discussed the variation in methods for monitoring growth with teachers of tested subjects and grades and non-tested subjects and grades. They noted that the lack of resources and guidance for student achievement measures in non-tested compared to tested subjects and grades contributes to the variation in

approaches. The researchers recognized the difficulty this disparity may present for implementation, noting that teachers of tested subjects and grades may feel that the use of statistical methods holds them accountable for more rigorous expectations while teachers of non-tested subjects and grades feel the process is unfair due to the workload of gathering and developing resources to document student learning in an SLO (Hall et al., 2014, p. 24-25). McCullough et al. (2015) confirmed this concern when a teachers' union representative expressed that having some teachers evaluated with statistical models and others evaluated with SLOs might "induce resentment and backlash" (p. 14).

Consideration is also needed for evaluating teachers of special populations. Kearns et al. (2015) and Joyce et al. (2016) discussed the variation of experiences for teachers of general education students compared to teachers of students with disabilities. Kearns et al. (2015) noted the need for teacher evaluation systems to consider multiple factors when measuring teacher effectiveness, stating "characteristics of the learner, complexity of learner needs, and lack of opportunity to learn all contribute to a high degree of variability in the sophistication with which students engage in academic content that is grade specific and chronologically appropriate" (p. 23). Joyce et al. (2016) noted the variation in state rules for inclusion of students with disabilities in general education teachers' evaluation ratings. Variation typically involved identification strategies for target populations, goals for students, criteria for teacher effectiveness ratings, and the weight of SLO scores in those ratings (p. 12).

Researchers also revealed variance of assessment models among states and districts. SLO may have used vendor-developed tests, teacher-made tests, or rubrics.

Teachers in Makkonen et al.'s (2015) study primarily used vendor-developed tests. However, teachers in Schmitt and Hutchins' (2015) study either used teacher-developed tests or rubrics. Schmitt and Hutchins (2015) found that "student growth on teacher-created multiple-choice assessments was significantly worse than on other assessments, as were the percentages of students who met growth targets" (p. 1). Briggs (2013) noted advantages of teacher-made tests are the involvement of teacher in the process, the possibility for immediate scoring and use of the assessment, and the use of results for instruction (p. 28).

The diversity of interpretations of SLOs presents a challenge in the comparison of effectiveness of SLOs among a variety of settings. However, the consensus among researchers that implementing an SLO process to monitor student growth in teacher evaluation systems shows promise in supporting teachers to improve their practice. Thus, the literature indicates a possible connection between SLO implementation to both reflective practice and PCK.

Learning Progressions

Briggs et al. (2015) recommended the use of learning progression frameworks as a foundation for SLO implementation in student growth monitoring. stating, "inferences about student growth … need not only learning objectives, but a framework that structures objectives into a progression of student learning" (p. 1). Briggs et al. (2015) argued that learning progression frameworks promote educators and students to look beyond correct and incorrect responses to the level of thinking that students demonstrate in the tasks they attempt. Hess (2012) clarified that "learning progressions, progress

maps, developmental continuums, and learning trajectories are all terms that have been used in literature over the past decade to generally mean research-based, descriptive continuums of how students develop and demonstrate deeper, broader, and more sophisticated understanding over time" (p. 2). A standards-based rubric represents a portion of a learning progression by delineating the levels of expectation from the prerequisite grade level standard(s) to the targeted grade level standard(s) and continues to the ensuing grade level expectations. It also represents the progression within the grade with the inclusion of surface and deep grade level understandings.

Researchers do not all define learning progressions in the same way. According to Duschl et al. (2011) and Clements (2011), the fact that the term "learning progressions" may be used to describe sequences that have a variety of components and take different forms can cause ambiguity in the interpretation of learning progression literature. Mosher (2011) noted that "the work on learning progressions ranges in grain size—from one day's lesson to the entire Pre-K-12 grade span" (p. 4). Most learning progressions include upper and lower anchors that describe learning goals (Duschl et al., 2011; Gotwals, 2018). Some learning progressions also include an instructional sequence and tasks (Clements, 2011; Duschl et al., 2011; Fonger et al., 2018) and/or student misconceptions (Kobrin et al., 2015). Therefore, different interpretations of learning progressions can lead to varied purposes for their use.

Kobrin et al. (2015) argued that "learning progressions can be used to inform development of standards, to guide curriculum development, to build large-scale assessments, to help teachers conduct formative assessment, and to help teachers in their

own professional development" (p. 59). Kobrin et al. (2015) emphasized that learning progression grain sizes can and should differ according to the purpose of the learning progression. For example, the learning progression that informed the development of CCSS-M have a relatively large grain size, as the learning progression spans many grade levels (Daro et al., 2011; Gotwals, 2018; Kobrin et al., 2015). In contrast, learning progressions that inform instruction should have a relatively small grain size (Gotwals, 2018; Kobrin et al., 2015). In the context of an SLO, learning progressions with a relatively small grain size support teachers to monitor student growth using formative assessment practices while also supporting teachers' professional growth (Briggs et al., 2015; Hess, 2012).

Several researchers agreed with the work of Black et al. (2011) and Fonger et al. (2018), who noted that learning progressions support alignment of curriculum, assessment, and instruction. Others emphasized that learning progressions support the assessment of standards (Daro et al., 2011; Duschl et al., 2011; Fletcher et al., 2017). Multiple researchers agreed with Furtak et al., (2018) who noted the usefulness of learning progressions for tracking student growth. Kingston et al. (2015) added that learning progressions support communication about progress to parents. Thus, learning progressions support educators in standards implementation.

Several researchers agreed with Lai et al. (2017) who suggested the need for a cognitive model in the design of a learning progression. Black et al. (2011) noted several possible cognitive taxonomies for learning progressions, such as Bloom's Taxonomy of Educational Objectives, Haladyna's Cognitive Operations Dimensions, and Biggs and

Collis's SOLO Taxonomy (p. 91). Multiple researchers emphasized that cognitive taxonomies in learning progressions guide students and teachers to focus on the increase in sophistication of thinking (Briggs & Peck, 2015; Clements, 2011; Daro et al., 2011; Fonger et al., 2018; Mosher, 2011). Alonzo (2017) argued that the cognitive model guides educators to provide actionable feedback. Therefore, researchers agree that learning progressions designed around a cognitive model provide structure for monitoring student growth.

Learning Progressions, Formative Assessment, and Instruction

Many researchers agreed with Furtak et al. (2018), who found that learning progressions support both the design and the interpretation of assessment. Graf and Arieli-Attali (2015) purported that learning progressions can support the development of assessment for complex thinking. Both Hess (2011) and Nichols (2011) agreed, noting that assessing deeper knowledge goes beyond facts and skills to the interconnection among ideas. Both Black et al. (2011) and Briggs and Peck (2015) argued that aligning both formative and summative assessments to a common learning progression supports alignment between the instruments. Gitomer (2011) noted that the combination of formative and summative assessments provides information at the individual and group levels. Graf and Arieli-Attali (2015) also argued that developing assessment tasks aligned to a learning progression supports diagnosis of student strengths and weaknesses (p. 201). Additionally, Kingston et al. (2015) and Thissen (2015) noted that assessment tasks might provide evidence of student thinking applicable to multiple learning progressions. Mosher (2011) noted the value of assessment items that discriminate among the levels of

a learning progression compared to general dimensions or topics. Thus, researchers have found learning progressions to support educators to interpret assessment data.

Researchers examined educators' formative and summative use of assessment evidence. Both Alonzo (2018) and Kobrin (2016) noted educators' tendency toward a dichotomous view of student understanding (right answers indicate understanding and wrong answers indicate a lack of understanding) without consideration of the level of thinking. However, both researchers found that the learning progressions supported educators to broaden their view of assessment evidence and consider degrees of understanding. In addition, Briggs et al. (2015) recognized the need for multiple items for each level of a learning progression to gather evidence of student thinking. Consequently, researchers have found learning progressions to support teachers' understanding of formative assessment practices.

Multiple researchers noted that formative assessment refers to the practice of using assessment tools and strategies to guide instructional decisions as opposed to designating an instrument as formative (Alonzo, 2018; Gotwals, 2018; Hegazy & Barton, 2017). Furtak et al. (2018) elaborated on this interpretation by including a description of the Formative Assessment Design Cycle as a five-step process in which teachers use learning progressions to guide development and interpretation of student work samples (p. 145). Both Alonzo (2018) and Gotwals (2018) explained that both teachers and students are involved in the formative assessment process.

Researchers in several studies mentioned the use of learning progressions to inform feedback (Alonzo, 2017; Alonzo, 2018; Briggs et al., 2015; Fletcher et al., 2017;

Gotwals, 2018; Graf & Arieli-Attali, 2015). Dunne (2011) stated that one "major function of the road maps and the construct maps is to locate assessment and learning in a constant or regular series of feedback cycles" (p. 135). Kobrin (2016) noted that acting on assessment data can be the most challenging part of the feedback cycle. Thus, incorporating a learning progression that promotes feedback into the teacher evaluation system may support teachers' use of assessment data in planning instruction.

Multiple researchers agreed with Penuel (2015), who argued that learning progressions support feedback to teachers regarding student readiness through a diagnostic. As Kobrin et al. (2015) noted, "it is important for a learning progression to clearly define prerequisites if it is to be used for curriculum development, formative assessment, and teacher development so that teachers understand their students' preconceptions in a domain" (p. 65). Alonzo (2011) stressed that teachers need to ascertain student misconceptions as well as their understanding and depth of thinking. Gotwals (2018) argued that learning progressions provide a structure that moves beyond the dichotomous interpretation of student performance to "levels of sophistication along a progression" (p. 160). Mosher (2011) also noted that teachers "take responsibility for monitoring students' progress and intervening on a timely basis when needed" (p. 1). Many agree with Hegazy and Barton (2017) that learning progressions support teachers to provide descriptive, actionable feedback to students. Shepard (2018) noted that feedback that guides students to understand how to improve calls for a qualitative rather than quantitative structure (p. 169) and Mosher (2011) argued that students need feedback that is focused on the particular difficulties they are experiencing (p. 1).

However, Kobrin (2016) found that many teachers needed additional support and guidance to use learning progressions to provide actionable feedback (p. 173). Thus, incorporating learning progressions in the teacher evaluation system may provide opportunity for dialogue among teachers and evaluators to address this challenge.

Researchers also noted that learning progressions support students in both self-reflection and providing peer feedback (Black et al., 2011; Hegazy & Barton, 2017). Dunne (2011) added that using learning progressions to support students in self-reflection can support student ownership of the learning process, leading to greater self-esteem and collaboration (p. 135). This assertion is supported by Hegazy and Barton (2017), who noted that students who self-regulate develop a stronger sense of self and increase their motivation (p. 13). Popham (2008) also described the cultural shift of a classroom from teacher-centered to student-centered when peers provide feedback. In such a setting, both formal and informal assessments "routinely supply the evidence students and teachers need to make appropriate learning related decisions" (p. 96). Hattie and Donoghue (2016) proposed a model that represents a progression from surface learning to deep learning, to transfer, arguing that teachers should choose appropriate instructional methods for each of the learning phases. Thus, learning progressions can provide a framework for synthesizing understandings of standards, assessment, and instruction.

Benefits of Learning Progressions

Researchers have found the incorporation of learning progressions benefits teachers in several ways. Many agreed with Gotwals's (2018) assertion that teachers who use learning progressions will improve their knowledge of assessment and formative

assessment practices. Furtak et al. (2018) found that the introduction of learning progressions guided teachers to "develop sets of formative assessment tasks that aligned to multiple learning progressions" (p. 153). Hess (2011) found that teachers who analyzed formative assessment data using learning progressions were able to design more effective assessments and instruction, while Sarama et al. (2017) noted that teachers learned to adjust groups and differentiate instruction for students' individual needs. However, Heritage (2011) asserted that teachers need training to effectively use learning progressions for formative application.

Researchers have shown that teachers using learning progressions also benefit from an increased understanding of their students. Clements (2011) argued that learning progressions focus attention on student thinking rather than correct or incorrect responses. Multiple researchers agreed with Confrey et al. (2015), who noted that learning progressions guide item development to reveal a range of student strategies and levels of understanding. Sarama et al. (2017) observed that teachers changed their beliefs about the content students could address as they reflected on their assessment evidence and learning progressions (p. 65). Therefore, use of learning progressions has the potential to shift educators' views of assessment from strictly right or wrong to revealing nuances of understanding or lack of understanding for the expectations articulated in the standards.

Numerous researchers agreed with Arieli-Attali and Cayton-Hodges (2014) and Krajcik (2011), who argued that working with learning progressions fostered teachers' increased understanding of content. Birkhead et al. (2017) noted that teachers' instruction

improved as their understanding of algebraic reasoning increased. Sarama et al. (2017) noted that teachers who applied learning progressions learned to describe student thinking and learning using explicit language and became more confident in their own understandings of early mathematical ideas. Gotwals (2018) argued that the ways learning progression levels are defined can support teachers to think beyond dichotomous interpretations of student responses and distinguish nuances in students' ideas. Heritage (2011) argued that the learning progression provides a structure for teachers to examine their content knowledge and collaborative discussions with peers about learning progressions can support increased PCK.

Teachers' use of learning progressions has also been found to promote teachers' reflective practice (Ariell-Attali & Cayton-Hodges, 2014; Kobrin et al., 2015; Kobrin, 2016; Sarama et al., 2017). Engelhard and Sullivan (2011) noted that reflecting on summative assessments aligned to a learning progression can be a formative experience for teacher learning of PCK. Furtak et al. (2018) articulated how teachers reflect on classroom practice by collaboratively examining student work samples and a corresponding learning progression. Collectively, the research indicated the potential for learning progression use to promote increases in teachers' reflective practice.

When teachers incorporate learning progressions into their practice, students also benefit. Fonger et al. (2018) and Sarama et al. (2017) both noted that teachers use of learning progressions led to differentiated assessment tools and practices, which allowed teachers to better elicit evidence of individual student needs. By locating students' positions along the learning progression, researchers argued that teachers can better

differentiate their instructional practices, tools, sequencing, and pacing (Dunne, 2011; Fonger et al., 2018; Gotwals, 2018; Heritage, 2011; Kobrin et al., 2015; Sarama et al., 2017). Learning progressions also guide teachers to provide targeted feedback that informs students about their thinking and learning processes (Confrey et al., 2015; Hegazy & Barton, 2017). As one goal of teacher evaluation is increased student achievement, these findings indicated that incorporation of learning progressions into the teacher evaluation system has the potential to support this goal.

Learning Progressions in Teacher Evaluation Systems

Briggs and Peck (2015) argued that teacher evaluation systems should not compare teachers based on student achievement without considering growth; however, they noted that quantifying student growth can be problematic (p. 75). Briggs and Peck (2015) also argued that learning progressions support the use of both norm-referenced and criterion referenced interpretations of student learning (p. 79). Using a learning progression in teacher evaluation systems provides a definition of growth beyond counting correct responses. Confrey et al. (2015) defined growth as a change in knowledge over time (p. 101). Hess (2011) noted that learning progressions could provide a clearer understanding of within-grade progress (p. 13). However, Maul (2015) argued that claims of change on particular attributes depend on clear descriptions of the attributes. Mosher (2011) described growth in terms of movement across levels of a learning progression over time.

Researchers have argued that using learning progressions in teacher evaluation systems can enhance teacher practices (Briggs & Peck, 2015; Kobrin et al., 2015).

Gotwals (2018) and Furtak et al. (2018) recommended that teachers use the learning progression to examine the nuances in student understandings as professional development in standards, assessment, and instruction. Kobrin et al. (2015) purported that learning progressions offer the promise of increasing teachers PCK, allowing teachers to develop a deeper understanding of how students develop more sophisticated thinking over time. Therefore, findings from these studies support the use of learning progressions to structure student growth monitoring in teacher evaluation.

Multiple researchers indicated that the incorporation of learning progressions into teacher practice is most effective when teachers manage the process of gathering and interpreting evidence of student thinking (Black et al., 2011; Furtak et al., 2018; Heritage, 2011; Mosher, 2011). Black et al. (2011) stated that the process must be "directly related to the instructional goals behind the construct maps" (p. 99). Furtak et al. (2018) argued that learning progressions "serve as centerpieces for teachers' ongoing engagement in the processes of alignment between curriculum, instruction, and assessment" (p. 143). Studies have also demonstrated the effectiveness of incorporating learning progressions into teacher evaluation systems when teachers work collaboratively to establish learning progressions, design tasks, and interpret assessment data (Briggs et al., 2015; Hess, 2012; Krajcik, 2011). Hess (2011) noted that collaborative analysis led to "designing more effective assessment and instruction" and "represented cultural shifts in school communities" (p. 153). Thus, research supports the incorporation of learning progressions in the teacher evaluation process to promote teachers' reflective practice and PCK.

Structure of the Observed Learning Outcome (SOLO) Taxonomy

Because researchers indicated that learning progressions are effective if they are framed around a cognitive model, many of the rubrics in this study use the SOLO taxonomy as the cognitive framework for applying learning progressions to assessment. Biggs and Collis (1982) presented SOLO as a model for considering both the quantity and quality of learning. The model builds on the work of Marton and Säljö (1976), who described the quality of learning as surface learning or deep learning. The SOLO model consists of five levels, with Levels 1-3 describing surface learning and Levels 4 and 5 representing deep learning.

> Level 1 Pre-structural: The learner offers no attempt to respond, or the attempt is irrelevant. Level 2 Uni-structural: The learner can provide one relevant datum in response to a cue. Level 3 Multi-structural: The learner provides multiple isolated data relevant to a cue. Level 4 Relational: The learner can describe interrelations between and among relevant data and use inductive reasoning. Level 5 Extended Abstract: The learner can provide multiple interrelations and hypotheses for relevant data using both deductive and inductive reasoning (Biggs & Collis, 1982, p. 24-25).

Multiple researchers noted that SOLO represents the progression from surface to deep learning (Hattie & Purdie, 1998; Newton & Martin, 2013). Smith and Colby (2007) clarified that a surface approach focuses on memorization and requires minimal engagement with a task, but a deep approach involves reflective thinking. Some researchers noted that students who demonstrate thinking at the extended abstract level

are extending beyond the learning target, which could include learning expectations for a higher grade or course level (Hattie & Purdie, 1998; Jurdak, & Mouhayar, 2015). Caniglia and Meadows (2018) added that, at the extended abstract level, students thinking may involve reflection and evaluation. Caniglia and Meadows (2018), Hattie and Purdie (1998), and İlhan and Çetin (2016) emphasized that a strength of SOLO is the ability to capture both quantitative and qualitative. The model distinguishes among levels of thinking instead of tracking only correct and incorrect responses (Chan et al., 2002; Hattie & Purdie, 1998; İlhan & Çetin, 2016; Jurdak & Mouhayar, 2015; Prakash et al., 2010; Wells, 2015). Authors noted versatility of SOLO in that it can apply across content areas and grade levels (Biggs & Collis, 1982; Caniglia & Meadows, 2018; Chan et al., 2002; Hattie & Purdie, 1998; Keskin et al., 2016; Wells, 2015). Thus, studies of SOLO show that it is a viable cognitive model for structuring a learning progression.

Several researchers have compared SOLO to other models. Newton and Martin (2013) argued that phenomenology and Bloom's taxonomy could also provide a structure for promoting deeper learning. However, they, along with Hattie and Purdie (1998) agreed that Bloom's taxonomy only judges the questions and not the student responses. Hattie and Purdie (1998) added that SOLO considers that questions and answers may be at different levels of complexity. Hattie and Purdie (1994), İlhan and Gezer (2017) and Newton and Martin (2013) found that teachers scoring based on SOLO showed more interrater reliability than when using Bloom's taxonomy. İlhan and Gezer (2017) added that Bloom's was "more open to random error" (p. 647). İlhan and Çetin (2016) noted that raters found SOLO-based rubrics to be more objective and reliable than standard

rubrics (rubrics not based on a cognitive model). Therefore, multiple research findings support the use of SOLO taxonomy to structure standards-based rubrics.

SOLO taxonomy has been found to serve multiple functions. Several researchers noted the role SOLO could play in defining instructional learning outcomes (ILO) (Brabrand & Dahl, 2009; Prakash et al., 2013; Rembach & Dison, 2016). Smith and Colby (2007) argued that SOLO could also support educators in developing their understanding of depth and complexity of learning expectations. They, along with Rembach and Dison (2016) and Prakash et al. (2010) emphasized the benefit of SOLO for clarifying expectations for students. Fonger (2017) argued that SOLO could be used for monitoring growth in student thinking.

Proponents of SOLO purported that it can support the analysis of questions as well as responses (Smith & Colby, 2007; Hattie & Purdie, 1998; Wells, 2015). Researchers noted that learning targets are generally at the relational level (Biber & Incikabi, 2016; Keskin et al., 2016; Prakash et al., 2010). However, multiple studies found that teacher questions were primarily at the surface levels (Biber & Incikabi, 2016; Caniglia & Meadows, 2018; Keskin et al., 2016; Smith & Colby, 2007). In addition, studies found a significant number of students performing at uni-structural and multi-structural levels and struggling at the relational level, demonstrating surface level understanding (Gagani & Misa, 2017; Keskin et al., 2016; Ozdemir & Goktepe-Yildiz, 2015; Smith & Colby, 2007; Soobard & Rannikmae, 2015). Both Jurdak and Mouhayar (2015) and Kusumawathie et al. (2017) found that the complexity level of tasks influenced student performance. Smith and Colby (2007) recommended that teachers

collaboratively examine student work samples using the SOLO taxonomy to analyze "how and why particular work samples represent various levels" (p. 208). Consequently, SOLO has been found to support assessment analysis and instructional planning.

Biggs and Tang (2011) argued that SOLO supports constructive alignment. Constructive alignment involves students working on tasks that align to the instructional learning outcome. This claim is supported by the research of Prakash et al. (2010), Rembach and Dison (2016), and Smith and Colby (2007). Hattie and Purdie (1998) evaluated intervention programs with respect to SOLO and found that the programs that were designed at the relational level were highly effective in all domains. In Kusumawathie et al.'s (2017) program evaluation, they found a strong relationship between curriculum inputs and both the development of a SOLO-based curriculum and the SOLO-based curriculum development process.

Researchers recommend that SOLO taxonomy be used in a variety of ways to support effective assessment practices. Both Hattie and Purdie (1998) and Leat and Nichols (2000) found that SOLO taxonomy can be used to guide diagnostic assessment. Smith and Colby (2007) argued for a formative application of SOLO taxonomy to assessment. Others agreed, arguing that SOLO provides a framework for providing students and teachers with descriptive feedback (Hattie & Purdie, 1998; Prakash et al., 2010; Rembach & Dison, 2016; Stewart, 2012). In addition, researchers have found SOLO useful for guiding instructional planning (Hattie & Purdie, 1998; Kusumawathie et al., 2017; Rembach & Dison, 2016; Smith & Colby, 2007). Some authors noted SOLO's value in promoting higher order thinking (Kusumawathie et al., 2017; Newton & Martin,

2013; Stewart, 2012; Wells, 2015). Rembach and Dison (2016) also argued that SOLO provides a framework for monitoring student growth. Thus, the SOLO taxonomy provides a structure for aligning learning outcomes or standards, curriculum, assessment, and instruction and serves as an appropriate cognitive framework for the learning progressions in this study.

Summary and Conclusions

This chapter began with a restatement of the problem and purpose for this research study and a synopsis of the literature search process that was used to gather background information regarding the conceptual framework and major topics of the study. It provided literature search strategies for locating sources and an outline of topics for the literature review. The chapter provided a detailed description of the conceptual framework and the relationship of reflective practice and PCK to the research problem, research question, and sub-questions. The chapter concluded with a synthesis of the literature regarding standards implementation, teacher evaluation, student growth, SLO, learning progressions, and rubrics.

Incorporation of student growth on newly adopted standards in teacher evaluation systems is a complex process that researchers noted can and should be mutually supportive. If legislative changes were made to support improvement for school systems, educator practices, and student learning, then the literature on standards implementation, teacher evaluation system revision, and student growth, indicated a need for greater emphasis on formative application of teacher evaluation systems. Studies of standard implementation and teacher evaluation systems called for further research regarding

structures and tools that support the improvement of teachers' PCK to support student growth. The literature review of teacher evaluation systems and SLO revealed inconsistency in SLO implementation models, leading to the disparity in educator perceptions of their experiences. This study provided information about educators' perceptions of their teacher evaluation experience using a standards-based rubric to monitor student growth, a practice that has yet to be found in the literature. Chapter 3 includes the methodology for this research study.

Chapter 3: Research Method

Introduction

For this basic qualitative inquiry study, I explored teachers' perceptions of using a standards-based rubric to structure SLOs in the teacher evaluation process to support their reflective practice and PCK. I ascertained whether and how teachers perceive the system to support their reflection on standards, and adaptation of assessment, and instruction when standards-based rubrics were incorporated into their teacher evaluation system. This chapter includes an explanation of the choice of basic qualitative inquiry as the research design for the study. The chapter also includes a description of the role of the researcher and the methodology, which includes sections for participant selection, instrumentation, procedures, and data analysis plan. The chapter concludes with a section on trustworthiness and ethical procedures.

Research Design and Rationale

This study addressed the following research questions and sub-questions:

RQ1: How do teachers perceive the experience of implementing SLOs structured by standards-based rubrics to support reflective practice?

RQ2: How do teachers perceive the experience of implementing SLOs structured by standards-based rubrics to support PCK?

SQ1: In what ways do teachers reflect on and adapt their mathematical content knowledge as they implement SLOs structured by standards-based rubrics?

SQ2: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLOs structured by standards-based rubrics?

SQ3: In what ways do teachers reflect on and adapt their instructional tools and practices as they implement SLOs structured by standards-based rubrics?

The concept of interest in this research study was how teachers perceive the use of standards-based rubrics to foster teachers' reflective practice and PCK. I explored educator reflections that stem from using standards-based rubrics and how these reflections influence teachers' PCK. Specifically, teachers were asked about the role the rubric played in supporting their reflective practice regarding their knowledge of mathematics standards, assessment tools and practices, and instruction.

I used a basic qualitative inquiry approach to explore teacher perceptions of the use of standards-based rubrics to monitor student growth in the teacher evaluation system and how the rubrics supported teachers' reflective practice and PCK. According to Denzin and Lincoln (2013), qualitative research is situated in the natural world and involves interpretation to "make the world visible" (p. 7). Interpretative research "assumes that reality is socially constructed" (Merriam & Tisdell, 2016, p. 9). Qualitative researchers acknowledge the existence of multiple realities, understanding that experiences are situational (Lichtman, 2013). A qualitative approach is an inductive process that can produce a rich description of interpretations of experiences with phenomena. Cooley (2013) argued that qualitative research is "the most robust and inclusive means of attempting to understand the complexities of education and processes

of schooling" (p. 248). As the purpose of this study was to examine how teachers perceive the use of standards-based rubrics to monitor student growth as an element of teacher evaluation to influence teachers' reflective practice and PCK, a qualitative approach was suitable. Qualitative inquiry is meaning based, used to understand the motives and qualities of experiences undergone by participants (Eisner, 2017). According to Patton (2015), a basic qualitative inquiry approach can be used to explore the participants' meaning of an experience, process, or event.

Basic qualitative studies are the most common form of qualitative research in educational settings (Merriam & Tisdell, 2016). According to Worthington (2010), the purpose of educational qualitative research is to improve our practice, and the basic qualitative research design is particularly well-suited to obtain an in-depth understanding of effective educational processes" (p. 2). Basic qualitative studies are designed on a foundation of constructivism; they are used to explore the realities that participants construct through their experiences. Merriam and Tisdell (2016) noted that researchers using basic qualitative inquiry seek to find "how people interpret their experiences, how they construct their worlds, and what meaning they attribute to their experiences (p. 24)."

Prior to identifying a basic qualitative inquiry approach for this study, other research types were considered, including phenomenology and case study. Phenomenological studies capture the essence of an experience (Shudak, 2018). By conducting in-depth interviews with participants, it is possible to ascertain the essence of the participants' collective experience with standards-based rubrics. The focus was on teacher perceptions of the experiences implementing SLOs structured around standards-

based rubrics to monitor student growth. Although case study was a potential fit because it allows the opportunity to tell the story of one teacher's experience and reflections using standards-based rubrics, I did not select the approach because it would not support examination of trends among multiple educators' experiences and reflections.

Quantitative methods were not considered, as they do not align with the purpose of the study. In the review of the literature, I found that most studies conducted to the student growth component of teacher evaluation systems have been quantitative in design. A quantitative approach could be used to examine the amount of growth, considering a variety of variables that impact the growth, but would lack the exploration of educators' perceptions of teachers' reflective practice during the SLO process. Such an approach would not have gathered patterns or themes regarding educators' experiences within rubric based teacher evaluation systems that might promote further development of PCK. Therefore, a quantitative approach would not adequately have addressed the research questions for this study.

Basic qualitative inquiry was selected because it was the best approach to address the purpose and research questions of this study. I investigated how teachers perceived using a standards-based rubric to structure SLOs in the teacher evaluation process to support their reflective practice and PCK. The research questions addressed teacher perceptions on their experiences using rubrics to implement the SLO process. The questions explored how the use of standards-based rubrics support teachers to reflect and deepen their PCK as it applies to knowledge of mathematics standards, assessment, and instruction.

Role of the Researcher

I served as an observer in this study. I did not engage in the activities with teachers and evaluators. I did not serve as a teacher or evaluator using standards-based rubrics to monitor student growth in a teacher evaluation system. Therefore, I was not a participant-observer.

Although I interviewed participants outside of my own organization, I had previously met some participants at conferences or other networking events in the past. Therefore, I developed an interview protocol for teachers to maintain focus on the research questions (Creswell, 2014). An interview protocol provides structure to somewhat standardize the interview process (Rubin & Rubin, 2012).

My personal experience with standards, rubrics, and classroom assessment created a potential for bias in interpretation of interview questions and responses. Therefore, I used the technique of interviewing the investigator (Chenail, 2011) to document any potential biases regarding the experience of using standards-based rubrics to monitor student growth. This technique allows a researcher to experience the interview protocol from the participant's perspective. The personal insights gained from this experience helped me to use bracketing to minimize bias (Merriam & Tisdell, 2016). After field testing the interview, I refined the interview protocol with more explicit language in main questions and follow up questions to help maintain focus on the research questions.

Methodology

In this section, I describe the overall methodological approach for investigating the research problem. This basic qualitative inquiry involved semi structured interviews with teachers. The methodology section includes information about participant selection, instrumentation, data collection, and data analysis.

Participant Selection Logic

The participant selection logic provides a framework for sampling and selecting subjects for this study. In this section, I describe the population of teachers, the sampling strategy, and criterion for selection. I also explain the sample size and determining factors for data saturation.

The population for this study included teachers who teach mathematics to K–8 students. I sought 10–12 teachers from a variety of districts (small and large; rural, suburban, and urban) from one midwestern state who have varying levels of teaching experience (early career, midcareer, and late career). Participating educators worked in school districts in which the joint committee has agreed SLO is an approved method for monitoring student growth for teacher evaluation. The participants had used standards-based rubrics to monitor student growth in mathematics in K–8 classroom settings to participate in interviews.

Purposeful sampling was used to identify participants for this study. I used snowball sampling to locate teachers who fit the profile and were interested in participating in this study (Merriam & Tisdell, 2016). I recruited participants by communicating with educators through the Illinois Association of Regional School

Superintendents, Illinois Association for Supervisors of Curriculum Development, and Illinois Council for Teachers of Mathematics networks. I made initial contact by emailing district leaders who use or were considering using SLOs to monitor student growth. I knew these leaders either from past experiences or referrals from contacts. I then scheduled follow-up phone calls to discuss the details of the study with any leaders who responded with interest. These leaders were asked to share the invitation to participate with their district colleagues (Appendix D), who then emailed to notify me of their interest.

I endeavored to select participating teachers to represent rural, suburban, and urban settings. Participants were teachers who used standards-based rubrics to monitor student growth within an SLO structure applicable to each districts' teacher evaluation process. I sent informed consent forms to individuals who fit the profile as teachers of mathematics to K-8 students.

Merriam and Tisdell (2016) acknowledged that determining sample size for basic qualitative inquiry studies depends on the information gathered. I interviewed 10 teachers. Sample size was impacted by the challenge of locating school districts with willing participants due to the transition to remote learning during the COVID-19 pandemic. I was able to locate more teachers in rural and suburban settings than urban settings who were willing to participate. Fusch and Ness (2015) argued that saturation is reached when no new data, themes, or coding emerge, and sufficient information has been obtained for the study to be replicated. Therefore, when interviews revealed no new data, themes, or codes, I determined saturation was reached.

Instrumentation

In this study, I used a basic qualitative inquiry approach. The instrumentation included a questionnaire for demographic and baseline data and semi structured one-on-one interviews. I developed all instruments to align with the conceptual framework and research questions for this study. Instruments were used to gather teacher perceptions from their experience using standards-based rubrics to monitor student growth in mathematics to examine whether such a structure promotes teachers' reflective practice and PCK.

After an extensive review of the literature, I identified a gap in the literature regarding educators' reflective practice regarding content knowledge and PCK in the context of teacher evaluation. No studies I found were conducted to examine the use of standards-based rubrics to monitor student growth in a teacher evaluation system. Also, I found no studies in which researchers considered how standards-based rubrics promote reflective practice or the development of PCK in the context of teacher evaluation. Thus, the instrumentation for this study was designed to address this gap in the literature.

Questionnaire for educators

The purpose of the questionnaire in this study was to gather basic information about the SLOs to better inform the interviewer in preparation for semi structured interviews. The questionnaire provided baseline information regarding teachers' perceptions of student growth and SLO so that I could refer to teachers' questionnaire responses during interviews when discussing their perceptions of any growth in PCK. Demographic data from the questionnaire were used in interview participant selection.

Table 1 provides the purposes for each question of the questionnaire. The questionnaire can be found in Appendix A.

Table 1

Question Purposes for Teacher Questionnaire

	RQ1	RQ2	SQ1	SQ2	SQ3	Other purpose
Question 1						Contact
Question 2						Demographic
Question 3						Demographic
Question 4			X			SLO
Question 5						SLO
Question 6				X		SLO
Question 7	X				X	SLO baseline
Question 8	X	X		X		Baseline
Question 9	X	X	X			Baseline
Question 10	X	X			X	Baseline
Question 11	X	X	X	X	X	Baseline

Teacher interview guide

The interview guide in Appendix B was developed for interviewing participating teachers. Questions 1–2 were intended to build rapport with the participants (Patton, 2015). Question 1 was used to gather demographic data and aligns to RQ2 in that it addresses content knowledge and PCK. Question 2 aligns to both RQ1 and RQ2, as it addresses background knowledge of CCSS-M and was intended to reveal teachers' reflective practice regarding their implementation of the standards.

Several questions (Questions 3–8) addressed the use of data to ascertain the level of formative assessment usage that the use of standards-based rubrics in an SLO process could promote, which addressed RQ2. Questions 3–5 specifically addressed the use of the rubrics in the SLO. Questions 3 and 4 align to both SQ2 and SQ3 in that they addressed both assessment and instructional practice. Questions 5–8 addressed the feedback process for the teacher and students, which addressed both SQ2 and SQ3. Question 8 examined whether and how the teacher experienced a collaborative aspect of the SLO experience and aligns to both SQ2 and SQ3. Questions 9 and 10 targeted the participants' reflections after the experience. Question 9 addressed all four research questions by inquiring about teachers' reflections regarding content knowledge and PCK regarding CCSS-M implementation, while Questions 10 and 11 align to SQ2 and SQ3 by addressing decisions regarding assessment and instruction. Table 2 shows the purposes for each of the questions in the teacher interview guide.

Table 2

Question Purposes for Teacher Interview Guide

	RQ1	RQ2	SQ1	SQ2	SQ3	Other purpose
Question 1						Demographic & establish rapport
Question 2		X	X			
Question 3	X	X		X	X	
Question 4	X		X	X	X	
Question 5	X			X		
Question 6	X	X	X	X	X	
Question 7	X	X	X	X		
Question 8	X	X	X	X	X	
Question 9	X	X	X			
Question 10	X	X	X		X	
Question 11	X	X		X	X	
Question 12	X	X				

Standards-Based Rubric Resources

A sample standards-based rubric is shown in Appendix C. In an SLO process, teachers select standards and accompanying performance level descriptors from the standards-based rubrics to develop their SLO plans, and evaluators approve the selections.

Other Data Sources

I also recorded field notes during and after each interview. Field notes included my observations of participant behaviors, such as hesitations, facial expressions, or gestures. To triangulate the data, I compared the field notes with the data from interview transcripts and questionnaire responses to address the four research questions addressed in this study. Field notes, interview transcripts, and questionnaire responses all addressed

the concepts of reflective practice and PCK in the context of teacher evaluation systems to ensure content validity.

Field Test

I designed a field test to determine interview questions that would elicit teacher reflections regarding the process of monitoring student growth with standards-based rubrics. The field test allowed me to practice one-on-one interviews both in person and on the telephone. I drafted interview questions for teachers and administrators and then recruited participants to field test the interview questions. I invited four teachers to participate in the field test. All participants were selected because they had experience using standards-based rubrics to monitor student growth. Teachers were invited via email and informed consent was obtained through email as well. The three elementary teachers all taught third grade at the same suburban school. Although all three worked on the same team, they were invited separately and interviewed separately at an off-site location to maintain confidentiality. The middle school teacher taught seventh grade math at a rural school.

I recorded and transcribed each interview using the phone application NoNotes.com. I coded the transcripts using open coding and analytic coding using the qualitative data analysis program NVivo (Saldaña, 2016). After completing two teacher interviews, I made refinements to the interview guide to elicit more data regarding teachers' knowledge of standards and practices and how these aspects of PCK may have changed during the SLO process. I used the updated interview guide with in the last two teacher interviews. The field test provided me with insights regarding key terminology

that elicit teacher reflections regarding PCK and adaptations to teacher practices. I noted which questions led teachers to share about how the rubrics influenced their assessment and instructional practices. I also noted which questions prompted teachers to share about their experience with the rubrics helping them to learn about the mathematics content and about their students' learning.

This field test of interview questions was conducted to develop and refine the interview questions. Participants did not complete the questionnaire prior to the interviews. I developed the questionnaire after conducting the field test, noting that I had limited data regarding educator practices and perceptions of standards implementation, SLO, and student growth prior to the interviews. I was aware of the SLO content for the three elementary teachers prior to conducting their interviews, which helped me to customize the questions to the specific content knowledge and PCK targeted in the SLO. However, that information had to be elicited during the interview with the middle school math teacher. I noted that having the information provided clarity that was absent in the middle-school math teacher interview, which led me to create the questionnaire.

Procedures for Recruitment, Participation, and Data Collection

The participant sample included elementary teachers in Illinois who educate K–8 students in mathematics. Participants were working in school districts that have approved SLO as a method for monitoring student growth for their teacher evaluations or were considering using SLO for this purpose. Participants included educators working in rural, urban, and suburban environments to compare experiences across contexts.

I first emailed contacts I had made through state leadership and mathematics organizations (Illinois Association of Regional School Superintendents, Illinois Association for Supervisors of Curriculum Development, and Illinois Council for Teachers of Mathematics) to inquire whether their school districts have approved SLO as a method for monitoring student data. Those that responded in the affirmative were contacted by phone to discuss participation in the research study. I also contacted the Illinois Association of Regional School Superintendents to obtain contact information for administrators trained by Regional Offices of Education for re-certification as teacher evaluators. Administrators were informed that participating teachers would not be compensated for their participation.

I administered the teacher questionnaire using a securely constructed Google Form that I developed. Data from the questionnaire was used to select interview participants. It was analyzed regarding teachers' perceptions and practices. I conducted the interviews and recorded using Zoom. Interviews lasted approximately 15–35 minutes and were held via Zoom. I recorded field notes after each interview. Interviews were transcribed by a transcription service, and transcripts were checked and summarized by the researcher and sent to participants for member checking. Within one week of completing the interview, the participant received a letter of thanks for their time and effort. Once participants have confirmed the accuracy of the transcript, they exited the study. Once the study was completed, I provided participants with a summary of the outcomes of the research and a link to the full document.

Data Analysis Plan

Researchers have articulated various approaches to qualitative data analysis for interviews. Rubin and Rubin (2012) described steps that include transcription, inductive coding, deductive coding, summarization, integration of ideas across multiple interviews, identification of trends or themes, and generalization beyond the individuals involved.

Merriam and Tisdell (2016) describe qualitative data analysis as inductive and comparative (p. 201). Elo et al. (2014) argued that qualitative content analysis could be both inductive and deductive. Merriam and Tisdell (2016) described that qualitative data analysis follows a logical sequence of (a) discovery, (b) discovery (inductive) and verifying (inductive and deductive), and (c) testing and confirming (primarily deductive) (p. 211).

Based on the approach described by Merriam and Tisdell (2016) and Rubin and Rubin (2012), for this study, I initially coded interviews using open coding and analytic coding with an inductive approach. I then examined the excerpts as sorted by codes and sub-codes that I organized using the QDA program NVivo. Once codes were established by analyzing the first few interviews, I applied these codes to the analysis of ensuing interview transcripts, using both inductive and deductive reasoning. As new codes were added in ensuing transcripts, I revisited the earlier transcripts to review for the additional codes. Throughout the coding process, I grouped related codes into categories. When I determined that data saturation has been reached and had collected sufficient information for the study to be replicated, I examined themes that emerged in the data. I included

relevant excerpts from the interview statements that have been aligned to each of the themes.

Issues of Trustworthiness

According to Shenton (2004), the trustworthiness of qualitative research can be established by attending to these four criteria: (a) credibility, (b) transferability, (c), dependability, and (d) confirmability. Elo et al. (2014) recommended that trustworthiness should be addressed by attending to these criteria in the preparation, organization, and reporting phases of a study (p. 3). This section includes descriptions for how each of these criteria were addressed in this study.

Credibility

Credibility ensures that the phenomenon of interest for the study is accurately represented. Tracy (2010) advised that credibility can be better established by showing through descriptive language rather than telling the reader what to think. This element was addressed in this study by focusing on the participants' experience and reflections from using standards-based rubrics for monitoring student growth. It was addressed through triangulation of the data from the questionnaires, with researchers' field notes, and interview data from the teacher perspectives.

Transferability

Transferability represents the idea that the context has been clearly articulated. The context for this practice study is monitoring student growth for mathematics as part of the teacher evaluation process. Shenton (2004) argued that contextualized factors make it difficult to ensure transferability in qualitative research. However, if I share detailed

information about the contexts for a qualitative study, readers may find connections to a setting that may transfer to their own contexts. For this study, I recruited participants from more than one district to support possible transferability for educators in a variety of settings, such as rural, suburban, and urban communities. In addition, I attempted to support transferability by clearly articulating participating teachers' years of experience and describing the context for each participant to support readers to make connections regarding early-career, mid-career, or late-career experiences.

Dependability

Dependability addresses the idea that the research procedures have been clearly presented so that the study can be replicated. The participant invitation, teacher questionnaire, and interview guide have been provided so that other researchers can replicate the recruitment and interview process for data collection. In addition, the data analysis plan is articulated so that researchers can use the same coding strategy.

Confirmability

Confirmability ensures that the findings stem from the thoughts and experiences of the participants and not from the biases or prejudices of the researcher (Toma, 2011). I made sure to focus the interviews on the participants' contributions to the dialogue instead of my own. I restated or rephrased to confirm the participants' messages but did not share my own experiences during the interview process. I reminded some teacher participants of their responses in the questionnaire when asking them to share reflections on changes in their PCK. Therefore, the transcripts contained the perceptions and experiences that participants shared and not my own. I also compared each participant's

responses on both the questionnaire and the interview transcripts to my interview field notes to triangulate the responses.

Ethical Procedures

Participation in this study was voluntary. I obtained permission from district administrators to contact teachers. No participants in this study were under the age of 18, and all participants were informed that they could leave the study at any point. I took care to make participants comfortable and verified the confidentiality necessary to support accuracy in data collection. Participants were informed of the purpose of this study during the recruitment process. The purpose was reiterated in email communication and at the beginning interviews. All participants received an informed consent form by email and acknowledgment of consent was collected verbally at the beginning of each interview. If participants withdrew from the study, their choice was kept confidential to prevent any possible negative repercussions.

Confidentiality and privacy of participants was upheld by using pseudonyms in for all participants (Janesick, 2011). I ensured the security of all files by using password-protected telephone and computer. Aside from sharing transcripts with each participant for member checking, I only shared data with the dissertation committee members. I obtained approval from the Walden University Institutional Review Board (IRB) before beginning the study (Approval #02-25-20-0629438). All data will be kept for 5 years and then securely destroyed to protect participants' confidentiality and privacy.

Summary

This chapter included the research design and rationale for this study, the role of the researcher, instrumentation, data collection, data analysis, trustworthiness, and ethical procedures. I used a basic qualitative inquiry approach to explore whether and how the use of standards-based rubrics in SLO to monitor student growth for teacher evaluation supported teachers' reflective practice and PCK. Participants were teachers who have used standards-based rubrics to monitor student growth for teacher evaluation. They were recruited through purposeful sampling that ensures maximum variability. I gathered data through semi structured interviews, which were analyzed using qualitative data analysis with open and analytic coding. I attended to confidentiality and ethical practices that respect participants rights throughout the process. Results of this study will be discussed in Chapter 4.

Chapter 4: Results

Introduction

The purpose of this study was to explore teacher perceptions of their experiences using standards-based rubrics to monitor student growth in an SLO process as an element of their teacher evaluation. In this chapter, I present an overview of the nature of this qualitative study including its setting and participant demographics. This chapter also includes an explanation of the data collection and analysis processes used to complete this study. In addition, Chapter 4 includes evidence of trustworthiness, a discussion of results and a summary of the findings. In this study, I used teacher surveys and semi-structured interviews to address the following research questions:

RQ1: How do teachers perceive the experience of implementing SLOs structured by standards-based rubrics to support reflective practice?

RQ2: How do teachers perceive the experience of implementing SLOs structured by standards-based rubrics to support PCK?

SQ1: In what ways do teachers reflect on and adapt their mathematical content knowledge as they implement SLOs structured by standards-based rubrics?

SQ2: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLOs structured by standards-based rubrics?

SQ3: In what ways do teachers reflect on and adapt their instructional tools and practices as they implement SLOs structured by standards-based rubrics?

Field Test

I conducted a field test to refine the interview questions for this study. There were four participants in the field test, representing two school districts. One district was in a suburban setting and the other was in a rural setting. One participant was a middle school mathematics teacher, while the other three participants were elementary classroom teachers. After conducting the field test, I refined the interview questions so that I would elicit data more specific to the research questions.

Table 3

Field Test Participant List

Pseudonym	Setting	Experience	Grade Level	Teacher Type
Ms. W	Suburban	20 years	3rd grade	Elementary classroom
Ms. X	Suburban	21 years	3rd grade	Elementary classroom
Mr. Y	Suburban	12 years	3rd grade	Elementary classroom
Ms. Z	Rural	28 years	7th grade	Middle-school mathematics

Study Setting

This study occurred in one midwestern state of the United States. Participants were elementary and middle-school teachers practicing in public-school settings during the study. The 10 interviewed teachers taught in four counties, four school districts, and seven public schools. Five teachers taught in suburban settings and five taught in rural settings. Seven of the teachers taught at kindergarten through Grade 5 schools, one taught at a pre-kindergarten through Grade 5 school, and two taught at a Grade 5 through 8 school. Six teachers taught in schools with more than 50% of the students classified as

low income, two taught in schools with between 30% and 49% of students classified as low income, and two taught in schools with fewer than 10% of students classified as low income. Two teachers taught in schools with more than 50% of the students classified as English language learners (ELLs), one teacher taught in a school with between 20% and 29% ELL students, three teachers taught in schools with between 10% and 19% ELL students, and two teachers taught in schools with less than 10% ELL students. Table 4 summarizes the number of teachers working in settings according to percentages of students classified as low-income and ELL.

Table 4

Number of Participants Working in Settings by Selected Student Populations

Student population	Fewer than 10%	10-19%	20-29%	30-39%	40-49%	50% or more
Low-income students	2			2		6
ELL students	3	4	1			2

Demographics

Eleven teachers responded to the call for participation. One volunteer did not complete SLOs for her teacher evaluation as she had originally planned and was excluded from the study. One of the participants I interviewed realized late in the interview, as she reflected on her experience, that the rubric she used did not include the standards and was, therefore, not a standards-based rubric. Because using a standards-based rubric was one of the criteria for participation in this study, I could no longer consider her a viable

participant. Therefore, Ms. A exited the study and I excluded Ms. A's questionnaire and interview responses from the data set.

Of the remaining nine participants, six were classified as general education classroom teachers, two were classified as middle-school mathematics teachers, and one was classified as both a special education and a bilingual teacher. There were eight female participants and one male participant. Three participants had fewer than 5 years of experience, four participants had 6 to 15 years of experience, and two participants had 16 to 25 years of experience in education. Each participant was given a pseudonym using an alphabetical system with participant one being given the pseudonym of Ms. A continuing to the letter J.

Table 5

Participant List

Pseudonym	Setting	Experience	Grade level	Teacher Type
Ms. B	Rural	4 years	4th grade	Elementary classroom
Ms. C	Rural	10 years	4th grade	Elementary classroom
Ms. D	Rural	3 years	4th grade	Elementary classroom
Ms. E	Rural	14 years	2nd grade	Elementary classroom
Ms. F	Rural	4 years	K–5	Special education & bilingual
Ms. G	Suburban	23 years	3rd grade	Elementary classroom
Mr. H	Suburban	17 years	3rd grade	Elementary classroom
Ms. I	Suburban	12 years	6–8th grade	Middle-school mathematics
Ms. J	Suburban	11 years	6–8th grade	Middle-school math coach

Participant Demographics

I acquired participant demographics (Table 5) through a Google Form link that each participant submitted (Appendix A) prior to their interviews. Nine female and one

male teacher were interviewed for the study. Interviews occurred between August 2020 and May 2021.

Ms. B. was a general education classroom teacher in her fourth year of teaching in the district. She taught fourth grade all 4 years in a rural public-school setting. Prior to having her own fourth-grade classroom, she worked in another district as a paraprofessional doing mathematics intervention for 1 year. She also worked as a long-term substitute in a first-grade class for approximately half of a year before that.

Ms. C was a general education classroom teacher with 10 years of experience in a rural public-school setting. At the time of the study, she was in her fifth year of teaching fourth grade. Prior to that, she worked as a substitute teacher for 2 years before becoming a seventh-grade literature and science teacher in the same school district.

Ms. D was a general education classroom teacher in her third year of teaching. She has spent her entire career teaching fourth grade in the same rural public school. At the time of the study, Ms. D was also hosting a student teacher in her classroom.

Ms. E. was a general education classroom teacher in her 14th year teaching at a rural public school. At the time of the study, she was teaching second grade. Prior to this role, she taught English to students, ages 2 to 92, in France.

Ms. F was a special education and bilingual teacher at a rural dual language public school. At the time of the study, she was in her fourth year of teaching. She was working with students in each grade, kindergarten through Grade 5. In prior years, she has worked with fewer grade levels, but has always served as a special education bilingual teacher at the same dual language school.

Ms. G was a general education classroom teacher in her 23rd year of teaching at a suburban public school. At the time of the study, she was teaching third grade. Previously, she taught fourth grade. She also obtained her reading specialist degree.

Mr. H was a general education classroom teacher with 17 years of experience total, 11 years in his current district. He taught third grade in a suburban public-school setting. Prior to teaching third grade, he also taught in second-, fourth-, and fifth-grade classrooms.

Ms. I was a general education classroom teacher for middle-school mathematics. She taught sixth, seventh, and eighth grade in a suburban school district. She taught for 12 years at a variety of grade levels. Prior to becoming a teacher, she worked for several years as an engineer.

Ms. J was a mathematics instructional coach in a middle-school setting. She coached teachers of Grades 5–8. Prior to this role, she taught sixth–eighth-grade mathematics in the same school for 8 years. She had taught middle-school math for 2 years in a large urban school district before moving to her current district.

Survey results provided information regarding the tools each participant used to monitor student growth for their teacher evaluation. Participating teachers used a standards-based rubric specific to the standards chosen for their SLO (see Table 4). All participants responded that they administered a pre-assessment to establish baseline data for their SLO. Some noted that the district provided a common pretest and one described less formal assessment instruments, such as informal assessments and observations.

Table 6

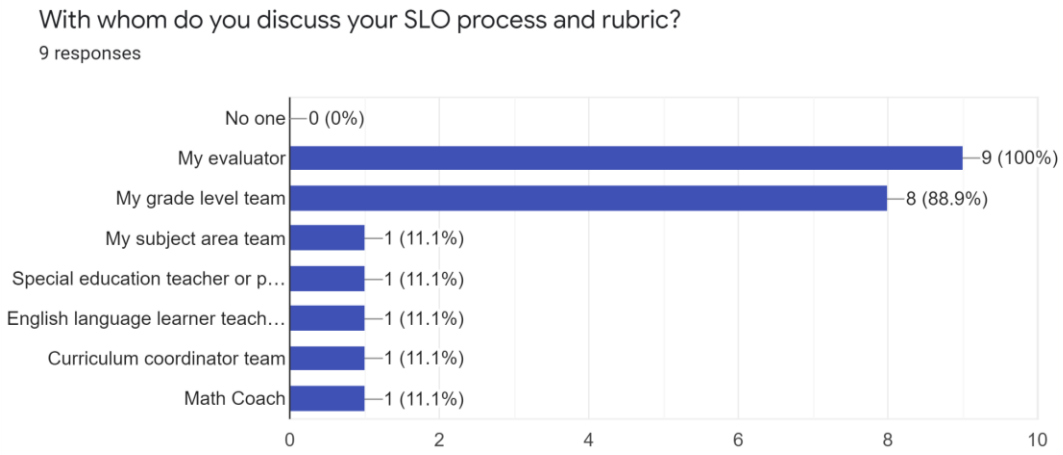
Student Learning Objective Standards-Based Rubrics Used by Participants

Pseudonym	Standards for student learning objectives	Rubric source
Ms. B	4.OA.1, 4.OA.2, 4.NBT.5	District developed tool
Ms. C	4.OA.1, 4.OA.2, 4.OA.3, 4.NBT.5	District developed tool
Ms. D	4.OA.1, 4.NBT.5	District developed tool
Ms. E	2.MD.7	District developed tool
Ms. F	2.NBT.1, 2.NBT.2, 2.NBT.3, 2.NBT.4, 2.NBT.8	District developed tool
Ms. G	3.NF.2ab, 3.NF.3abd	District provided tool
Mr. H	3.OA.1, 3.OA.2, 3.OA.3, 3.OA.4, 3.OA.5, 3.OA.6, 3.OA.7, 3.OA.8	Example rubric provided to district
Ms. I	6.NS.1	Example rubric provided to district
Ms. J	7.NS.1, 7.NS.2, 7.NS.3	Example rubric provided to district

All participants responded that they discussed their SLO process and rubric with at least one other person (see Figure 2). All nine participants discussed their SLO with their evaluators. Eight participants responded that they also discussed their SLO process and rubric with grade-level team members and one middle-school participant responded that the SLO process and rubric were discussed with her math department team. The following colleagues were also consulted by a member of this participant pool, receiving one response each: special education teacher or paraprofessional, ELL teacher, math coach, and curriculum coordinator team.

Figure 2

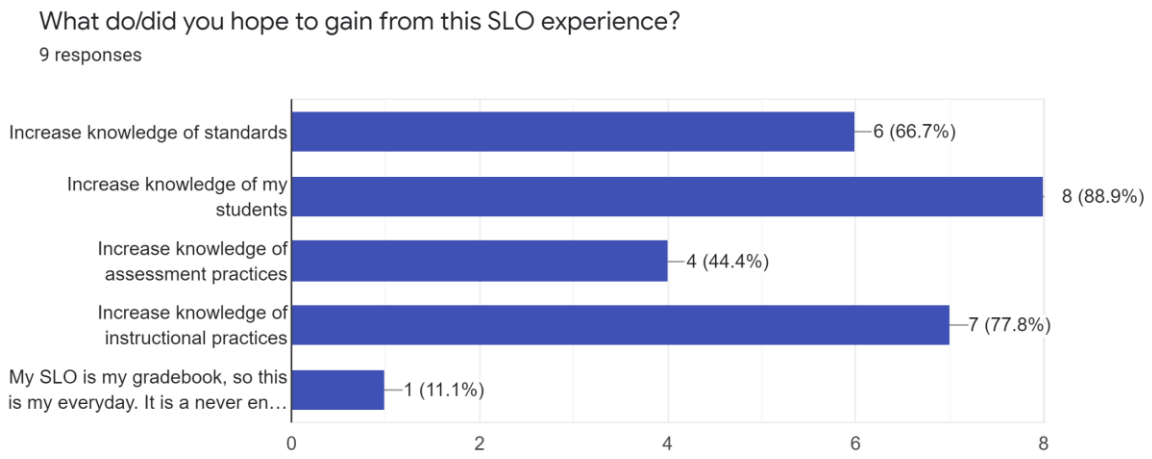
Participant Responses to Question 7



Participant responses to Question 11 regarding their goals for the SLO experience are shown in Figure 3. Some participants indicated that they had one goal, while others had multiple goals for the experience. Six of the participants indicated a desire to increase their knowledge of standards. Eight participants indicated a desire to increase their knowledge of their students. Four participants indicated a desire to increase their knowledge of assessment practices, and seven participants indicated a desire to increase their knowledge of instructional practices. Ms. D responded that her SLO is her gradebook, so the everyday process is "a never-ending experience of collecting evidence of her students' knowledge and skills."

Figure 3

Participant Responses to Question 11



Data Collection

I received notice from IRB that my study was approved in February 2020 (approval #02-25-20-0629438). On February 26, 2020, I requested the contact person for each partner organization to distribute an invitation to participate to teachers of kindergarten through eighth grade mathematics students in the respective school districts (Appendix D). Shortly after the invitations were distributed, all the school districts in the state were required to transition to remote learning in response to the COVID-19 pandemic. One participant from a suburban setting volunteered in March 2020. Once consent was obtained, she was sent the teacher questionnaire. However, she was not interviewed until September 2020, due to the amount of work she needed to do to transition to remote learning.

I reconnected with the partner districts again in August 2020 to redistribute the invitation to participate. Three of the partner districts allowed me to field questions about

participating in the study in staff meetings, one in person and two virtual. Seven participants volunteered for the study between August and October 2020 from two different districts. However, two of those volunteers had not yet completed their SLOs. As five of the participants were from rural public schools in the same county, I reached out to contacts from other regional offices to recommend other sites using rubrics to monitor student growth. In February 2021, one additional site in a suburban setting provided two more participant volunteers. In May 2021, I was able to interview two middle-school mathematics teachers: the first had volunteered in the fall and the other was referred to me by a participant.

Once consent was obtained via email, each participant completed a teacher questionnaire. The tool provided demographic information and baseline data regarding their SLO. This information was used to determine whether participants should be interviewed and provided information to which I could refer for specificity of questions in the interview (Appendix A). I then scheduled interviews with each participant.

All interviews were conducted via Zoom and recorded in Zoom. Interviews ranged in length from 12 to 34 minutes, with an average of 21 minutes. I then uploaded each recording to Nonotes.com for transcription. I also watched the video recordings to make field notes for each interview. Once I received the transcripts back, I checked each while re-viewing the video recording. After making my own corrections, I emailed each transcript to the participant for transcript checking. Participants were invited to make corrections, revisions, or clarifications. Once I received confirmation from participants that the data reflected their perspectives, I summarized each interview.

Data Analysis

From September 2020 to May 2021, I conducted ten semi structured interviews using the interview guide (Appendix B). I created a summary of each of the nine interviews that was included in the study prior to coding. The summaries allowed me to reflect on the content of each interview in its entirety and to anticipate prospective codes. Each interview was hand-coded in the first cycle of coding using descriptive coding in the QDA program NVivo. After coding each transcript, I applied the codes that I had identified to each of the ensuing transcripts, adding additional codes as they emerged in the data. As any new codes were identified, I conducted a second cycle of coding prior transcripts to see if the new codes applied. I then examined codes for redundancy, merging codes where appropriate. When no new codes emerged, I determined that data saturation had been reached. After the second cycle of coding, I organized the codes into categories that aligned to the research questions for this study (see Table 6). I identified themes through the iterative process of (a) Coding responses made by participants in their interview and questionnaire, (b) Eliminating redundancies and categorizing related codes, (c) Clustering categories into subthemes, and (d) Grouping subthemes into the overarching themes.

Table 7

Example of Data Analysis Process for Identifying Themes From Codes

Excerpt	Code	Category	Subtheme	Theme
We **meet with our grade level peers**, and we discuss, usually our summative data.	Discuss with grade level	Discuss data with peers	Collaborative dialogue with peers	Fosters collaborative dialogue and descriptive feedback
I can actually give **legitimate feedback** and tell them what they were able to do.	Feedback	Descriptive feedback	Descriptive feedback to students	
"This is what your child is currently doing." Tells me a lot more than, "Oh, he got a B or a C or a D."	Communicate with parents	Descriptive feedback	Descriptive feedback to parents	
So then, that helps me **focus** on like, "Oh, these are the standards I need to focus on. These are the skills and strategies that we need to focus on." And plan my lesson plan that way.	Focus - instruction		Focus Instruction on standards	
We **met as a grade level team to evaluate the standards** and see how we are going to assess students	Team evaluate standards together	Common understanding	Promotes common understanding of standards	Promotes standards-based focus
I feel like I learned them in and out because I had a partner who was very, she really **focused on the verbiage of standards**. So, we would sit and have conversations of, well it says that they have to, they have to demonstrate it. What does demonstrate mean?	Unpack the standard	Knowledge of standards	Promotes depth of understanding of standards	

Evidence of Trustworthiness

Credibility

I ensured credibility through multiple methods. The first method was to triangulate the data from the questionnaire, transcript, and field notes for each participant. I addressed any discrepancy by checking with the participant to verify their intended response. The second method was using the strategy of transcript checking. Participants reviewed transcripts of their interviews to verify their accuracy. Each participant was emailed an electronic copy of the transcript to verify that it truthfully reflected his or her perceptions of the experience. My analysis included direct quotations from the transcripts to show the reader the results rather than to describe in my own words, as recommended by Tracy (2010).

Transferability

I addressed the issue of transferability by recruiting participants from different settings. I was able to gather perceptions from participants working in both rural and suburban settings. Participants ranged in levels of experience from 3 to 23 years. Participants also taught a variety of grade levels. Two participants taught primary students, 6 participants taught intermediate elementary students, and 2 participants taught middle school students. I have provided a description of each participant's background and experience to support readers in finding connections to their own context.

Dependability

I addressed the issue of dependability by including a detailed description of the research procedure so that the study could be replicated. I also included the participant

invitation, teacher questionnaire, and interview guide so that other researchers can replicate the recruitment and data collection process. I articulated the iterative data analysis process of interpreting data from codes to categories to themes so that another researcher can use them in a future study (see Table 7).

Confirmability

I addressed the issue of confirmability by ensuring that the findings stem from the ideas shared by participants and not my own. I confirmed participants' questionnaire responses by referring to their answers during the interview. I also ensured confirmability by restating or rephrasing the participants messages without sharing my own experiences. I took field notes immediately following each interview by viewing the recording and compared these notes to the transcripts to further triangulate the data. If I found any discrepancies between notes, responses, and transcripts, I consulted the participant to verify their intended response.

Study Results

The conceptual frame for this study was a dual lens of reflective practice (Schön, 1983) and PCK (Shulman, 1986). Both of these frameworks informed these results. The teacher questionnaire (Appendix A) and interview questions (Appendix B) were based on the constructs from these two frameworks and were used to guide the analysis of these results. Table 8 includes the themes and subthemes identified from the data and their aligned research questions.

Table 8

Themes and Subthemes That Addressed Research Questions

Themes and subthemes	Research questions
Theme 1: Fosters collaborative dialogue and descriptive feedback. Fosters collaborative dialogue with evaluator. Fosters collaborative dialogue with peers. Fosters collaborative dialogue with specialist teachers (Special education, English language learner, MTSS, etc.) Supports teachers to provide descriptive feedback to students. Supports teachers to communicate with parents.	RQ1: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support reflective practice? RQ2: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support pedagogical content knowledge? SQ2: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLO structured by standards-based rubrics?
Theme 2: Promotes standards-based focus. Promotes common understanding of standards. Promotes teachers' depth of understanding of standards and prerequisites. Promotes students' depth of understanding of mathematics. Supports transition to standards-based system	RQ1 RQ2 SQ1: In what ways do teachers reflect on and adapt their mathematical content knowledge as they implement SLO structured by standards-based rubrics?
Theme 3: Supports evidence-based assessment. Increases teachers' knowledge of student learning. Fosters evidence-based assessment of student growth. Fosters evidence-based assessment to establish baselines for each student. Fosters evidence-based assessment of mastery.	RQ1 RQ2 SQ2: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLO structured by standards-based rubrics?
Theme 4: Supports student-centered instruction. Guides grouping of students. Guides targeted intervention. Guides differentiation of instruction. Guides reflection on instructional tools and strategies. Fosters mathematical language.	RQ1 RQ2 SQ3: In what ways do teachers reflect on and adapt their instructional tools and practices as they implement SLO structured by standards-based rubrics?
Theme 5: Encourages students' reflective practice. Encourages students to use feedback. Encourages students to self-monitor their growth.	RQ1
Theme 6: Cultivates a positive teacher evaluation experience. Supports in becoming a better teacher. Improves the evaluation experience.	RQ1 RQ2

Research Question 1

Research Question 1 was as follows: How do teachers perceive the experience of implementing student learning objectives (SLO) structured by standards-based rubrics to support reflective practice? To answer RQ1, questions 2-12 of the Teacher Interview Guide (Appendix B) were asked. Teachers shared different types of reflections, some of which aligned to RQ2: 1) Reflections on standards (RQ2, SQ1), 2) Reflections on assessment tools and practices (RQ2, SQ2), 3) Reflections on instructional tools and practices (RQ2, SQ3), 4) Reflections on student learning, and 5) Reflections on the teacher evaluation experience. Thus, all 6 themes that emerged in the data address RQ1. All 9 participants contributed to the data for reflective practice. Table 9 includes information about the number of participants who contributed to each of the 6 themes and the number of mentions for each theme.

Table 9

Themes for RQ1

Theme	Number of participants	Mentions
Fosters collaborative dialogue and descriptive feedback	9	39
Promotes standards-based focus	9	38
Supports evidence-based assessment	9	36
Supports student-centered instruction	9	27
Encourages students' reflective practice	9	23
Cultivates a positive teacher evaluation experience	9	17

Research Question 2

Research Question 2 was as follows: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support PCK? The first sub-question for RQ2 was: In what ways do teachers reflect on and adapt their mathematical content knowledge as they implement SLO structured by standards-based rubrics? Questions 2, 4, and 6-10 of the Teacher Interview Guide (Appendix B) addressed sub-question 1. The theme of promoting a standards-based focus addresses the first sub-question of RQ2. Table 10 includes the number of participants and mentions that contributed to this theme and its subthemes.

Table 10

Theme and Subthemes for SQ1 Mathematical Content Knowledge

Theme/subtheme	Number of participants	mentions
Theme 2: Promotes standards-based focus	9	38
Promotes teachers' depth of understanding of standards and prerequisites.	9	23
Promotes common understanding of standards.	8	13
Promotes students' depth of understanding of mathematics.	8	12
Supports transition to standards-based system.	5	9

The second sub-question for RQ2 was: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLO structured by standards-based rubrics? Questions 3-8, and 11 of the Teacher Interview Guide (Appendix B) addressed sub question 2. Two themes emerged from the data to address the second sub-question: (a) fosters collaborative dialogue and descriptive feedback, and (b) supports evidence-based assessment. Table 11. addresses the number of participants and mentions that apply to the themes and subthemes for assessment tools and practices.

Table 11

Themes and Subthemes for SQ2 Assessment Tools and Practices

Theme/subtheme	Number of participants	Mentions
Theme 1: Fosters collaborative dialogue and descriptive feedback	9	39
Supports teachers to provide descriptive feedback to students.	9	15
Supports teachers to communicate with parents	6	8
Theme 3: Supports evidence-based assessment	9	36
Fosters evidence-based assessment of student growth	9	19
Increases teachers' knowledge of student learning	9	13
Fosters evidence-based assessment to establish baseline	9	12
Fosters evidence-based assessment of mastery	5	9

The third sub-question for RQ2 was: SQ3: In what ways do teachers reflect on and adapt their instructional tools and practices as they implement SLO structured by standards-based rubrics? Questions 3, 4, 6, 8, 10, and 12 of the Teacher Interview Guide (Appendix B) addressed sub-question 3. The theme of supporting student-centered instruction emerged from the data regarding how the use of standards-based rubrics in the SLO process supported teachers to adapt their instructional tools and practices. Table 11 addresses the number of participants and mentions that apply to the theme and subthemes for instructional tools and practices.

Table 12

Themes for SQ3 Instructional Tools and Practices

Theme	Number of participants	Mentions
Theme 4: Supports student-centered instruction	9	27
Guides reflection on instructional tools and strategies	9	19
Guides targeted intervention	8	16
Guides grouping of students	8	10
Guides differentiation of instruction	6	9
Fosters mathematical language	4	4

Theme 1: Fosters Collaborative Dialogue and Descriptive Feedback

All nine participants experienced reflective dialogues with at least one other educator. Five subthemes emerged in the data for this theme: (a) fosters collaborative dialogue with evaluators, (b) fosters collaborative dialogue with peers, (c) fosters collaborative dialogue with specialist teachers (Special education, English language learner, MTSS, etc.), (d) supports teachers to provide descriptive feedback to students, and (e) supports teachers to communicate with parents. All teachers experienced an increase in descriptive feedback to students, although only those who had been using the rubrics for more than one year expressed that their feedback to parents had become more descriptive.

Subtheme 1: Fosters Collaborative Dialogue with Evaluators

Participants described their experiences reflecting on student data with their evaluators. Some of the conversations included reflective questioning from the evaluator, as in Ms. B's experience,

> She will usually ask how did the unit go? She'll ask—we usually tend to focus on the kids who didn't meet the goal—and we'll say, like why do you feel they didn't

> meet the goal? What could we do better? What could we do to better meet their needs? Is it a case of we just need more evidence? Things like that.

Other participants shared that their evaluators provided suggestions to help them address concerns; as Ms. F noted,

> They definitely do give us feedback. I'm able to go with them. Sometimes I do get stuck on like, "Hey, I've tried all this and I can't get them to this level." And they're very good about giving us constructive feedback and helping us plan or come up with a potential plan of next steps based on the data and stuff we have collected.

Ms. J reflected how the experience meeting with her evaluator changed from the former process of using mirrored pretest/post-test to this rubric based process, stating,

> In sharing the growth data at the very beginning, I sat down with my evaluator. We looked at the rubric together and had discussions about how we would show the growth through the rubric. And then, with each student in the post-assessment, they had a rubric attached and we were able to show where they started and then how they moved through. And with it being a new process, it was definitely a more complex discussion with my evaluator as opposed to before where I would just fill out a spreadsheet with their numbers in there and how many they got right or wrong and the percentage who grew. Now it was more of a conversation.

Overall, teachers shared that implementing the rubric based SLO fostered collaborative dialogue with their evaluators.

Subtheme 2: Fosters Collaborative Dialogue with Peers

All nine participants described how they share and discuss their students' performance with colleagues. Some shared how they meet regularly to reflect on data with peers. Ms. D described how she meets informally with a teaching partner to discuss data-informed instruction, asserting

> I would discuss with my teammate about how I felt like those—because I have been really fortunate to have such a great teammate that we work really collaboratively on those things. So, I'm not necessarily talking with my administrator about the instructional choices. It's more so talking to my teammate about the instructional decisions we make together and changing it based on the data.

Ms. I described how she and the other sixth through eighth grade mathematics teachers in her building meet as a department saying,

> We meet once a week to talk about this information, the information that we've seen, which kids have kind of mastered it, which kids are still kind of struggling. And then at the time, we would say, "Hey, did you notice on step seven, this is in here. That's a very awesome tool. It really helped my kids get a better understanding of this."

In Ms. E's district, grade level teachers meet periodically across the district to reflect on data after completing a unit. She noted,

> We meet with our grade level peers, and we discuss, usually our summative data. And if our class as a whole was struggling, why did we think they were

> struggling? What could we do to improve on that next time, or go back and reteach each to help them reach the levels that they need to be successful in the next grade level? So those constant conversations of reflection, actually.

Ms. G met with the other grade level teachers in her building to address the diverse needs of their classes. However, she also noted that data is shared among other grade level teachers stating,

> I would say that that's a lot of times where the conversation kind of starts, because those are usually the ones you're most concerned about. On either end, those ones that already know everything, and they're bored to tears in your class because they already know it all, and then you have the ones that are having a hard time making sense of anything. So, the conversation usually starts around the outliers, and then kind of comes in towards the middle. And then we also use that information to articulate across grade levels as well.

In sum, participants shared that the process of gathering data to monitor growth on the standards-based rubric led them to discuss data with peers; as Ms. J noted, “We try to feed off one another.”

Subtheme 3: Fosters Collaborative Dialogue with Specialist Teachers

Some teachers described how using the rubric provided opportunities to collaborate with intervention, English language, and special education staff to serve students’ needs. Ms. G noted how sharing data supported her to collaborate with team teachers.

> All of the teachers that pushed into the classroom are privy to that information as well. And then we team-teach together. And they work on—We all have our specialties. So, you have the EL, the special Ed teachers, and reading specialists, or whatever the case might be. I know we're talking about math right now. But then they come in and they support in that sense, yes.

Mr. H described how sharing information led him to collaborative dialogue with intervention services team members saying,

> When it's more MTSS or intervention service, we're starting to look at how to work together with kids that have needs and intervention. The timing on that and delivery of it is a challenge, but that's part of our conversation and it does happen. So that's good.

Ms. F, special education and bilingual teacher, described how sharing data allowed her to support classroom teachers in the interpretation of evidence collected from special education students by stating,

> Sometimes, a student will be able to improve in that standard, but they'll need some accommodations for it. I have run into teachers who are like, "Well, he was able to do this for me this way, but then he wasn't able to do it this way." I'm like, "But technically he's still showed you that he could do it. It's just with an accommodation." And I think that's one of the things that they need to keep in mind that all students learn differently. And just because they aren't able to show to you one way or show mastery, a specific way doesn't mean that they have not mastered it.

Therefore, some participants engaged in collaborative dialogue with specialist teachers as they shared data collected for their rubric based SLO.

Subtheme 4: Supports Teachers to Provide Descriptive Feedback to Students

All teachers referred to the rubric as a resource for supporting their feedback. Some teachers emphasized how the rubric's language supported them in their conversations with students about their work. Ms. G noted,

> I could look at the progression, and I could talk to them about where they were at, and where I knew that they were capable of going when applying themselves. And so, I had actual language that I could use with them that was tied into the standards that was friendly enough for a third grader to understand to show them ultimately what my goal was for them.

Ms. C mentioned how the rubric supported her in one-on-one conferences with students, stating that she would "talk about their goals and talk about how they are doing and just really making it more individualized."

Many of the teachers compared the feedback they provided with the rubric to past practice, as Ms. B did when she stated,

> I think it gave me a better opportunity to provide feedback, for sure. It also—instead of just saying good job, I can actually give legitimate feedback, and tell them what they were able to do. Or if they weren't able to, we can really, actually talk about it now. Where before, I felt like it was more like, oh, "great job, you got an A." Or "great job you did 100%." Where now I can say, I really like how, like we're doing adding and subtracting right now, for instance. I can say, "When

> you're borrowing across zeros, it seems to be where you really struggle is with the regrouping across zeros." Or before it would have been, "oh, you got that answer wrong."

Both Ms. J and Ms. I described similar experiences to Ms. B. Ms. J noted how her feedback became more specific due to this process. She asserted,

> Before I would give feedback and it would be something as simple as like "great job" or "watch your signs" where like now it's "okay this is what your step is or this is the particular struggle that you're having." So, I think I was able to just like hone-in more on that specific.

While Ms. I shared how she used the rubric to coach students, saying,

> It's like you can show them something like "this is how much you've covered so far. And if you just look at something in this sense, like a little bit more, maybe explain yourself a little bit better…you might be able to clarify better what you're trying to say and hit this part of the rubric better." So, yeah, I use it as a feedback tool for them.

Overall, teachers articulated that using the standards-based rubric supported them to provide descriptive feedback in written and oral forms.

Subtheme 5: Supports Teachers' Communication with Parents

In addition to descriptive feedback to students, a few teachers shared how they have transitioned to providing more descriptive information when reporting progress to parents. Ms. D used the rubric to describe how success would be reported to parents, stating, "I have plenty of evidence of you being able to do 4 by 1 digit, but you haven't

mastered Trailblazer. You haven't earned your 3 on your report card until you've shown all of these skills." Ms. F articulated how she describes progress for parents on report cards when she stated,

> Being able to show those standards on the report card and say, "This is what your child is currently doing." Tells me a lot more than, "Oh, he got a B or a C or a D or whatever it is on there." Now, I know like, "Oh, we're looking at the standard." He is approaching the standard, or she is a little behind a grade level, but he will get there eventually. So, we know that he hasn't mastered it possibly. But eventually he will, and we'll continue to work on it. So, it makes me appreciate that a lot more than as if we were just getting a grade for it.

Ms. G agreed and asserted, "It also helps with parent communication, too, when you are very direct, and you have concrete things to show where you came up with the goals and why you're doing what you're doing." Therefore, some participants described how the rubric supported them in providing descriptive feedback to parents and students.

Theme 2: Promotes Standards-Based Focus

All nine teachers described how using an SLO process structured by a standards-based rubric supported them in focusing their instruction and assessment on the chosen standards. Ms. E expressed how the change has shifted her approach from implementing a textbook program to focusing on teaching to standards stating,

> So, my teaching has completely been streamlined since we started using these. We used to use Everyday Math and just the spiral math every day. And now what we have is, we know when our unit will start, what standards we're teaching

> within each unit, when the formative will be given. So, everything is very focused compared to what it used to be.

Several participants shared how the rubrics supported them to stay focused in their instructional planning and the execution of lessons. Ms. F referenced how the process supported her focus by stating,

> It's easy to deviate sometimes. Some of our plans, depending how the day is going or stuff like that. And sometimes I feel like, "Oh, we haven't gotten there because we found this, or we got distracted by this and stuff." So, I do think that it definitely does help. And it's a way to kind of keep us as teachers focused on what we need to teach or the standards we need to teach for certain units.

Some participants expressed feelings of self-assurance using the rubric to teach to standards. Mr. H shared this sentiment and acknowledged his feelings of security when he stated,

> I think it gives me a sense of confidence about what I'm basing my teaching on because I know what we're heading towards. I know what I'm looking for. It helps me to be more focused on some things and more expansive on others because I know where things fit together.

The theme of a standards-based focus also emerged in participants' comments about assessment. As Ms. F described, "it helps me to be more focused when it comes to my assessments as to what standards we're meaning to assess." Ms. C exemplified this intentional focus on standards when developing assessments stating, "our assessments are standards-based, so we make sure that we are creating … creating problems for the

students that are standards-aligned and aligned to the rubric." Thus, using the rubric to structure SLO has supported these participants to focus on standards in both instruction and assessment. Four subthemes emerged from the data in this theme of standards-based focus: (a) promotes common understanding of standards, (b) promotes teachers' depth of understanding of standards, (c) promotes students' depth of understanding of mathematics, and (d) supports transition to standards-based system.

Subtheme 1: Promotes Common Understanding of Standards

Most participants described how the use of standards-based rubrics supported educators in developing their collective understanding of standards. Through their collaborative dialogue with fellow educators about the rubrics, they described how they came to some consensus about the interpretation of mastery for their respective performance levels. Ms. G argued that teachers might have had different interpretations before discussing the rubric, stating,

> Because we're talking--it was all just different information. I mean, we all came with our own, I don't want to call it agendas, but we all came with our own thoughts and ideas, and there was nothing that was necessarily concrete that we were discussing like we would be like, "Oh, we're going to talk about our fraction unit," and then we could all talk about it. But this really focuses our discussion like this standard, this 3.NF.3 is about or comparing ordering of fractions, and so we could start and have a conversation just around that standard.

Ms. C articulated how grade level PLC meetings were a venue for developing common assessments stating,

> We did a lot of PLCs. We met as a grade level team to evaluate the standards and see how we are going to assess students. There's been a lot of changes throughout the years. We've been really tweaking our formative assessments and our summatives. We don't have a set curriculum that we follow so, a lot of us teachers have worked together to develop our curriculum and how we are going to … how we are going to achieve the standards throughout our different units.

Ms. I expressed that the organization of the rubric supported her team in reaching a consensus. She stated,

> In seventh grade, we're covering rational numbers. So, what does that mean? What does that mean to everybody in the whole world? And so, this tool kind of breaks it down. This is what it means. You need to do this, and the kids have to meet this minimum criterion and the students have to know this and they've got to know these minimum criteria.

In contrast, Ms. B met less formally with a teaching partner but still discussed the standards and rubric to reach a consensus.

> I feel like I learned them in and out because I had a partner who was very—she really focused on the verbiage of standards. So, we would sit and have conversations of, well it says that they have to, they have to demonstrate it. What does demonstrate mean? And the standard, and things like that. So, for our standards that we are covering, I feel like I've understood them a lot better through this process. Because how can you say something is demonstrating

> mastery or can be gathered as evidence if it's not specific to that standard and what it actually says? So, we do a lot of standards analyses.

Overall, participating teachers developed a shared understanding of standards through these rubric based discussions.

Subtheme 2: Promotes Teachers' Depth of Understanding of Standards and Prerequisites

Both Ms. I and Ms. B alluded above to how the dialogue that supported them and their colleagues to reach consensus also deepened their understanding of the standards examined in their SLO. Two categories emerged under this subtheme of the depth of understanding: (a) promoting teachers' depth of understanding of grade-level standards, (b) promoting teachers' depth of understanding of prerequisite expectations.

Teachers noted that using an SLO structured around a standards-based rubric deepened their understanding of their targeted grade-level standards. Most agreed with Ms. B's statement above, such as Ms. C, who asserted, "I think I've gained a greater understanding of the standard itself and the different teaching methods that it takes to teach those standards."

The structure of the rubric itself supported some teachers to deepen their understanding; as Ms. G shared, "The rubric really helped guide me into discussing the different levels of teaching and learning and explaining where I was going and why." She went on to say that the rubric "made it so that it was easier to digest, and I guess I could see where kids were at and where I wanted to push them, even if pushing them beyond third grade."

The process of monitoring growth along the rubric led Ms. J to reflect on implementing the mathematics standards. She declared,

> They're definitely more involved and I think a lot of times we just did not always get to that complex aspect of it, and we were more just in the simple. So, it definitely took a lot more. but I also felt that by using the rubric, it broke it down more, too. So, it wasn't just like I didn't have to just give them, okay, let's add, subtract, multiply, and divide together. We were able to take each part and break that up and kind of just focus on one. And we really looked at it to make sure that they were ready to move on from one before we went to the other.

Similarly, Ms. E reflected that the rubric also supported her to translate this deeper understanding into her instruction, saying,

> With each standard, that you could break it apart and try to make each piece of each standard a moment for that child to be successful and how to progress through that standard instead of just throwing it out all at once and would be in one big pile at the beginning of a unit. So really being able to take it apart so that the kids can access it in pieces to be successful.

In general, teachers deepened their understanding of standards through their implementation of rubric based SLO for teacher evaluation.

Not only did teachers deepen their understanding of the targeted standards, but several participants described how the inclusion of prerequisites on the rubric guided them to deepen their understanding of the connection between prerequisite expectations

and grade-level content. Ms. B articulated how she began her SLO by examining background knowledge thus,

> So, the prerequisite standards that are on our rubric are what I would start with. So, if those kids don't have the prerequisite standards for the pretest, for instance, I just start with a blank slate. If they don't have it at that, then that would be what I would cover in intervention, or something like that. So, it informs what I'm going to be teaching.

Ms. D shared this sentiment, adding how she analyzed the descriptions in the rubric to break apart expectations into skills and track who needs to address prerequisite content.

> So, for multiplication, those same skills—and you will see for each level that we have two skills—I still break that apart so I don't just have level 2. So, I know specifically on different assessments if they're doing—If they can interpret a multiplication equation as a comparison, or they're doing the reverse where they're taking the multiplicative comparison and putting it—as they find which one it is. But so, you'll see over here is my pretest data, and you'll see some students, because of quarantine and things—other ones that are lacking some of that. But then, from that point on, I can plan instruction that's aligned to our common formative assessments.

Ms. F articulated how the rubric connects prerequisites to the targeted standards stating, “for the place value ones, it's very easy to see how they build on each other.” She went on to describe how this connection has informed her implementation, noting, “That's definitely the biggest takeaway is that I don't need to deviate too much—just try to build

on from those rubrics and go on there." Thus, participants deepened their understanding of grade-level standards and the prerequisite standards on which the targeted grade-level standards build.

Subtheme 3: Promotes Students' Depth of Understanding of Mathematics

A few shared how implementing this process has supported their students to deepen their understanding. Ms. C articulated how the process evolved from being teacher-centered to being student-centered thus,

> When we first started, we had our performance descriptors; and now we give the standards to the students so that they have more ownership, and they understand what they're learning. At first, it was more of a teacher-based rubric that we were following and were looking at how are we going to assess these students, but now the students actually know what are the standards, what are the objectives, what am I going to be learning in this unit and how am I going to show growth throughout the unit?

Knowing the standards may support some students to deepen their thinking, as Ms. G observed,

> I realized from my own experience that I want kids to know the why, what they're doing and why they're doing it, and I want them to understand conceptually what's going on. And they can learn those algorithms later on when they have the concepts down to speed things up, but they need to understand what they're doing.

Ms. E shared this sentiment, reflecting, "I believe that what we are teaching now really gives students a deeper understanding of what they're actually doing when it comes to

math. And they can explain why things are happening better than I ever could have as a little kid." Overall, teachers noted how implementing the rubric based SLO supported them in facilitating a deeper understanding of mathematics among their students.

Subtheme 4: Supports Transition to Standards-Based System

Five of the participants described some form of standards-based change within their school systems. Ms. D described her experience of change, saying, "The rubric is like the performance descriptor, it's literally the heart of everything. Everything is based off of it." She also described that she uses "standards-based tests" and "standards-based grading" practices aligned to the standards-based rubric. She noted, "My grade book is my SLO data…It's all in one. So, my grade book is the rubric itself or the standards performance descriptor itself." Ms. B agreed, saying, "My standards-based rubric is completely my SLO. So, for my SLO, I just turn in my grade book." Ms. B elaborated, saying, "our SLO is—I know a lot of teachers kind of think of it as like a whole extra thing we have to do, and it's such a pain—But for us, it's pretty woven throughout the whole year."

With her shift in focus to standards-based teaching, Ms. F observed a change in her approach. She described it thus,

> Before, when I was a para, I used to be all over the place. So now, that I'm an actual teacher and using standards-based grading and performance descriptors, it's so much easier to just follow that and build on those instead of being all over the place and trying to teach the student.

Similarly, Ms. I observed a transition to standards-based thinking. She reflected on the shift from following a textbook program to following the standards by stating,

> It's very easy to go to your book and turn to Chapter 11, okay, there you go. And did you learn, and how come you didn't learn? We went over that. So, coming out of the book, the textbook, even though the textbook is set up and it's saying you are working on this topic, you're working on this standard, the book says you are. There's not that conscious thought. The conscious thought is not there because the book says this is the standard you're working on. I've got to go and look at the standard and see what the standard is saying. Like how complicated is this going to be, where do I have to take them to go? And I can honestly say that I hadn't done that before. I would just teach the concept and not really go to the standard and try to get some more information from it.

Thus, 5 of the 9 teachers discussed observing a transition to standards-based focus in their system.

Theme 3: Fosters Evidence-Based Assessment

Eight of the participating teachers described assessment as a process of gathering evidence rather than as an event. All nine participants used observation and student work samples as evidence. Four subthemes emerged in their assessment descriptions: (a) increases teachers' knowledge of student learning, (b) foster evidence-based assessment of student growth, (c) foster evidence-based assessment to establish a baseline, and (d) foster evidence-based assessment to show mastery of the standards.

Several participating teachers concurred with Ms. C's interpretation of assessment as "looking for different pieces of evidence." She went on to describe her process thus,

> I want them to show me that they're able to perform a task, whether it is hands-on, using manipulatives or a worksheet or a game or an activity, I want them to show me that they're able to perform that standard multiple times with proficiency.

The following statement by Ms. G shows that she also gathers evidence to assess and then aligns the evidence to the rubric:

> I think, overall, it just really has helped me take anecdotal notes on my students. Their learning, my teaching. It's very specific and, I guess, helps communicate in a clear way, so that everyone's kind of on the same page, and it's kind of like proof. I could say you're developing, and I can actually show you this progression rubric and say why you're developing versus secure, versus beginning, and show you the different levels, and show you where the goal is, and it's more concrete.

In general, participants shared the interpretation of assessment as a process of gathering evidence.

Subtheme 1: Increases Teachers' Knowledge of Student Learning

All nine participants described how the process of using the standards-based rubric to monitor student growth increased their knowledge of students. Many articulated how the precision of the performance level descriptions supported them to identify students' location in the learning progression, such as Ms. D, who stated,

> I'm just so much more aware of where my students are on those standards because we're constantly--like those are the objectives. That's constantly what we're

> discussing. And it's more so like, “Okay, we have the standard broken down, especially for addition and subtraction.” Level 3 is you could do an algorithm. You can perform the standard algorithm with multi-digit whole numbers, but you're regrouping, is what—you're still making mistakes, regrouping. Whereas then level 4 is you can do it with no errors in regrouping.

Ms. B described how the specificity involved in this process differed from the previous practice of strictly using letter grades when she described,

> The way that we track data really, I feel, gives me a good idea of what they can and can't do. So, like before, for me, I'm thinking of our standards-based grading also. So, before when they would get you know, an A or B, or they can do it, or they can't. It didn't really tell me skills, specific skills. So, now I feel like when I'm doing an SLO and standards-based grading, I can hone in on specific skills of whether they can or can't do it. So, I feel like I have a much better idea of skill specific ideas.

Ms. J shared this sentiment, adding how the rubrics supported her to diagnose gaps when she stated,

> Since I've been using the rubrics it's definitely not the everybody moves on at the same time. I feel that with the rubrics, I have a better understanding of where the kids or the students truly are in the standard. And it's not necessarily that they're missing a whole standard, they may just be missing a part of it. And even in looking at the rubrics, I've sometimes found that they understand the grade-level standard concept, but they're missing some of those prerequisites. They get the

> process, but it's the computation that's maybe an issue for them. It's definitely had me realize that everybody isn't always at the same place at the same time and kind of just a better way to meet the needs of the students.

In addition to the content knowledge awareness, Mr. H shared his improved understanding of students' dispositions toward learning through this process. He noted, "What I learned about my students, some are very invested in getting to that next level meeting each goal…when they saw that there were goals to tackle, they were on board." Therefore, all nine participants experienced an increased knowledge of their students by implementing the rubric based SLO to monitor growth for teacher evaluation.

Subtheme 2: Fosters Evidence-Based Assessment of Student Growth

Although eight of the nine participants described assessment as "gathering evidence," all nine teachers interpreted student growth as movement across the rubric. They all used the rubric to track the progress of each student from their baseline level. Ms. I described the process thus,

> Our pretests determine where the majority of the kids are, what level they're going to start within the progression, and then as they are all progressed, we're all trying to get them to progress to grade-level. And once they get to that part and we finish off whatever is the final CFA and then give the summative assessment to see if they have improved, mastered the content level.

Ms. E used a similar process, noting how the rubric defined the sequence for her unit when she stated, "I use my rubrics to gauge how I start my unit. So, how I want them to progress through the standards. So, I'll start with the lower levels and then move up

through the standards in that order instead of jumping around." She further described how the students were able to see their growth. "I started using the rubrics with my students. So, with a student-friendly rubric, so they could see what the goal was for the unit where they start. So, we'd mark where they start and then how they progress through it."

Ms. F also provided a rubric to her students, noting how students used it for personalized goal setting. She stated,

> I actually use the rubric with my students. I like to show them like, "Hey, so we took the pretest. This is where you landed on the pretest. Let's look through all these performance descriptors and come up with a goal. Like what do you want your goal to be?" I feel like it helps make it more personal and it makes some feel like they're in power and they're the ones choosing where they want to be. So, I'm able to use that. And we have different data points, different assessments throughout the unit where we're able to meet up like a quick five-minute meet up with them and be like, "Oh, you just took this assessment. This is what you've showed me you've mastered and what you can do." And I think it definitely helps motivate them and helps them feel more in control of their learning.

Some participants shared how the rubric supported them to promote growth for all students, including those in need of more challenge or enrichment. Ms. D noted that the rubric included beyond level expectations for such students stating,

> You'll see that we do have the 3rd grade standards for level 2, but it's not just one standard. It mixes the numbers in base 10 with the operations in algebraic thinking to fully acquire 4th grade and we don't report level 5, but we have that

> included there in case, for SLO, we have a student come in on the pretest at level 3. We could still show two levels of growth for that student by challenging them with this, the next grade level standard.

All nine participants shared observations of how this process represented a transition to evidence-based growth. Mr. H described the change in assessment, saying, “I think [assessment] is more graduated--before I think it was more binary. They get it, they don't get it versus what is it they get and how does that help them get the next thing?” Ms. B agreed, adding her perspective on summative assessment and growth when she stated,

> Before, I would always do the pretest and it was like the percentage they got on the pretest, and then the summative…That's drastically different than how I do it, actually. And so, I get their starting level from their pretest, and then gather along the way. So, that summative while, you know, it is the summative, it really to me is just a piece of evidence. So, I can see the growth throughout the unit, rather than from this test to this test.

Ms. D reflected on the validity of assessments for measuring standards-based growth when she stated,

> Using district common assessments like this or classroom-based assessments are much truer and more valid set of data because it's genuinely what students are being taught and we’re able to align our instruction to the assessments. I mean, it's universal backwards design that is good pedagogy, like we know. Whereas these outside assessments aren't—Those are not aligned with our instruction, so they're

> not always valid. They might be normed and being able to give us that type of data, but it is important to consider the type of data that they're using and whether or not it's valid or not.

Mrs. I agreed with Ms. D, adding, “So, using a tool like this is more—I think it represents student growth better.” Mrs. I further described the shift in her thinking that resulted from using the standards-based rubrics, saying,

> I do have to tell you, I had not the highest opinion about rubrics…I'm thinking if you show them a way to be mediocre, they'll achieve that. So, that's what I always thought about the rubrics. But in this case, the more you discuss—okay, so you just have to be able to look at it from a different perspective, like, look what you can do, you can do this now. And it's not so much if you do this, this is your grade. It's more like, look what you're capable of doing.

Overall, teachers perceived the process of monitoring growth on the rubric to view assessment as an ongoing process of collecting evidence of that growth.

Subtheme 3: Fosters Evidence-Based Assessment to Establish Baselines

All nine participants used pre-assessment at the beginning of the unit to establish a baseline for each student. Ms. E described how the pre-assessment addressed prerequisite skills stating,

> The rubric we use at the beginning with a pretest to see where they are entering the unit to see if they have the prerequisite skills that they need from the very beginning, and maybe possibly they already can meet of all the standards that we want them to.

Ms. F agreed, noting how easily she aligned the data to the performance level descriptions when she said, “I feel like the rubric actually it makes it really easy for us to be able to put them in a level because our rubrics have a description of the standard and sort of an example of they should be able to do this.” Ms. J had a similar experience, noting how the rubric supported her in addressing gaps.

> To gather the baseline data, we looked at the prerequisite standards because the first two parts of the rubric that I used were the prerequisite standards. So, we used that to determine if the students were prepared to move on to the grade-level standard, and then, once we were able to see where they were on there, we did our best to close those gaps and then moved on to the grade level.

Mr. H also noted that the baseline data provided guidance when students already had grade-level knowledge. He stated,

> Using some of the assessments, the readiness, and the pretest, pre-assessment. Just to see where kids are at in terms of what skills they need to come into the unit. Then what is going to be asked to them ultimately so that they don't need to repeat things they already know.

Thus, teachers used the rubric to guide gathering evidence to establish a baseline for each student to begin implementing their SLO.

Subtheme 4: Fosters Evidence-Based Assessment of Mastery

Participants shared how the rubric supported them to focus assessment on the targeted standards. Ms. F noted, “I do think that my assessments are a lot more clear and more direct to the point. Because I have, right in front of me, what standards I need to

assess, and I don't need to include all those other standards that are not being assessed." In addition, some shared how they continued to gather evidence of mastery over time, as Ms. B stated, "I usually will have to gather more evidence from them if they haven't demonstrated mastery within the unit."

Ms. J described how designing the summative assessment provided clarity for instructional planning when she said,

> In beginning my SLO, I looked at the standards-based rubric to see what was necessary for the students to be at the highest level and then use that rubric to develop the assessment, because I knew that that was what the standard required of them. And then went from there and knowing the assessment was able to go back and then develop the instructional activities necessary as we move through with small formative assessments throughout the process to see them moving through each part of the standard.

Ms. D articulated how she identified the importance of assessing the full rigor of the standard when she expressed, "when we're hitting that standard, we're hitting it to its full integrity." She went on to provide an example of addressing the full rigor for word problems:

> When I'm asking word problems in a unit, yes, I'm in my multiplication unit, but I should still be including division and all of the other operations in those because that's the core of that standard—being able to choose which operation you're using in which context. You're not actually conceptually understanding it if you're just doing multiplication because you're in your multiplication unit.

Overall, participants described how the rubric guided them to assess mastery of the targeted standards by capturing the full rigor of each chosen standard. As Ms. G stated, “I can now look at the different standards, what the expectation is, but then I can also look at the level of depth that they are understanding the end.”

Theme 4: Supports Student-Centered Instruction

All nine participants articulated how they reflected on using a standards-based rubric to monitor student growth to guide their instructional planning. Most shared reflections similar to Ms. D, who contemplated, “So, if students struggled with this type of question, how might I change that in my instruction for the future?” Five subthemes emerged from their comments on this theme: (a) guides grouping of students, (b) guides targeted intervention, (c) guides differentiation of instruction, (d) guides reflection on instructional tools and strategies, and (e) fosters mathematical language.

Subtheme 1: Guides Grouping of Students

Eight of the nine participants discussed how reflecting on data with the rubric guided them in grouping their students. As Ms. G described,

> Initially, what I do is we would look at kind of low, medium, high, and we would do small group instruction. But they're not stuck there. So, for instance, if they do poorly on the pretest and they're put in the low group, but then they start to catch on, they're going to move. If they maybe had a bad day and had a bad test, and all of a sudden, they're in this group, and they're like, "Oh, wow," they know a lot more than they showed on that test, they would get moved. Some kids move faster and slower. So, the groups change. They're very fluid throughout the process.

Mr. H. echoed this approach, adding that “Each [group] enters the teaching sequence at a location that best supports their learning.”

Ms. B reflected on data in determining whether small group or whole group instruction was appropriate, stating, “it definitely influences which groups I'm going to be pulling. Which kids I'm going to have to work a little extra with, whether I have to pull small groups or teach a whole class.” While Mr. H reflected on his pacing of instruction for his groups stating, “What lessons make sense for the groups that I'm working with, the two groups or three groups at different times. Where do they come into the unit and what speed can they work? Can they take on new information?” Many teachers shared how they used data in planning intervention groups. Ms. D described, “I know exactly how many students in my class I still need to pull in small groups and do more practice with regrouping, even though that unit has ended.” Overall, participants found using the rubric helpful in informing their small group planning.

Subtheme 2: Guides Targeted Intervention

Teachers shared that once they identified which students had similar needs to form groups, they also used that same information to plan instruction for addressing those identified needs. Participants shared Ms. E’s perspective, “It helps me know where they are successful. And where they need that extra reteaching or extra time to be able to be successful with each standard.” Ms. C noted that she could deliver targeted interventions in mini lessons to small groups or individuals, saying,

> I think the rubric really influences my interventions. If I'm seeing that students are not showing growth in that they're not understanding different concepts, that's

> when I take a minute and I provide them with those mini lessons, I'm meeting with them small group, maybe even one-on-one to make sure that they're on the right track.

Some teachers articulated how intervention planning has changed since using the rubrics. Ms. I described how her thought process changed, saying,

> I look a little bit closer at what went wrong when they go wrong…Because I used to generalize a lot more and now it's more specific. Like looking specifically, what could the problem be? Why did these five kids get this question wrong? What is missing that they weren't able to even pull some old background information and then you scaffold it to make them understand?…I used to look at it as just say, oh, you got it wrong…Now it's more specific looking at the problem.

Participants described how the rubrics supported specialist teachers in providing targeted interventions. Ms. F described how she used the rubric to plan instruction aligned to each student's individualized education plan goals, stating,

> I always look at the rubric to try to find the standards that my students are working on. Because, as you know, in special ed. we have to select standards to line them up with their goal. So, I always want to make sure that the goal that I'm working on is part of a standard on the rubric. That's always the first step for me. And then, once I go from there, I see of like Alyssa's — in our performance descriptors I go and see, "Oh, is this a level one? Is this a pre-skill or whatever

> level it's on? And then, built their lesson plan or their plan for intervention from there.

Ms. G described how collaborating with special education and English language teachers to analyze the performance level data provided targeted intervention thus,

> Usually, we were trying to get kids a double dose of [small group] math instruction, and we tried to be—We would look at their learning styles, but most of the time when they were low, visual was a huge part, tactile is a huge part. And then using that with a double dose to hopefully kind of close the gap a little bit so they could move along with their peers.

Overall, teachers shared that they believed rubrics supported them to plan and implement targeted interventions to address the needs of their students.

Subtheme 3: Guides Differentiation of Instruction

Three of the nine participants described how the rubric supported them in differentiating instruction to ensure both low and high students showed growth. Mr. H examined his data and asked himself, "For the kids that are already adapted to the standards coming into it, what are some directions for what will challenge them going forward to both ends of it? Ms. G agreed that the rubric supported differentiating for the variety of students in her classroom. She stated,

> I think that the rubric really helps see kids and their different levels of thinking. And so, you have very visual kids, you have kids that are like calculators. And so, with this rubric with the different progressions, I was able to kind of look at where they're at and what they're thinking and how I could push them, and how I could

> differentiate to meet their needs and it was written in a way that helped me identify different kids and their needs for differentiation.

Ms. J commented that she looked forward to using the rubric more for "developing more differentiated activities, like really looking at ones that worked and didn't work and trying to develop it so more students could move further along in that progression." Thus, some participants found the rubric helpful in supporting differentiation for support and enrichment purposes.

Subtheme 4: Guides Reflection on Instructional Tools and Strategies

All nine participants described their experiences using the rubric for reflecting on instructional tools and strategies to meet the needs of their students. Ms. C explained, "I think it helps me reflect a lot on my teaching and the tools that I'm using and the assessments that we use in the district and how effective they are, showing student growth and being more standards-based." Ms. E agreed, describing examples of tools and strategies when she stated,

> It can help me adjust my teaching by seeing how they learn best. Do they learn best with the manipulatives, with games, with the songs, or different strategies we might learn or use in the classroom? And just yeah, I don't know. I use how they progress. I can see if they're successful based on certain strategies I use in the classroom to know that's going to work better for them in another unit or just moving them through that standard, I guess.

Ms. D described how using the rubric "influences me as an educator because I use my number sense. And I'm always doing 'think-alouds' of my number sense to try to help

students acquire that as well.” Ms. F reflected on how she has adapted her SLO over time, stating,

> The first time I did my SLO, I didn't include as many visuals or hands-on stuff. When then I realized like, "Oh, I need to modify that because we really needed that." So, I think every time — and it depends on the group of students. Every group of students is different. So, I'm constantly modifying my SLOs for the units as needed.

Overall, participating teachers articulated how using the rubric to monitor student growth supported them to continuously reflect on instructional tools and strategies to suit the needs of their students.

Subtheme 5: Fosters Mathematical Language

Four of the participating teachers described how using the rubric to structure their SLO supported them in fostering mathematical language with their students. Ms. G appreciated that the rubric provided “actual language that I could use with them that was tied into the standards and was friendly enough for a third-grader to understand to show them, ultimately, what my goal was for them.” Ms. E mentioned how providing the rubric to students led them to engage in mathematical discourse. She stated,

> [Students were] having those conversations with each other, “Oh, you know, I was so close, but I forgot to do this on the number line.” Or whatever it was or if I'm talking about telling time, “Oh, I forgot to–I messed up the AM and PM,” and I have to have that to be at a level 3. So, it does give them also that vocabulary.

Ms. F also gave the rubric to students but shared how her bilingual and special education students sometimes needed modified rubrics. She noted,

> Sometimes I do have to modify the rubric and use simpler language. But I try to use the academic language because the point is we want them to be able to understand what the standards mean and what they are. So, I try to keep some of the academic language but just simplify it down more to a level where they can understand it.

Ms. D agreed with Ms. F, stating, "I have a lot of language learners in my building. So, everything I'm doing, very gradually, very color-coded, and step-by-step with making sure that I'm holding them accountable to the mathematical language of the standards." Thus, some participants found the rubric supportive of students acquiring and using mathematical language.

Theme 5: Encourages Students' Reflective Practice

All nine participating teachers described instances of using the rubrics to monitor student growth led students to engage in reflective practice. Two subthemes emerged in this theme: (a) students' reflection leading to their use of feedback and (b) students self-monitoring of growth. Ms. C articulated both of these ideas when she stated,

> It builds a different classroom culture, I think, where students really are focused on the objectives and how they are learning and what they're doing and they're having those conversations with each other. They're able to have conversations with me and it really creates that mindset of, "I know what we are doing in the classroom, and I know how I'm going to meet my goals."

Subtheme 1: Encourages Students to use Feedback

As noted previously, all nine teachers revealed that the rubric supported them to provide descriptive feedback. Eight of the nine teachers provided examples of students using this descriptive feedback to show growth. Most shared similar experiences to Ms. G, who expressed, “I would say the majority really take it and listen to what I have to say. And I think when the goals are clear to them, they know what they're working for.” Ms. F noted that students look forward to feedback when she stated, “I've had students where we'll have an assessment and they'll be like, ‘Oh, do we get to see what level we're on now?’ So, they always kind of look forward to seeing like that they move up a level.”

Ms. B described how students discussed her feedback with each other when she noted, “I've heard my kids talking amongst themselves and saying, like, ‘I'll be a level three if I could just get to regroup. And if I could just learn to do this. I just got to prove this to her.’” Ms. I described how some students questioned her about the feedback she provided, saying, “they'll take that feedback, and they'll go, okay, what do you mean by that? And then they'll actually pull up some part of their work and they'll go ‘Is this what you mean? Is this what you're talking about?’” Thus, most teachers perceived the descriptive feedback they provided to be used by many of their students and helpful to student learning and engagement.

Subtheme 2: Encourages Student Self-Monitoring of Growth

Eight of the nine teachers shared instances of students self-monitoring their growth while implementing their SLO. Teachers agreed with Ms. C, who noted, “We give the standards to the students so that they have more ownership, and they understand

what they are learning." Ms. E described how giving second-graders rubrics empowered them to curate their evidence when she stated,

> When I started giving them the student-friendly versions, they would have their own—we called it their PD file, but their performance descriptor file—and every few days we would take it out. And if we did one of the standards at level 2, if they hadn't met that, and they just were able to complete it in the classroom, they were able to put a little smiley mark under that level themselves. So, they were responsible and sometimes they'd look, and they'd say, "Oh, you know, I'm so close to level three but I didn't do this. I need to do that next time." So, it empowered them to understand what they were learning.

Ms. J described a similar experience with her middle school students when she gave them the rubrics. She noted,

> I would show them [the rubric] that would be attached to their pre-assessment, where they could go back and look at the pre-assessment just to have an idea, and then attach to the post-assessment, because that was how they were evaluated. But then in the same sense, we also had "I can" charts that were created that aligned with the performance rubric. So, the students had access to those as we move through the unit where it was kind of like an "I can" checklist, that they could then mark that off as well as they move through the learning targets.

Ms. C further shared her observations of the impact on students who engage in this process, stating,

> I think using this student-centered approach to where they can actually be a part of their learning and understanding the learning and the objectives and the standards, it really…it just really opens up their minds to what they're learning, and they take part in what they are learning, and they are able to set goals for themselves.

Overall, those teachers who showed students their progress on the rubric found the process empowering students to accept ownership of their learning and encourage them to be engaged in self-monitoring.

Theme 6: Cultivates a Positive Teacher Evaluation Experience

Participants reported positive teacher evaluation experiences when they used standards-based rubrics to monitor student growth in an SLO process. Two subthemes emerged in this theme: (a) becoming a better teacher, and (b) improving the evaluation experience.

Subtheme 1: Becoming a Better Teacher

Six participants observed that using standards-based rubrics to monitor student growth for teacher evaluation supported teachers to improve their practice. Ms. G articulated how the process supported her to reflect and improve by saying,

> I would look at the standard, I would see where kids are doing well, and not so well, and then I would change things based on what I saw in my own teaching that maybe, to me, I was looking at myself going, "Okay. Well, I need to do a better job at teaching this because a lot of the kids did not do well on that, and because of that, it shows that it was probably my instruction, versus if 1 kid or 2

> kids didn't do well on something. So, then, if that was the case, then I would go back and kind of peel the standard apart more and figure out…talk to colleagues and whatnot and figure out how I could improve my teaching.

Ms. I and Mr. H also shared reflections on their growth. Ms. I stated, “It makes me feel like, wow, what was I doing all those years!” and Mr. H shared, “It just helps me feel like a more flexible teacher.” Ms. B also found the process to support her improvement, so much so that she recommended, “I think it's definitely worth it to use them. I think it's beneficial to use them through all your math units, not just for an SLO.” Ms. C agreed, noting what evaluators might see when observing teachers who experience this process. She stated,

> I think that the evaluator, or the principal, they would be able to see that classroom environment where students are involved in their learning, and they would be able to see that teachers are more reflective, and they're seeing how can I make changes and how can I better assess my students.

Both Ms. I and Ms. E encouraged others to try this process. Ms. E suggested,

> My advice would be that, even though at the beginning it may seem like just another hoop to jump through that it can benefit you as a teacher because you really are taking that data to understand your students and how they're progressing through each standard. And if they're not what you need to do to help make them successful. So, I would say, to trust the process and to use that to become a better teacher.

Therefore, several of the participants found using standards-based rubrics to monitor student growth for teacher evaluation to be beneficial for becoming a better teacher, which they perceived as a positive change.

Subtheme 2: Improving the Evaluation Experience

Eight participants expressed perceptions that using standards-based rubrics to monitor student growth in an SLO process for teacher evaluation was a positive evaluation experience. Ms. G described her experience, saying,

> I think that for me, it's very clear, it's not foggy. I'm not questioning anything. I know and I can show you, I can tell you what I know and how I know it and I can tell you the why, and having some clear, a lot more-- How do I describe it? It's clear cut.

Both Ms. D and Ms. J expressed how much they enjoyed their evaluation experience. Ms. J stated, “I definitely enjoyed it and I'm looking forward to continuing to implement it and further the implementation of it.” Ms. D shared both her joy and satisfaction with student growth, saying,

> I really love the performance descriptors. I couldn't imagine because this has been my career. I couldn't imagine using a different system… I do enjoy it. And I think it's effective. And I've definitely seen growth in my students using this… So doing it well in the classroom leads to that kind of growth that can be reflected on those standardized tests.

Ms. I agreed with Ms. B that the process is worthwhile and encouraged others to try this approach when she advised,

> I like the organization part of it… Next year, I'm much more confident in what I would do, and I know where everything is at. I know the things I like. I know the things I would change. So, two years for sure to do this… I would just suggest that people be open enough to try it.

Overall, several teachers expressed their perceptions of an improved teacher evaluation experience using a standards-based rubric to monitor student growth and encouraged other school systems to consider implementing this process for teacher evaluation.

Summary

The chapter included descriptions of setting, demographics, data collection process, data analysis process, and evidence of trustworthiness. The chapter also included the results from the field test and the conducted study to address the research questions:

RQ1: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support reflective practice?

RQ2: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support PCK?

The themes that emerged from the coding process were: (a) fosters collaborative dialogue and descriptive feedback, (b) promotes standards-based focus, (c) supports evidence-based assessment, (d) supports student-centered instruction, (e) encourages students' reflective practice, and (f) cultivates a positive teacher evaluation experience. The next, and final, chapter includes a discussion of conclusions and recommendations of this study.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The purpose of this basic qualitative inquiry was to explore teacher perceptions of their experience using standards-based rubrics to monitor student growth in an SLO process for teacher evaluation in one mid-western state. Using the dual lens of reflective practice and PCK, I examined teacher perceptions of how the use of standards-based rubrics influence their reflective practice, their understanding of mathematics standards, their assessment tools and practices, and their instructional tools and practices.

Based on the data analysis, I identified themes that answer the following research questions:

RQ1: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support reflective practice?

RQ2: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support PCK?

The following themes were identified: (a) fosters collaborative dialogue and descriptive feedback, (b) promotes standards-based focus, (c) supports evidence-based assessment, (d) supports student-centered instruction, (e) encourages students' reflective practice, and (f) cultivates a positive teacher evaluation experience (See Table 12).

This study was conducted to address the lack of knowledge regarding how structuring a goal-based approach for monitoring student growth, such as an SLO process, could support teachers to reflect on their knowledge of standards, assessment, and instruction in the teacher evaluation context. The findings indicate that teachers who

used a standards-based rubric to structure the SLO process engage in reflective practice throughout the experience. The findings also indicate that teachers perceive an increase in their knowledge of standards and their knowledge of students' levels of understanding of targeted expectations. Additionally, teachers reported their reflection on assessment led them to also reflect on instructional tools and strategies.

Table 13

Research Questions and Sub-questions with Resulting Themes

Research questions	Themes
RQ1: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support reflective practice?	Fosters collaborative dialogue and descriptive feedback. Promotes standards-based focus. Supports evidence-based assessment. Supports student-centered instruction. Encourages students' reflective practice. Cultivates a positive teacher evaluation experience.
RQ2: How do teachers perceive the experience of implementing SLO structured by standards-based rubrics to support pedagogical content knowledge?	
SQ1: In what ways do teachers reflect on and adapt their mathematical content knowledge as they implement SLO structured by standards-based rubrics?	Promotes standards-based focus.
RQ2: SQ2: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLO structured by standards-based rubrics?	Fosters collaborative dialogue and descriptive feedback. Supports evidence-based assessment.
RQ2: SQ2: In what ways do teachers reflect on and adapt their assessment tools and practices as they implement SLO structured by standards-based rubrics?	Supports student-centered instruction.

Interpretation of the Findings

This study was focused on teachers engaging in reflective practice to influence their PCK in mathematics when monitoring student growth as an element of teacher evaluation. I interpreted these findings given the empirical literature and the dual lens conceptual framework of reflective practice and PCK for this study. Since legislation passed requiring teacher evaluation systems to include monitoring of student growth, educators have been challenged to implement growth-monitoring systems that support teachers' continuous improvement (Milanowski et al., 2016). Findings in this study indicate that participating teachers experienced a structured implementation of a goal-based approach for monitoring student growth. All participants shared how the process supports their reflective practice and PCK.

Interpretation of Findings and Empirical Literature

Theme 1: Fosters Collaborative Dialogue and Descriptive Feedback

The first key finding of this study was that teachers who used standards-based rubrics to monitor student growth for teacher evaluation engage in collaborative dialogue with evaluators, peers, and specialist teachers and provide descriptive feedback to students and parents. This finding was consistent with the recommendation of Roussin and Zimmerman (2014), who advocated for reflective conversation to promote a collaborative model. This theme also added to the findings of Darling Hammond et al. (2012), who encouraged school districts to consider alternatives to VAMs that would foster collaboration among educators. The result also confirmed the finding of Plecki et al. (2016) that SLO implementation supports collaboration among educators and the

finding of Paufler et al. (2020) that teachers value reflective conversations between teachers and evaluators. However, this theme differed from Paufler et al.'s (2020) finding that "teachers expressed disparate views regarding the impact of [the teacher evaluation system] on their professional practice and to a lesser extent, on student achievement" (p. 6). Although more than half of the teachers in Paufler et al. (2020) indicated that the process encouraged their reflective practice, almost half of the teachers indicated no real impact or a generally negative impact on their professional practice.

This theme confirms Kingston et al.'s (2015) finding that learning progressions support communication with parents and Van den Heuvel-Panhuizen et al.'s (2021) work in which they noted that mathematics teachers found oral and written descriptive feedback helpful for students' learning. The effectiveness of descriptive feedback is interesting in relation to Wisniewski et al.'s (2020) meta-analysis on the effect of feedback on student learning. Wisniewski et al. argued high-information feedback is effective when it not only identifies errors but supports students to understand causes for the errors and how to avoid future mistakes. Teachers in this study described how the rubric supports them in providing such high-information feedback to students and parents.

Theme 2: Promotes Standards-Based Focus

The second key finding that emerged from the data was that implementing standards-based rubrics to monitor student growth for teacher evaluation promotes a standards-based focus. Findings in this study support the positions of Timar and Carter (2017) and Urick, et al. (2018), who found that standards implementation requires both

administrators and teachers to study the standards to deepen their understanding of the expectations. Pak et al. (2020) supported these claims, recommending that administrators provide adaptive leadership rather than technical leadership in standards-implementation. Pak et al. (2020) stated, “Because adaptive processes necessitate ongoing learning and reflection, educational leaders should embed multiple curriculum-focused learning opportunities throughout the implementation process” (p. 12). Teachers in this study experienced reflective rubric-based conversations with their colleagues and evaluators to provide such standards-based learning opportunities.

Participants in this study also emphasized how the rubric supports them in focusing their assessment and instruction on the standards. This finding is consistent with Darling-Hammond et al. (2012), who found an increased focus on the depth of thinking required to meet standards. This finding confirms Furtak et al. (2018), who argued that learning progressions support focus for teachers to align curriculum, instruction, and assessment. Nonetheless, this finding differs from the findings of Slotnik et al. (2014), who found that teachers implementing SLO (which were not rubric based) struggled to connect teacher evaluation and standards implementation. Teachers in this study reported that the rubric provides a structure to align curriculum, instruction, and assessment to the standards, which is consistent with what other researchers have found with implementation of learning progressions (Black et al., 2011; Briggs et al., 2015; Hess, 2012; Fonger, et al., 2018). Thus, this finding added to the literature regarding the use of SLO and learning progressions for supporting standards implementation.

Theme 3: Supports Evidence-Based Assessment

The third key finding of this study was that using standards-based rubrics to monitor student growth for teacher evaluation supports teachers in implementing evidence-based assessment practices. This finding is consistent with the findings of Darling-Hammond (2016), who noted that teachers who analyze student work along standards improved their abilities to evaluate the effectiveness of their instruction and adjust practices to address their students' needs. In contrast, Garet et al. (2017) found no impact on teachers' interest in improving practices when student growth for teacher evaluation was measured using standardized tests.

Teachers in this study emphasized that using the rubrics to monitor their students' growth provides them actionable, standards-based data, thereby increasing their knowledge of student learning. This was consistent with the experience of teachers in Slotnik et al. (2013), who expressed the baseline data collection was enlightening, and the recommendation of Briggs (2013), who argued teacher-developed assessment could lead to the use of assessment results for instruction. Actions of teachers in this study were consistent with the recommendation of Leo and Coggshall (2013) to gather evidence of student learning throughout lessons to identify to what degree students met learning targets. This finding also reinforced Lin et al.'s (2020) recommendation to set data-based targets to be achievable and realistic for students. This theme supports Herman et al.'s (2011) argument that assessments for teacher evaluation that are standards-based would be likely to support instruction. It also confirms Plecki et al.'s (2016) finding that SLOs support teachers to consider alternate forms of assessment, and the finding of

McCullough et al. (2015) that SLO could build teachers' assessment capacity and data-driven instruction.

Theme 4: Supports Student-Centered Instruction

The fourth key finding of this study was that implementing SLO with standard-based rubrics to monitor student growth for teacher evaluation can support student-centered instruction. This finding is consistent with Derrington (2016), who described the positive impacts of teacher evaluation when teachers use data to plan instruction. The finding supports Prizovskaya's (2018) recommendation that teachers receive guidance using assessment for instructional decisions and Kearns et al. (2015), McCullough et al. (2015), Plecki et al. (2016), and Slotnik et al. (2015) who asserted that SLO can support instructional improvement.

This finding is also consistent with Zenouzagh (2019), who found that teachers who engaged in collaborative discussions of student performance led to "teacher change from mere delivery of teaching to learners to a more learner-focused teaching" (p. 354). It also confirms the finding of Farrell and Vos (2018) that teachers who engage in regular reflective practice change their instruction over time. However, the finding differs from that of Veugen et al. (2021) who found that teachers in their study had difficulty tailoring instruction to their students' needs, even though they had been trained in the formative assessment cycle. The teachers in Veugen et al. (2021) did not reflect on standards-based rubrics for monitoring student growth. In general, literature supports using SLO for promoting student-centered instruction. This finding adds to the literature by providing a

structure for enacting formative assessment practices within the SLO process for teacher evaluation.

Theme 5: Encourages Students' Reflective Practice

The fifth key finding of this study was that teachers perceive using standards-based rubrics to monitor student growth in an SLO process for teacher evaluation can encourage their students to engage in reflective practice. This finding adds to Rodgers (2018), who argued that providing descriptive feedback between students and teachers promotes students' agency. As Rosen and Parise (2017) recommended that teacher evaluation systems be used to identify needs for professional learning as a school improvement strategy and Bergin (2015) argued that the purpose of teacher evaluation is to increase student learning, this finding supported both of their recommendations. This theme confirms Lang, et al.'s (2014) finding that teachers engaging in collaborative dialogue around formative assessment data in mathematics benefits students. This finding also confirms Dunne's (2011) finding that using learning progressions supports students' self-reflection, thereby promoting student ownership of learning.

Theme 6: Cultivates a Positive Teacher Evaluation Experience

The final key finding of this study was that using standards-based rubrics in an SLO process to monitor student growth for teacher evaluation can cultivate a positive teacher evaluation experience. This finding supports Papay's (2012) argument that teacher evaluation can serve formative purposes. It also confirms the findings of Mette et al. (2015) and Raudenbush (2015) who acknowledge that teachers who experienced collaboration with their evaluators had positive perceptions of teacher evaluation. It

confirms Golberg's (2018) finding that teachers positively perceived experiences using standards-based performance indicators and rubrics in the teacher evaluation process as well as the findings of Bradley-Levine et al. (2017) and Tripamer et al. (2014), who found educators to have positive perceptions of teacher evaluation when it was associated with professional learning. Tripamer et al. (2014) also noted that using multiple assessment pieces was an indicator of teachers' positive perceptions of the evaluation experience. In addition, this theme adds to the findings of Smith and Holloway (2020) that teachers in districts with a focus on standardized tests experienced a decrease in satisfaction, while teachers in districts that do not focus on standardized tests can experience an increase in satisfaction.

All participants in this study engaged in an evaluation process that served both summative and formative purposes with the emphasis on formative evaluation. Therefore, the findings confirms Ford and Hewitt's (2020) finding that "teachers find meaning and satisfaction in evaluation processes that are more open to teacher input" (p. 21). Teachers in this study expressed satisfaction with their evaluation experience, citing examples of how it supported their professional growth. Ford and Hewitt (2020) added, "feedback which points the way to better teaching and learning on the part of students will, in the long term, sustain teachers' intrinsic motivation for the work" (p. 22). Some participants in this study shared how they were looking forward to using this process again and advise other districts to use standards-based rubrics due, in part, to the success they observed in their students' performance.

In contrast, many researchers found that teachers reported negative experiences in teacher evaluation systems using statistical models for monitoring student growth (Ford et al., 2017; Ford et al., 2018; Jiang et al., 2015; Lavigne, 2014). Hewitt (2015) noted that those teachers using value-added models had negative perceptions of their experiences and were opposed to their use in teacher evaluation. This theme also differs from Ford et al. (2017) who found that teachers using SLO were stressed to create a system for monitoring growth in the CCSS. Thus, this finding adds to the literature regarding teacher perceptions of evaluation experiences.

Interpretation of Findings and the Dual Lens Conceptual Framework

As noted in Figure 1, this study applied the dual lens of reflective practice and PCK to examine how teachers perceived the standards-based rubrics to influence their reflective practice and implementation of standards, assessment practices, and instructional practices. Participants articulated how reflecting with evaluators, colleagues, and specialist teachers supports them to interpret standards, design and implement assessments, and plan instruction. The rubric serves as a tool for discussing standards and how students can show progress toward meeting them. Teachers reported reflectively discussing the tool with colleagues when designing assessments, examining student work, and planning for team-teaching. They also reported that, in discussing the final results of their SLO with their evaluators, they reflected on how they used the tool to modify their instruction and might better serve their students in the future. Thus, teachers described the rubric as a catalyst for reflective dialogue.

Teachers shared that the rubric supported their engagement in both reflection-in-action and reflection-on-action. Participants described how the rubric guided them to reflect on the targeted standards when analyzing assessment data, planning instruction, and implementing instruction. They articulated how this focus helped them to reflect on their knowledge of mathematics concepts as well as their knowledge of assessment and instructional strategies for mathematics. In addition, teachers shared that the process of using the rubric to monitor growth supports them to better understand their students' knowledge, skills, and understandings. They described reflecting on which students were demonstrating understanding and ready to move on to the next steps in the learning progression and which needed additional guidance on the chosen targets. Participants shared that this knowledge of students supports planning of appropriate instruction for each student to show growth along the learning progression. Consequently, teachers perceived the process of monitoring growth with the rubrics to increase their PCK and enhance their knowledge of students.

Limitations of the Study

This study had a number of limitations, all of which are typical in qualitative research. One limitation of this study was due to the method of purposeful sampling. Participants were volunteers from one midwestern state who were invited because they fit the criteria of having experience using standards-based rubrics to monitor student growth for teacher evaluation. Because the participants were volunteers, there was no way to control for self-selection bias. Results may be skewed toward the positive because teachers who had positive experiences were more willing to share their observations.

Another limitation of the study was the content area and grade band focus. Teachers only shared their experiences monitoring student growth in mathematics. The information may or may not be applicable to applying standards-based rubrics in other content areas. In addition, the participant pool was limited to elementary and middle school teachers of mathematics and, therefore, results may not be generalizable to preschool, high school, or higher education settings.

An additional limitation of this study was due to the fact that it was conducted in the midst of the COVID-19 pandemic. Recruitment of participants was a challenge while in the midst of the great change educators were managing in their transition to remote or partially remote teaching. Several school districts suspended teacher evaluation for the 2020-2021 school year, which limited the number of viable participants that may have volunteered if conditions were different.

Another limitation to note was the study design. This study was not a case study and, therefore, did not provide supplemental documentation to examine, such as student work samples, data tracking documentation, or other SLO documentation. Because the study was a basic qualitative design and not a mixed method approach, student data were not analyzed that might have provided additional information about the effectiveness of the use of standards-based rubrics for supporting student growth. Because state testing was suspended in 2020 due to the COVID-19 pandemic, this data would not have been available to analyze. As such, the data were limited to teacher perceptions of their students' growth collected via questionnaire and interview questions.

A final limitation was my own proclivity toward rubric-based assessment and growth monitoring. As a trainer in standards implementation, teacher evaluation, and assessment literacy my preference for monitoring student growth with standards-based rubrics that provide qualitative descriptors may have influenced my data collection and analysis for this topic. Although I used the techniques of following an interview protocol, restating participant statements for clarification, and interviewing the investigator to support bracketing to minimize bias, my personal beliefs could have influenced my interpretation of data.

Recommendations

There are several recommendations for further research that have emerged from this study. As noted in the limitations section of this chapter, this study was focused solely on teacher perceptions of experiences monitoring student growth in mathematics. Therefore, one recommendation would be to repeat this study for other content areas to determine if rubrics also promote teachers to engage in reflective practice supporting their PCK in such areas as ELA, science, social studies, fine arts, or physical education. Another recommendation would be to replicate the study with pre-school, higher education, or high school teachers of mathematics to determine whether similar results would be found in other settings.

Another recommendation for future research would be to expand the focus from teacher perspectives of the evaluation experience to the evaluators' perspectives of the evaluation experience. Although teachers in this study expressed positive perceptions of their interactions with evaluators, additional research could ascertain whether evaluators

perceive the experience to promote improvement in teachers' instructional effectiveness. Research could include analysis of teacher ratings or compare evaluators' experiences with SLO structured by standards-based rubrics to experiences with SLO that did not use standards-based rubrics for monitoring growth.

Although teachers expressed positive perceptions of their experience, the research questions for this study were not targeting teacher satisfaction with the evaluation experience. Therefore, additional research could expand on the theme of cultivating a positive evaluation experience by exploring research questions that address teacher and/or evaluator satisfaction with teacher evaluation systems structured by standards-based rubrics for monitoring growth.

Teachers in this study perceived the experience to foster their students' reflective practice; however, the research questions for this study were not focused on students' reflective practice. Additional research could explore students' perceptions of their experiences. Further study could also include a quantitative approach to examine the impact of implementing teacher evaluation systems that monitor student growth with standards-based rubrics on students' state assessment results.

Because this study was limited to one midwestern state of the United States, the study could be replicated in other states or nations where SLO are approved or promoted for monitoring student growth for teacher evaluation. The volunteers in this study were from rural and suburban settings, but no volunteers emerged from urban settings. Further study could explore whether teachers in urban settings have similar experiences to the teachers in this study.

Future researchers considering exploration of standards-based rubrics may find it necessary to define the construct for teachers and evaluators as to what constitutes a standards-based rubric. Because one volunteer for this study did not realize that her rubric was not actually standards-based until she described the tool in her interview, it is possible for educators to use a rubric that is not standards-based without recognizing that the rubric must include the actual language of the standards. For this study, participants needed to have used a rubric that included a sequence of performance level descriptors that represented a learning progression of one or more standards. Participants in this study shared a common definition of standards-based growth as movement across the performance levels of the rubric. However, because Briggs et al (2015) noted a lack of clarity regarding definitions of student growth and Close, et al. (2020) noted a variety of definitions of SLO, further research may explore educators' definitions of standards-based rubrics, student growth, or SLO.

Implications

Evidence from this study indicates that the monitoring of student growth in an SLO process for teacher evaluation can promote educators' reflective practice and PCK when standards-based rubrics are used as the structure. All participants shared multiple examples of the reflection-in-action and reflection-on-action that were prompted by the rubric-based process. Teachers shared examples of how such reflections support their PCK by deepening their knowledge of mathematics expectations, supporting them to use evidence-based assessment practices, and guiding their planning and implementation of student-centered instruction. Other studies of SLO implementation found that the lack of

clarity and structure impeded teachers' use of assessment data for instructional planning (Riordan, et al., 2015; Riordan, et al., 2016; Slotnik et al., 2014; Woulfin et al., 2016). However, the teachers in this study emphasized how the standards-based rubrics provide clarity for both assessment and instruction. Therefore, the findings of this study indicate that using standards-based rubrics may address this previously identified challenge in SLO implementation.

Participants in this study shared a common definition of student growth as movement across the performance levels of the standards-based rubric, which agrees with Mosher's (2011) interpretation of growth. The rubric-based approach to monitoring growth defines student growth qualitatively, in accordance with Maul's (2015) recommendation that teachers use performance level descriptions to monitor growth on attributes. This interpretation differs from the strictly quantitative measurement of growth used in statistical models. Teachers reported that the qualitative descriptions provided them guidance for evidence-based assessment, descriptive feedback, and student-centered instruction. Although the findings from this study confirm other SLO research findings, the lack of a consistent interpretation of student growth in SLO contexts continues to pose a challenge for interpreting and comparing SLO research findings (Close, et al., 2020). However, the findings from this study indicate a clear connection between teachers' use of standards-based rubrics to monitor growth and other researchers' recommendations of interpreting growth as movement along learning progressions (Briggs et al., 2015; Fletcher et al., 2017; Furtak et al., 2018; Herman et al., 2011; Hess, 2012).

Amrein-Beardsley and Holloway (2017) found no evidence that using statistical models for monitoring growth, such as VAM and student growth percentiles, enhances teachers' abilities or increases student growth. However, the teachers in this study perceived the process of monitoring student growth to enhance their reflective practice, their PCK, their knowledge of students, and their students' reflective practice. Thus, the findings of this study may be considered in the design or refinement of teacher evaluation systems. Educators seeking to improve the effectiveness of teacher evaluation systems for promoting teachers' knowledge of standards, knowledge of students, assessment practices, or instructional practices, may consider implementing student growth monitoring structured by standards-based rubrics in their school districts. The findings from this study may be used to promote social change in student growth monitoring practices for teacher evaluation at individual, team, school, district, or state levels. Researchers have found that teachers' reflective practice (Lang et al., 2014), increased content and PCK (Hill, et al., 2005), and improved knowledge of students (Hill & Chin, 2018) have all been linked with student achievement gains. Therefore, designing teacher evaluation systems to foster teachers' reflective practice, content knowledge, PCK, and knowledge of students also have the potential for supporting student growth.

Conclusion

Using standards-based rubrics to monitor student growth has the potential to influence the design of formative teacher evaluation systems. This study explored teacher perceptions of their experiences using standards-based rubrics to monitor student growth in an SLO process for teacher evaluation. All participants perceived the experience to

support their reflective practice and increase PCK in mathematics. Teachers reported that they engaged in collaborative dialogue with peers and evaluators, prompting reflection on standards, assessment, and instruction. In addition, teachers perceived the experience to increase their knowledge of students, promote their students' reflective practice, and cultivate a positive teacher evaluation experience. Engaging in teacher evaluation with this structure supported these teachers to improve their practice and supported their students to improve their understanding of the targeted standards.

Multiple researchers have argued that supportive teacher evaluation structures lead to positive changes in teacher practices (Ford et al., 2018; Goe et al., 2017; Roberson-Kraft & Zhang, 2018). Researchers have also argued that using an SLO process for monitoring student growth has potential for supporting both teacher and student growth (Marion et al, 2012; Slotnik et al., 2014, Slotnik et al., 2015). However, researchers have identified challenges in SLO implementation due to a lack of structure (Crouse et al., 2016; Plecki et al., 2016; Slotnik et al., 2015). With no evidence that statistical models support teacher growth or student growth, a viable alternative that supports teacher growth and student growth must be found. Participants in this study indicated that using standards-based rubrics to monitor student growth provides the necessary structure to make implementing an SLO process manageable and meaningful for teachers and students. This is especially significant in response to challenges that have emerged from the COVID-19 pandemic, as this system supports teachers in assessing learning gaps and guiding instruction for learning gaps that may have arisen during pandemic learning. The findings of this study present one possible strategy for improving

the teacher evaluation system with a focus on leveraging actionable data around standards-based growth to benefit teachers and students.

References

Alexander, N. A., Jang, S. T., & Kankane, S. (2017). The performance cycle: The association between student achievement and state policies tying together teacher performance, student achievement, and accountability. *American Journal of Education*, *123*(3), 413–446. https://doi.org/10.1086/691229

Alonzo, A. C. (2011). Learning progressions that support formative assessment practices. *Measurement*, *9*(2–3), 124–129. https://doi.org/10.1080/15366367.2011.599629

Alonzo, A. C. (2017). Tracing the assessment triangle for formative assessment: Not all learning progressions are created equal. *Measurement: Interdisciplinary Research and Perspectives*, *15*(3–4), 163–167. https://doi.org/10.1080/15366367.2017.1404369

Alonzo, A. C. (2018). Exploring the learning progression–formative assessment hypothesis. *Applied Measurement in Education*, *31*(2), 101–103. https://doi.org/10.1080/08957347.2017.1408625

American Alliance for Theater & Education. (2018). *National standards.* https://www.aate.com/national-standards

American Educational Research Association. (2015). AERA statement on use of value-added models for evaluation of educators and educator preparation programs. *Educational Researcher, 44*(8), 448–452. https://doi.org/10.3102/0013189X15618385

Amrein-Beardsley, A., & Holloway, J. (2017). Value-added models for teacher evaluation and accountability: Commonsense assumptions. *Educational Policy,*

33(3), 516–542. https://doi.org/10.1177/0895904817719519

Andrade, H. L. (2013). Classroom assessment in the context of learning theory and research. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment*, (pp. 17–34). SAGE Publications. https://doi.org/10.4135/9781452218649.n2

Arieli-Attali, M., & Cayton-Hodges, G. (2014). Expanding the CBAL™ mathematics assessments to elementary grades: The development of a competency model and a rational number learning progression. *ETS Research Report Series, 2014*(1), 1–41. https://doi.org/10.1002/ets2.12008

Arieli-Attali, M., & Liu, Y. (2016). Beyond correctness: Development and validation of concept-based categorical scoring rubrics for diagnostic purposes. *Educational Psychology, 36*(6), 1083–1101. https://doi.org/10.1080/01443410.2015.1031088

Aspen Institute. (2013). *Implementation of the Common Core State Standards: A transition guide for school level leaders.* https://www.aspeninstitute.org/wp-content/uploads/2013/09/ImplementationCommonCoreStateStandards.pdf

Avalos-Bevan, B. (2018). Teacher evaluation in Chile: Highlights and complexities in 13 years of experience. *Teachers and Teaching, 24*(3), 297–311. https://doi.org/10.1080/13540602.2017.1388228

Backes, B., Cowan, J., Goldhaber, D., Koedel, C., Miller, L. C., & Xu, Z. (2018). The common core conundrum: To what extent should we worry that changes to assessments will affect test-based measures of teacher performance? *Economics of Education Review, 62*, 48–65.

https://doi.org/10.1016/j.econedurev.2017.10.004

Balch, R., & Springer, M. G. (2015). Performance pay, test scores, and student learning objectives. *Economics of Education Review*, *44*, 144–125. https://doi.org/10.1016/j.econedurev.2014.11.002

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher 44*(2), 77–86. https://doi.org/10.3102/0013189X15574904

Barrett-Tatum, J., & Smith, J. M. (2018) Questioning reform in the standards movement: professional development and implementation of common core across the rural South, *Teachers and Teaching, 24*(4), 384–412. https://doi.org/10.1080/13540602.2017.1401534

Bell, A., & Mladenovic, R. (2015). Situated learning, reflective practice and conceptual expansion: effective peer observation for tutor development. *Teaching in Higher Education*, *20*(1), 24–36. https://doi.org/10.1080/13562517.2014.945163

Bergin, C. (2015). *Using student achievement data to evaluate teachers.* [White paper]. https://nee.missouri.edu/documents/NEE_White_ PaperStudent_Achievement_in_TeacherEvaluation2015_5_20.pdf

Berliner, D. C. (2018). Between Scylla and Charybdis: Reflections on and problems associated with the evaluation of teachers in an era of metrification. *Education Policy Analysis Archives, 26*(54), 1–25. https://doi.org/10.14507/epaa.26.3820

Biber, A. C. & Incikabi, L. (2016). Problems posed by prospective elementary

mathematics teachers in the concept of functions: An analysis based on SOLO Taxonomy. *Mersin University Journal of the Faculty of Education, 12*(3), 796–809. https://doi.org/10.17860/mersinefd.282381

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press.

Biggs, J. B., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). Society for Research into Higher Education & Open University Press.

Birkhead, S., Gerasimova, D., Suh, J., & Seshaiyer, P. (2017). Improving knowledge of algebraic learning progressions through professional learning in collaborative vertical teams. In E. Galindo, & J. Newton (Eds.) *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Hoosier Association of Mathematics Teacher Educators.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81-90. https://doi.org/10.1177/003172171009200119

Black, P., Wilson, M., Yao, S.Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement, 9*(2/3), 71-123. https://doi.org/10.1080/15366367.2011.591654

Bolyard, C. (2016). Test-based teacher evaluations: Accountability vs. Responsibility. *Philosophical Studies in Education, 46,* 74-82. https://files.eric.ed.gov/fulltext/EJ1076457.pdf

Boston, M. D., Henrick, E. C., Gibbons, L. K., Berebitsky, D., & Colby, G. T. (2017). Investigating how to support principals as instructional leaders in mathematics. *Journal of Research on Leadership Education, 12*(3), 183-214. https://doi.org/10.1177/1942775116640254

Bowen, T. (2017). Assessing visual literacy: A case study of developing a rubric for identifying and applying criteria to undergraduate student learning, *Teaching in Higher Education, 22*(6), 705-719. https://doi.org/10.1080/13562517.2017.1289507

Brabrand, C., & Dahl, B. (2009). Using the SOLO taxonomy to analyze competence progression of university science curricula. *Higher Education, 58,* 531-549. https://doi.org/10.1007/s10734-009-9210-4

Bradley, J. (2014). From gotcha to growth: How principals promote learning in the context of teacher evaluation. *Journal of Staff Development, 35*(6), 10-14. https://eric.ed.gov/?id=EJ1046805

Bradley-Levine, J., Romano, G., & Reichart, M. (2017) Teacher leaders' influence on teachers' perceptions of the teacher evaluation process. *International Studies in Educational Administration, 45*(3), 65-84.

https://www.research.ed.ac.uk/portal/files
/50477872/Torrance_Murphy_ISEA2017PolicyFudgeAndPracticeRealities.pdf

Braun, H. (2015). The value in value added depends on the ecology. *Educational Researcher, 44*(2), 127-131. https://doi.org/10.3102/0013189X15576341

Bridich, S. M. (2015). The invisible schism: Teachers' and administrators' differing perceptions of education reforms. *Education Policy Analysis Archives, 24*(87) 1-22. http://dx.doi.org/10.14507/epaa.24.219

Briggs, D. C. (2013) Teacher Evaluation as Trojan Horse: The Case for Teacher-Developed Assessments, *Measurement: Interdisciplinary Research and Perspectives*, 11(1-2), 24-29. https://doi.org/10.1080/15366367.2013.784153

Briggs, D.C., Diaz-Bilello, E., Peck, F., Alzen, J., Chattergoon, R., & Johnson, R. (2015). *Using a learning progression framework to assess and evaluate student growth.* Center for Assessment Design Research and Evaluation (CADRE) and Center for Assessment. https://files.eric.ed.gov/fulltext/ED561889.pdf

Briggs, D. C., & Peck, F. A. (2015). Using learning progressions to design vertical scales that support coherent inferences about student growth. *Measurement, 13*(3/4), 75-99. https://doi.org/10.1080/15366367.2015.1042814

Brown, N. J. S., Afflerbach, P. P., and Croninger, R. G. (2014). Assessment of critical-analytic thinking. *Educational Psychology Review, 26*, 543-560. https://doi.org/10.1007/s10648-014-9280-4

Busi, R., & Jacobbe, T. (2018). The impact of analyzing student work on preservice teachers' content knowledge and beliefs about effective mathematics teaching. *IUMPST: The Journal, 1*, 1-20. https://doi.org/10.1007/s11858-020-01145-x

Callahan, K. & Sadeghi, L. (2015). Teacher perceptions of the value of teacher evaluations: New Jersey's ACHIEVE NJ. *NCPEA International Journal of Educational Leadership Preparation, 10*(21), 46-57. https://files.eric.ed.gov/fulltext/EJ1060978.pdf

Camburn, E. M., & Han, S. W. (2015). Infrastructure for teacher reflection and instructional change: An exploratory study. *Journal of Educational Change, 16*(4), 511-533. https://doi.org/10.1007/s10833-015-9252-6

Camburn, E. M., & Han, S. W. (2017). Teachers' professional learning experiences and their engagement in reflective practice: a replication study. *School Effectiveness and School Improvement, 28*(4), 527-554. https://doi.org/10.1080/09243453.2017.1302968

Caniglia, J. C., & Meadows, M. (2018). An application of the SOLO taxonomy to classify strategies used by pre-services teachers to solve "one question problems."

Australian Journal of Teacher Education, 43(9), 75-89. https://doi.org/10.14221/ajte.2018v43n9.5

Cannata, M., Rubin, M., Goldring, E., Grissom, J. A., Neumerski, C. M., Drake, T. A., & Schuermann, P. (2017). Using teacher effectiveness data for information-rich hiring. *Educational Administration Quarterly, 53*(2), 180-222. https://doi.org/10.1177/001316X16881629

Cardno, C. Bassett, M., & Wood, C. (2017). A tale of two secondary schools' efforts to embed 'teaching as inquiry' within an appraisal system. *Leading & Managing, 23*(1), 12-24. https://doi.org/10.3316/informit.069791868368680

Carney, M. B., Brendefur, J. L., Thiede, K., Hughes, G., & Sutton, J. (2016). Statewide mathematics professional development: Teacher knowledge, self-efficacy, and beliefs. *Educational Policy, 30*(4), 539-572. https://doi.org/10.1177/0895904814550075

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal, 97*(1), 3-20. https://doi.org/10.1007/s10649-021-10103-7

Casalaspi, D. (2017). The making of a "legislative miracle": The Elementary and Secondary Education Act of 1965. *History of Education Quarterly, 57*(2), 247-277. https://doi.org/10.1017/heq.2017.4

Center for Assessment. (2017). *SLO Toolkit: Student learning objective toolkit.* http://www.nciea.org/library/recent-publications/slo-toolkit

Champ, C. H. (2015). Measuring teacher effectiveness: The impact of institutional culture on initial implementation of New York's annual professional performance review. *AASA Journal of Scholarship and Practice, 12*(2), 34-43. https://www.aasa.org/uploadedFiles/Publications/JPS-Summer2015.pdf#page=34

Chan, C. C., Tsui, M. S., & Chan, M. Y. C. (2002). Applying the structure of the observed learning outcomes (SOLO) taxonomy on student learning objectives: An empirical study. *Assessment & Evaluation in Higher Education, 27*(6), 511-527. https://doi.org/10.1080/0260293022000020282

Chapman, O., & Koh, K. (2017). Preservice teachers' development of knowledge of authentic assessment mathematics tasks. In Galindo, E., & Newton, J., (Eds.). *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Hoosier Association of Mathematics Teacher Educators.

Chenail, R. J. (2011). Interviewing the investigator: Strategies for addressing instrumentation and researcher bias concerns in qualitative research. *The Qualitative Report, 16*(1), 255-262. https://doi.org/10.46743/2160-3715/2011.1051

Clements, D. H. (2011). Learning trajectories: Foundations for effective research-based education. In Wiest, L. R., & Lamberg, T. (Eds.). *Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* University of Nevada, Reno.

Close, K., Amrein-Beardsley, A., & Collins, C. (2020). Putting teacher evaluation systems on the map: An overview of state's teacher evaluation systems post–Every Student Succeeds Act. *Education Policy Analysis Archives, 28*(58), 1-27. https://doi.org/10.14507/epaa.28.5252

Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The Common Core State Standards and implementation research. *Educational Researcher, 45*(4), 243-251. https://doi.org/10.3102/0013189X16651080

Coon-Kitt, M. J., Nolan, J. F., Lloyd, G. M., Romig, G. (2015). Professional development school triads inquiring about student work in elementary mathematics. *School-University Partnerships, 8*(2), 33-40. https://files.eric.ed.gov/fulltext/EJ1085180.pdf

Confrey, J., Jones, R. S., & Gianopulos, G. (2015). Challenges in modeling and measuring learning trajectories. *Measurement, 13*(3/4), 100-105. https://doi.org/10.1080/15366367.2015.1055131

Cooley, A. (2013). Qualitative research in education: The origins, debates, and politics, of creating knowledge. *Educational Studies, 49*, 247-262. https://doi.org/10.1080/00131946.2013.783834

Council of Chief State School Officials and National Governors Association. (2010). Common core state standards. http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage.

Croft, S. J., Roberts, M. A., & Stenhouse, V. L. (2015). The perfect storm of education reform: High-stakes testing and teacher evaluation. *Social Justice, 42*(1), 70-92. http://www.jstor.org/stable/24871313

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.

Cuervas, R., Ntoumanis, N., Fernandez-Bustos, J. G., & Bartholomew, K. (2018). Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? *Journal of School Psychology, 68,* 154-162. https://doi.org/10.1016/j.jsp.2018.03.005

Cushing, E., & Meyer, C. (2014). *Balancing autonomy and comparability: State approaches to assessment selection for student learning objectives*. Center on Great Teachers and Leaders.

Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education, 57*(3). 300-314. https://doi.org/10.1177/0022487105285962

Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning.* Council of Chief State School Officers.

Darling-Hammond, L. (2016). Research on teaching and teacher education and its influences on policy and practice. *Educational Researcher, 45*(2), 83-91. https://doi.org/10.3102/0013189X16639597

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher, 44*(2), 132-137. https://doi.org/10.3102/0013189X15575346

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, R. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15. https://doi.org/10.1177/003172171209300603

Daro, P., Mosher, F. A. & Corcoran, T. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction.* CPRE

Research Report #RR-68. Consortium for Policy Research in Education. https://doi.org/10.12698/cpre.2011.rr68

Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., Van Petegem, P. (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

Deneen, C. C., & Brown. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education, 3,* 1-14. https://doi.org/10.1080/2331186X.2016.1225380

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2013). *Strategies of qualitative inquiry.* Sage.

Derrington, M. L. (2016). Implementing teacher evaluation: Lattice of leadership. *Journal of Research on Leadership Education, 11*(2), 181-199. https://doi.org/10.1177/1942775116658689

Derrington, M. L., & Kirk, J. (2017). Linking job-embedded professional development and mandated teacher evaluation: Teacher as learner. *Professional Development in Education, 43*(4), 630-644. https://doi.org/10.1080/19415257.2016.1231707

Dewey, J. (1916/2011). *Democracy and education.* Simon & Brown.

Dewey, J. (1938). *Experience and education.* Simon & Schuster.

Dewey, J. (1910). *How we think.* D. C. Heath & Co.

Dewey, J. (1933/1998). *How we think: A restatement of the relation of reflective thinking to the educative process.* Heath.

Dewey, J. (1904). The relation of theory to practice in education. In Murray, C. A. (Ed.) *The third yearbook of the National Society for the Scientific Study of Education*. Public School Pub. Co. https://archive.org/details/r00elationoftheorynatirich

Doherty, K. M., & Jacobs, S. (2015). *State of the states 2015: Evaluating teaching, leading and learning.* National Council on Teacher Quality.

Donaldson, M. L., & Woulfin, S. (2018). From tinkering to going "rogue": How principals use agency when enacting new teacher evaluation systems. *Educational Evaluation and Policy Analysis, 20* (10), 1-26. https://doi.org/10.3102/0162373718784205

Dunne, T. T. (2011). Road maps for learning: A bird's eye view. *Measurement, 9*(2/3), 134-137. https://doi.org/10.1080/15366367.2011.599639

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education, 47*(2), 123-182. https://doi.org/10.1080/03057267.2011.604476

Earl, K., & Ussher, B. (2016). Reflective practice and inquiry: Let's talk more about inquiry. *Teachers and Curriculum, 16*(2), 47-54. https://doi.org/10.15663/tandc.v16i2.139

Egalite, A. J., Fusarelli, L. D., & Fusarelli, B. C. (2017). Will decentralization affect educational inequity? The Every Student Succeeds Act. *Educational Administration Quarterly, 53*(5), 757-781. https://doi.org/10.1177/0013161X17735869

Elementary and Secondary Education Act of 1965: H. R. 2362. Pub. L. No. 89-10. (1965).

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE Open*, 1-10. https://doi.org/10.1177/2158244014522633

Endacott, J. L., Collett, V., Goering, C. Z., Turner, R., Denny, G. S., & Wright, G., Jennings-Davis, J. (2016). On the frontline of CCSS implementation: A national study of factors influencing teachers' perceptions of teaching conditions and job satisfaction. *Cogent Education, 3*(1), 1-25. https://doi.org/10.1080/2331186X.2016.1162997

Enderson, M.C., Grant, M.R., & Liu, Y. (2018). Supporting mathematics coaches' learning of probability through professional development tasks. *International Journal of Research in Education and Science (IJRES), 4*(2), 613-632. https://doi.org/10.21890/ijres.428976

Engelhard, G., & Sullivan, R. K. (2011). An ecological perspective on learning progressions as road maps for learning. *Measurement, 9*(2/3), 138-145. https://doi.org/10.1080/15366367.2011.603615

Eisner, E. W. (2017). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice.* Teachers College Press.

Estaji, M., & Dezfoolian, S. (2018). EFL teacher's pedagogical knowledge base as a predictor of teacher's reflectivity: Comparing different components and perceptions. *International Journal of Instruction, 11*(3), 491-510. https://doi.org/10.12973/iji.2018.11334a

Everson, K. C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? *Review of Educational Research, 87*(1), 35-70. https://doi.org/10.3102/0034654316637199

Farrell, T. S. C., & Vos, R. (2018). Exploring the principles and practices of one teacher of L2 speaking: The importance of reflecting on practice. *Iranian Journal of Language Teaching Research, 6*(1), 1-15. https://files.eric.ed.gov/fulltext/EJ1165502.pdf

Filippi, J. R. & Hackman, D. G. (2019). Leading Common Core State Standards implementation: Lessons from one successful superintendent. *Leadership and Policy in Schools.* https://doi.org/10.1080/15700763.2017.1398334

Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher, 43*(2), 100-107. https://doi.org/10.3102/0013189X14521864

Firestone, W. A. & Donaldson, M. L. (2019). Teacher evaluation as data use: What recent research suggests. *Educational Assessment, Evaluation and Accountability, 31*(30), 289-314. https://doi.org/10.1007/s11092-019-09300-z

Fives, H., & Buehl, M. M. (2016). Teachers' beliefs, in the context of policy reform. *Policy Insights from the Behavioral and Brain Sciences,* 3(1), 114-121. https://doi.org/10.1177/2372732215623554

Fletcher, N., Ebby, C., & Hulbert, B. (2017). Developing formative assessment tools and routines for additive learning. In E. Galindo, & J. Newton (Eds.) *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Hoosier Association of Mathematics Teacher Educators.

Floden, R. E., Richmond, G., & Andrews, D. C. (2017). Responding to the challenge of new standards. *Journal of Teacher Education, 68*(3), 236-238. https://doi.org/10.1177/0022487117702380

Fonger, N. L. (2017). Characterizing sophistication in representational fluency. In E. Galindo, & J. Newton (Eds.). *Proceedings of the 39th annual meeting of the*

North American Chapter of the International Group for the Psychology of Mathematics Education. Hoosier Association of Mathematics Teacher Educators.

Fonger, N. L., Stephens, A., Blanton, M., Isler, I., Knuth, E., & Gardiner, A. M. (2018). Developing a learning progression for curriculum, instruction, and student learning: An example from mathematics education. *Cognition and Instruction, 36*(1), 30-55. https://doi.org/10.1080/07370008.2017.1392965

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). https://doi.org/10.14507/epaa.28.5024

Ford, T. G., Urick, A., & Wilson, A. S. P. (2018). Exploring the effect of supportive teacher evaluation experiences on U.S. teachers' job satisfaction. *Education Policy Analysis Archives, 26*(59), 1-32. https://doi.org/10.14507/epaa.26.3559

Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy, 31*(2), 202-248. https://doi.org/10.1177/0895904815586855

Fuller, E. J., Hollingsworth, L., Pendola, A. (2017). The Every Student Succeeds Act, state efforts to improve access to effective educators and importance of school leadership. *Educational Administration Quarterly, 53*(5), 727-756. https://doi.org/10.1177/0013161X17711481

Furtak, E. M., Circi, R. & Heredia, S. C. (2018). Exploring alignment among learning progressions, teacher-designed formative assessment tasks, and student growth: Results of a four-year study. *Applied Measurement in Education, 31*(2), 143-156. https://doi.org/10.1080/08957347.2017.1408624

Fusch, P. I., & Ness, L. R. (2015). Are We There Yet? Data Saturation in Qualitative Research. *The Qualitative Report, 20*(9), 1408-1416. https://cpb-us-e1.wpmucdn.com/sites.nova.edu/dist/a/4/files/2015/09/fusch1.pdf

Gabriel, R. (2017). Rubrics and reflection: A discursive analysis of observation debrief conversations between novice Teach for America teachers and mentors. *Action in Teacher Education, 39*(1), 85-102. https://doi.org/10.1080/01626620.2016.1245636

Gagani, R. F. M., & Misa R. O. (2017). Solo Based-cognition levels of inductive reasoning in geometry. *Alberta Journal of Educational Research, 63*(4), 344-356. http://doi.org/10.11575/ajer.v63i4.56331

Gagnon, D. J., Hall, E. L., & Marion, S. (2017). Teacher evaluation and local control in the US: An investigation into the degree of local control afforded to districts in defining evaluation procedures for teachers in non-tested subjects and grades. *Assessment in Education: Principles, Policy & Practice*, 24(4), 489-505. https://doi.org/10.1080/0969594X.2016.1167669

Gareis, C. R. & Grant, L. W. (2015). *Teacher-made assessments: How to connect curriculum, instruction, and student learning* (2nd ed.). Routledge.

Garet, M. S., Wayne, A. J., Brown, S., Rickles, J., Song, M., & Manzeske, D. (2017). *The Impact of Providing Performance Feedback to Teachers and Principals* (NCEE 2018-4001). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garrison, M. J. (2011). On the origin and political significance of test-based teacher evaluation and compensation. *Journal of Inquiry & Action in Education, 4*(1) 48-67. Retrieved from https://files.eric.ed.gov/fulltext/EJ1134640.pdf

Geiger, T., & Amrein-Beardsley, A. (2017). Administrators gaming test- and observation-based teacher evaluation methods: To conform to or confront the system. *AASA Journal of Scholarship & Practice, 14*(3), 45-54. https://aasa.org/uploadedFiles/Publications/JSP-Fall-2017.pdf#page=45

Gill, B., Bruch, J., and Booker, K. (2013). *Using alternative student growth measures for evaluating teacher performance: What the literature says*. U.S. Department of Education. https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=369

Gill, B., English, B., Furgeson, J., & McCullough, M. (2014). *Alternative student growth measures for teacher evaluations: Profiles of early-adopting districts.* (REL 2014–016). U.S. Department of Education, Institute of Education Sciences,

National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid- Atlantic. http://ies.ed.gov/ncee/edlabs.

Gilles, J. F. (2017). “It’s not a gotcha”: Interpreting teacher evaluation policy in rural school districts. *Rural Educator, 38*(2), 11-22. https://doi.org/10.35608/ruraled.v38i2.224

Gillies, D. (2016). Visiting good company: Arendt and the development of the reflective practitioner. *Journal of Educational Administration and History, 48*(2), 148-159. https://doi.org/10.1080/00220620.2016.1144576

Gitomer, D. H. (2011). Road maps for learning and teacher evaluation. *Measurement, 9*(2/3), 146-148. https://doi.org/10.1080/15366367.2011.603616

Goe, L., Wylie, E. C., Bosso, D, & Olson, D. (2017). *State of the states’ teacher evaluation and support systems: A perspective from exemplary teachers.* ETS Research Report Series, *2017*, 1-27. https://doi.org/10.1002/ets2.12156

Golberg, C. (2018). The Massachusetts educator evaluation system and teacher perceptions of professional growth. *The Graduate Review, 3,* 69-90. http://vc.bridgew.edu/grad_rev/vol3/iss1/15

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce [Special issue]. *Educational Researcher, 44*, 87–95. https://doi.org/10.3102/0013189X15574905

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principal's human capital decisions and the emergence of teacher observation data. *Educational Researcher, 44*(2), 96-104. https://doi.org/10.3102/0013189X15575031

Gotwals, A. W. (2018). Where are we now? Learning progressions and formative assessment, *Applied Measurement in Education, 31*(2), 157-164. https://doi.org/10.1080/08957347.2017.1408626

Graf, E. A., & Arieli-Attali, M. (2015). Designing and developing assessments of complex thinking in mathematics for the middle grades. *Theory into Practice, 54*(3), 195-202. https://doi.org/10.1080/00405841.2015.1044365

Griggs, V., Holden, R., Lawless, A., & Rae, J. (2018). From reflective learning to reflective practice: Assessing transfer. *Studies in Higher Education, 43*(7), 1172-1183. https://doi.org/10.1080/03075079.2016.1232382

Haertel, E. H. (2013). *Reliability and validity of inferences about teachers based on student scores.* Lecture presented at the National Press Club. Washington, D.C.

Haj Sassi, S. B. (2016). Teachers' perceptions of self-observation as an appraisal tool. *Arab World English Journal, 7*(1), 341-357. https://doi.org/10.2139/ssrn.2804017

Haladyna, T. (2011). Using student achievement tests to evaluate teachers—A very bad idea. *Nonpartisan Education Review / Essays, 7*(2). http://www.nonpartisaneducation.org/Review/Essays/v7n2.pdf

Hall, E., Diaz-Bilello, E., & Marion, S. (2015). *Considerations for establishing performance standards for educator evaluation*. National Center for the Improvement of Educational Assessment. https://www.nciea.org/sites/default/files/publications/Establishing_Performance_Standards_for_EES_EH2015.pdf

Hall, E., Gagnon, D., Thompson, J., Schneider, M. C., & Marion, S. (2014). *State practices related to the use of student achievement measures in the evaluation of teachers in non-tested subjects and grades.* National Center for the Improvement of Educational Assessment.

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *Npj Science of Learning, 1*, 1-13. https://doi.org/10.1038/npjscilearn.2016.13

Hattie, J. A., & Purdie, N. (1994). Using the SOLO taxonomy to classify test items. In B. Dart & G. Boulton-Lewis (Eds.), *Teaching and Learning in Higher Education* (pp. 145-176). Australian Council of Educational Research.

Hattie, J. & Purdie, N. (1998) The Solo model: Addressing fundamental measurement issues. In Dart, B. & Boulton-Lewis, G. M. (Eds.) *Teaching and learning in higher education.* Australian Council of Educational Research.

Hegazy, H., & Barton, G. (2017). Formative assessment in the middle years: A review of literature and alignment with the guiding principles for junior secondary. *Adolescent Success, 17*(2), 6-19. https://doi.org/10.1016/j.stueduc.2007.01.003.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* SAGE.

Heritage, M. (2011). Commentary on road maps for learning: A guide to the navigation of learning progressions. *Measurement, 9*(2/3), 149-151. https://doi.org/10.1080/15366367.2011.599647

Herman, J. L., Heritage, M., & Goldschmidt, P. (2011). *Developing and selecting assessments of student growth for use in teacher evaluation systems* (extended version). University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hess, K. K. (2011). It's time for more focus on educator involvement in developing and using learning progressions. *Measurement, 9*(2/3), 152-154. https://doi.org/10.1080/15366367.2011.605039

Hess, K. K. (2012). *Learning progressions in K-8 classrooms: How progress maps can influence classroom practice and perceptions and help teachers make more*

informed instructional decisions in support of struggling learners. (Synthesis Report 87). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Hewitt, K. K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives, 23*(76). https://doi.org/10.14507/epaa.v23.1968

Hewitt, K. K., & Amrein-Beardsley, A. (2016). The use of student growth measures for educator accountability at the intersection of policy and practice. In A. Amrein-Beardsley & K. Kappler Hewitt (Eds.), *Student growth measures: Where policy meets practice.* Palgrave Macmillan.

Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal, 55*(5), 1076-1112. https://doi.org/10.3102/0002831218769614

Hill, H. C., Rowan, B., Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal 42*(2), 371-406. https://doi.org/10.3102/00028312042002371

Hocket, J. A., & Doubet, K. J. (2014). Turning on the lights: What pre-assessments can do. *Educational Leadership, 71*(4), 50-54.

Holdheide, L., Browder, D., Warren, S., Buzick, H., & Jones, N. (2012). *Using student growth to evaluate educators of students with disabilities: Issues, challenges, and next steps.* National Comprehensive Center for Teacher Quality. https://eric.ed.gov/?id=ED543814

Holloway-Libell, J. (2016). What counts as good teaching? How a student growth percentile model has defined teacher quality at one urban middle school. In A. Amrein-Beardsley & K. Kappler Hewitt (Eds.), *Student growth measures: Where policy meets practice* (pp. 137-152). Palgrave Macmillan.

Huber, S. G., & Skedsmo, G. (2016). Teacher evaluation—accountability and improving teaching practices. *Educational Assessment, Evaluation, and Accountability, 28,* 105-109. https://doi.org/10.1007/s11092-016-9241-1

Humphrey, D. C., Gallagher, H. A., Yee, K. M., Goss, G. K., Campbell, A. Z., Cassidy, L. J., & Mitchell, N. M. (2012). *Teacher Incentive Fund: First implementation report, 2006 and 2007 grantees.* U.S. Department of Education Office of Planning, Evaluation and Policy Development.

İlhan, M., & Çetin, B. (2016). The identification of the views of raters on standard rubrics and rubrics based on the SOLO taxonomy. *Journal of Theory and Practice in Education, 12*(1), 1-16. https://dergipark.org.tr/tr/download/article-file/262379

İlhan, M., & Gezer, M. (2017). A comparison of the reliability of the Solo- and revised Bloom's Taxonomy-based classifications in the analysis of the cognitive levels of

assessment questions. *Pegem Journal of Education and Instruction*, *7*(4), 637-662. https://doi.org/10.14527/pegegog.2017.023

Illinois State Board of Education [ISBE]. (2013). Fact sheet: Common Core State Standards. https://www.isbe.net/Documents/ccs-fact-sheet-0813.pdf

Illinois State Board of Education [ISBE]. (2014). Illinois teacher survey of standards implementation. https://www.isbe.net/documents/ils-impl-survey-results0214.pdf

Illinois State Board of Education [ISBE]. (2015). Student learning objective guidebook. https://www.isbe.net/Documents/slo-guidebook.pdf

Jacob, B. (2017). The changing federal role in school accountability. *Journal of Policy Analysis & Management*, *36*(2), 469-477. https://doi.org/10.1002/pam.21975

Janesick, V. J. (2011). *Stretching exercises for qualitative researchers* (3rd Ed.). Sage Publications.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students [Special issue]. *Educational Researcher*, *44*, 105-116. https://doi.org/10.3102/0013189X15575517

Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school? [Special issue]. *Educational Researcher, 44*(2),117–126. https://doi.org/10.3102/0013189X15573351

Joyce, J., Harrison, J. R., Murphy, D. (2016). Evaluating students with disabilities and their teachers. Use of student learning objectives. *Teachers College Record, 118*(14), 1-22. http://www.tcrecord.org/Content.asp?ContentId=21542

Jurdak, M. E., & Mouhayar, R. R. E. (2014). Trends in the development of student level reasoning in pattern generalization tasks across grade level. *Educational Studies in Mathematics, 85*(1), 75-92. https://doi.org/10.1007/s10649-013-9494-2

Kaimal, G., & Jordan, W. (2016). Do incentive-based programs improve teacher quality and student achievement? An analysis of implementation in 12 urban charter schools. *Teachers College Record, 118*(7). https://eric.ed.gov/?id=EJ1108526

Katz, D. S. (2016). Growth models and teacher evaluation: What teachers need to know and do. *Kappa Delta Pi Record, 52*(1), 1-16. https://doi.org/10.1080/00228958.2016.1123039

Kearns, J. F., Kleinert, H. L., Thurlow, M. L., Gong, B., & Quenemoen, R. (2015). Alternate assessments as one measure of teacher effectiveness: Implications for our field. *Research and Practice for Persons with Severe Disabilities, 40*(1), 20-35. https://doi.org/10.1177/1540796915585105

Keskin, Y., Keskin, S. C., & Kirtel, A. (2016). Examination of the compatibility of questions used by social studies teachers in the class with the program achievements according to the SOLO taxonomy. *Journal of Education and Training Studies, 4*(4), 68-76. https://doi.org/10.11114/jets.v4i4.1286

Kingston, N. M., Broaddus, A., & Lao, H. (2015). Some thoughts on "using learning progressions to design vertical scales that support coherent inferences about student growth." *Measurement, 13*(3/4), 195-199. https://doi.org/10.1080/15366367.2015.1105084

Kobrin, J. L., Larson, S., Cromwell, A., & Garza, P. (2015). A framework for evaluating learning progressions on features related to their intended uses. *Journal of Educational Research and Practice, 5*(1), 58-73. https://doi.org/10.5590/JERAP.2015.05.1.04

Kobrin, J. L. (2016). Learning progressions in action in a middle school: A case study. *Professional Development in Education, 42*(1), 171-173. https://doi.org/10.1080/19415257.2014.963885

Kraft, M. A. & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administrators Quarterly, 52* (5), 711-753. https://doi.org/10.1177/0013161X16653445

Krajcik, J. (2011). Learning progressions provide road maps for the development and validity of assessments and curriculum materials. *Measurement, 9*(2/3), 155-158. https://doi.org/10.1080/15366367.2011.603617

Kuh, L. P. (2016) Teachers talking about teaching and school: Collaboration and reflective practice via Critical Friends Groups. *Teachers and Teaching, 22*(3), 293-314. https://doi.org/10.1080/13540602.2015.1058589

Kusumawathie, P. H., Mohamad, N., & Azam, F. (2017). Teachers' perceptions of classroom practices based on SOLO taxonomy in secondary school system. *European Journal of Alternative Education Studies, 2*(2), 16-39. https://doi.org/10.5281/zenodo.883653

Lachlan-Hache, L. (2015). *The art and science of student learning objectives: A research synthesis.* American Institutes for Research.

Lachlan-Hache, L., Cushing, E., & Bivona, L. (2012). *Student learning objectives as measures of educator effectiveness: The basics.* AIR.

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

Lai, E. R., Kobrin, J. L., DiCerbo, K. E. & Holland, L. R. (2017). Tracing the assessment triangle with learning progression-aligned assessments in mathematics. *Measurement: Interdisciplinary Research and Perspectives, 15*(3-4), 143-162. https://doi.org/10.1080/15366367.2017.1388113

Lalor, J., Lorenzi, F., & Rami, J. (2014). Developing professional competence through assessment: Constructivist and reflective practice in teacher-training. *Eurasian Journal of Educational Research, 58,* 45-66. https://doi.org/10.14689/ejer.2015.58.6

Lang, L. B., Schoen, R. R., LaVenia, M., & Oberlin, M. (2014). *Mathematics formative assessment system – Common Core State Standards: A randomized field trial in kindergarten and first grade.* Presented to SREE Spring 2014 Conference. Society for Research on Educational Effectiveness.

Lash, A., Makkonen, R., Tran, L., & Huang, M. (2016). *Analysis of the stability of teacher-level growth scores from the student growth percentile model*. (REL 2016–104). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. http://ies.ed.gov/ncee/edlabs

Lavigne, A. L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record, 116,* 1-29. http://www.tcrecord.org/Content.asp?ContentId=17294

Lavigne, A. L., & Chamberlain, R. W. (2017). Teacher evaluation in Illinois: School leaders' perceptions and practices. *Educational Assessment, Evaluation, and Accountability*, *29*(2), 179-209. https://doi.org/10.1007/s11092-016-9250-0

Leat, D., & Nichols, A. (2000). Brains on the table: Diagnostic and formative assessment through observation. *Assessment in Education: Principles, Policy & Practice, 7*(1), 103-121. https://doi.org/10.1080/713613327

Lenhoff, S. W., Pogodzinski, B., Mayrowetz, D., Superfine, B. M., & Umpstead, R. R. (2018) District stressors and teacher evaluation ratings. *Journal of Educational Administration, 56*(2), 146-p. https://doi.org/10.1108/JEA-06-2017-0065

Leo, S. F., & Coggshall, J. G. (2013). Creating coherence*: Common Core State Standards, teacher evaluation, and professional learning* [Revised Edition]. https://gtlcenter.org/sites/default/files/CreatingCoherence.pdf

Lichtman, M. (2013). *Qualitative research in education: A user's guide* (3rd Ed.). Sage.

Lillejord, S, Elstad, E, & Kavli, H. (2018) Teacher evaluation as a wicked policy problem. *Assessment in Education: Principles, Policy & Practice, 25(3),* 291-309. https://doi.org/10.1080/0969594X.2018.1429388

Lin, S., Luo, W., Tong, F., Irby, B. J., Alecio, R. L., Rodriguez, L., & Chapa, S. (2020). Data-based student learning objectives for teacher evaluation. *Cogent Education, 7*(1), 1-23. https://doi.org/10.1080/2331186X.2020.1713427

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*, 41-54. https://doi.org/10.1016/j.stueduc.2019.01.006

Loeb, H., Knapp, M. S., & Elfers, A. M. (2008). Teachers' response to standards-based reform: Probing reform assumptions in Washington state. *Education Policy Analysis Archives, 16(*8), 1-32. https://doi.org/10.14507/epaa.v16n9.2008

Longo-Schmid, J. (2016). Teachers' voices: Where policy meets practice. In A. Amrein-Beardsley & K. Kappler Hewitt (Eds.), *Student growth measures: Where policy meets practice* (pp. 49-72). Palgrave Macmillan.

Lopez, P., & Wise, D. (2015). Leading change for the implementation of Common Core State Standards in rural school districts. *NCPEA Educational Leadership Review of Doctoral Research 2*(1), 47-56. https://files.eric.ed.gov/fulltext/EJ1105742.pdf

Makkonen, R., Tejwani, J. and Rodriguez, F. Jr. (2015). *A Descriptive Study of the Pilot Implementation of Student Learning Objectives in Arizona and Utah*. U.S. Department of Education. https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=451

Malunda, P., Onen, D., Musaazi, J. C. S., & Oonyu, J. (2016). Teacher evaluation and quality of pedagogical practices. *International Journal of Learning, Teaching, and Educational Research, 15*(9), 118-133. https://www.ijlter.org/index.php/ijlter/article/view/753/pdf

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

National Center for the Improvement of Educational Assessment.

Marion, S. (2015). *Two sides of the same coin: Competency-based education and student learning objectives.* The National Center for the Improvement of Educational Assessment.

Marion, S. (2016). *Considerations for state leaders in the design of school accountability systems under the Every Student Succeeds Act*. National Center for the Improvement of Educational Assessment.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology, 46,* 4-11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Matlock, K. L., Goering, C. Z., Endacott, J., Collet, V. S., Denny, G. S., Jennings-Davis, J. & Wright, G. P. (2016). Teachers' views of the Common Core State Standards and its implementation. *Educational Review*, *68*(3), 291-305. https://doi.org/10.1080/00131911.2015.1070333

Matthews, M. E. (2017). The influence of the pedagogical content knowledge framework on research in mathematics education: A review across grade bands. *Journal of Education 193*, (3), 29-37. https://doi.org/10.1177/002205741319300305

Maul, A. (2015). Learning progressions, vertical scales, and testable hypotheses: Promising intuitions and points for clarification. *Measurement, 13*(3/4), 118-122. https://doi.org/10.1080/15366367.2015.1055144

McCullough, M., English, B., Angus, M. H., & Gill, B. (2015). *Alternative student growth measures for teacher evaluation: Implementation experiences of early adopting districts.* Institute for Educational Sciences (IES), 1-53. https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=445

McDuffie, A. R., Drake, C., Choppin,, J., Davis, J. D., Magana, M. V., & Carson, C. (2015). Middle school mathematics teachers' perceptions of the Common Core State Standards for mathematics and related assessment and teacher evaluation systems. *Educational Policy, 31*(2), 139-179. https://doi.org/10.1177/0895904815586850

McGuinn, P. (2012). Stimulating reform: Race to the Top, competitive grants, and the Obama Education Agenda. *Educational Policy, 26*(1), 136-159. https://doi.org/10.1177/0895904811425911

McMillan, J. H. (2016). National board certified teachers' perspectives on using measures of student learning for teacher evaluation. *The Educational Forum, 80*(1), 48-60. https://doi.org/10.1080/00131725.2015.1102366

Measured Progress. (2014). *A system for using student academic growth in the evaluation of teaching effectiveness in the non-tested subjects and grades*. Author.

Meierdirk, C. (2016). Is reflective practice an essential component of becoming a professional teacher? *Reflective Practice, 17*(3), 369-378. https://doi.org/10.1080/14623943.2016.1169169

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation.* (4th Ed.). Jossey-Bass.

Mette, I. M., Range, B. G., Anderson, J., Hvidston, D. J., & Nieuwenhuizen, L. (2015). Teacher perceptions of teacher supervision and evaluation: A reflection of school improvement practices in the age of reform. *NCPEA Education Leadership Review, 16*(1), 16-30. https://files.eric.ed.gov/fulltext/EJ1105545.pdf

Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions for Adult & Continuing Education, 74*, 5-12. https://doi.org/10.1002/ace.7401

Milanowski, A., Scott, J. A., Miller, J., Finster, M., Doll, M., Lewandowski, H., Roseland, D., & White, B. R. (2015). Evaluation of the Performance Evaluation Reform Act: Interim report. https://www.isbe.net/Documents/pera-final-report-160630.pdf#search=performance%20evaluation%20reform%20act

Milanowski, A., Ristrow, L., Finster, M., McKithen, C., Doll, M., Lewandowski, H., Roseland, D., White, B. R., Carl, B., Gawade, N., Kang, H., Marlin, D., Meyer, R., Ponisciak, S., Steele, C., & Wang, Y. (2016). *Evaluation of the Performance Evaluation Reform Act: Interim report.* Illinois State Board of Education.

https://www.isbe.net/Documents/pera-final-report-160630.pdf#search=performance%20evaluation%20reform%20act

Mintrop, R., Ordenes, M., Coghlan, E., Pryor, L., & Madero, C. (2018). Teacher evaluation, pay for performance, and learning around instruction: Between dissonant incentives and resonant procedures. *Educational Administration Quarterly, 54*(1), 3-46. https://doi.org/10.1177/0013161X17696558

Mosher, F. A. (September, 2011). The role of learning progressions in standards-based education reform. *CPRE Policy Briefs: Reporting on Issues and Research in Education Policy and Finance.* Consortium for Policy Research in Education. http://www.cpre.org/sites/default/files/policybrief/1218_lppolicybriefwebready.pdf

Muñoz, M. A., & Dossett, D. H. (2016). Multiple measures of teaching effectiveness: Classroom observations and student surveys as predictors of student learning. *Planning and Changing, 47*(3/4), 123–140. https://eric.ed.gov/?id=EJ1145378

Munroe, A. (2017). Measuring student growth within a merit-pay evaluation system: Perceived effects on music teacher motivation career commitment. *Contributions to Music Education 42*, 89-105. https://www.jstor.org/stable/26367438

Murphy, A. F., & Torff, B. (2016). Growing pains: The effect of Common Core State Standards on perceived teacher effectiveness. *The Educational Forum, 80*(1), 21-33. https://doi.org/10.1080/00131725.2015.1102999

National Art Education Association. (2018). *National visual arts standards.* https://www.arteducators.org/learn-tools/national-visual-arts-standards

National Association for Music Education. (2018). *Standards.* https://nafme.org/my-classroom/standards/

National Association of State Boards of Education. (2016). *Next Generation Science Standards*. http://www.nasbe.org/project/next-generation-science-standards/

National Council for the Social Studies. (2018). *College, career, and civic life (C3) framework for social studies state standards.* https://www.socialstudies.org/c3

Newton, G., & Martin, E. (2013). Research and teaching: Blooming, SOLO taxonomy, and phenomenology as assessment strategies in undergraduate science education. *Journal of College Science Teaching, 43*(2), 78-90. https://doi.org/10.2505/4/jcst13_043_02_78.

Nichols, P. (2011). Fulfilling the promise of the learning triangle. *Measurement, 9*(2/3), 163-165. https://doi.org/10.1080/15366367.2011.60570

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Olteanu, C. (2017). Reflection-for-action and the choice or design of examples in the teaching of mathematics. *Mathematics Education Research Journal, 29*(3), 349-367. https://doi.org/10.1007/s13394-017-0211-9

Ozdemir, A. S., & Goktepe Yildiz, S. (2015). The analysis of elementary mathematics preservice teachers' spatial orientation skills with SOLO model. *Eurasian Journal of Educational Research, 61,* 217-236. https://doi.org/10.14689/ejer.2015.61.12

Pak, K., Polikoff, M. S., Desimone, L. M., & García, E. S. (2020). The adaptive challenges of curriculum implementation: Insights for educational leaders driving standards-based reform. *AERA Open,* (2), 1-15. https://doi.org/10.1177/2332858420932828

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*(1), 123-141. https://doi.org/10.17763/haer.82.1.v40p0833345w6384

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Patel, N. (2012). Changing the culture of teaching: Hamilton County's project COACH teacher evaluation system. *District Management Journal, 10*, 22-29. https://dmj.dmgroupk12.com/articles/changing-the-culture-of-teaching-hamilton-county-s-project-coach-teacher-evaluation-system

Patton, M. Q. (2015). *Qualitative research & evaluation methods* [4th ed.]. SAGE.

Paufler, N. A., King, K. M., & Zhu, P. (2020). Promoting professional growth in new teacher evaluation systems: Practitioners' lived experiences in changing policy contexts. *Studies in Educational Evaluation, 65*, 1-9. https://doi.org/10.1016/j.stueduc.2020.100873

Pellegrino, J. W., & Chudowsky, N. (2003). Focus article: The foundations of assessment, *Measurement, Interdisciplinary Research and Perspectives*, *1*(2), 103-148. https://doi.org/10.1207/S15366359MEA0102_01

Pense, S. L., Freeburg, B. W., & Clemons, C. A. (2015). Implementation of Common Core State Standards: Voices, positions, and frames. *Career and Technical Education Research, 40*(3), 157-173. https://doi.org/10.5328/cter40.3.157

Penuel, W. R. (2015). Learning progressions as evolving tools in joint enterprises for educational improvement. *Measurement, 13*(3/4), 123-127. https://doi.org/10.1080/15366367.2015.1055145

Performance Evaluation Reform Act (PERA) of 2010, Public Act 096-0861. (2010). http://www.ilga.gov/legislation/publicacts/96/PDF/096-0861.pdf

Pham, K., & Heineman, A. (2014). Partners with a purpose. *Journal of Staff Development, 35*(6), 40-43, 47. http://search.proquest.com/docview/1651865820?accountid=14780

Pivovarova, M., & Amrein-Beardsley, A. (2018) Median Growth Percentiles (MGPs): Assessment of Intertemporal Stability and Correlations with Observational

Scores, *Educational Assessment*, 23(2), 139-155. https://doi.org/10.1080/10627197.2018.1449634

Plecki, M. L., Elfers, A. M., St. John, E., & Yeh, T. L. (2016). Practitioners' response to Washington's required use of student growth measures in teacher evaluation. In A. Amrein-Beardsley & K. Kappler Hewitt (Eds.), *Student growth measures: Where policy meets practice* (pp. 95-116). Palgrave Macmillan.

Polly, D. (2017). Elementary school teachers' uses of mathematics curricular resources. *Journal of Curriculum Studies, 49*(2), 132-148. https://doi.org/10.1080/00220272.2016.1154608

Popham, W. J. (2018). *Assessment literacy for educators in a hurry.* ASCD.

Popham, W. J. (2013). Can classroom assessments of student growth be credibly used to evaluate teachers? *English Journal, 103*(1), 34-39. https://www.jstor.org/stable/24484058

Popham, W. J. (2007). The lowdown on learning progressions. *Educational Leadership, 64*(7), 83-84. https://www.ascd.org/el/articles/the-lowdown-on-learning-progressions

Popham, W. J. (2008). *Transformative assessment.* Association for Supervision and Curriculum Development.

Prakash, E. S., Narayan, K. A., & Sethuraman, K. R. (2010). Student perceptions regarding the usefulness of explicit discussion of "Structure of the Observed Learning Outcome" taxonomy. *Advances in Physiology Education, 34*, 145-149. https://doi.org/10.1152/advan.00026.2010.

Pressley, T., Roehrig, A. D., & Turner, J. E. (2018). Elementary teachers' perceptions of a reformed teacher-evaluation system. *The Teacher Educator, 53*(1), 21-43. https://doi.org/10.1080/08878730.2017.1391362

Prizovskaya, V. (2018). Identifying New Jersey Teachers' assessment literacy as precondition for implementing student growth objectives. *Journal of Education and Learning, 7*(1), 184-207. https://doi.org/10.5539/jel.v7n1p184

Raudenbush, S. W. (2015). Value added: A case study in the mismatch between education research and policy. *Educational Researcher, 44*(2), 138-141. https://doi.org/10.3102/0013189X15575345

Reform Support Network. (2010) *Measuring student growth for teachers in non-tested grades and subjects: A primer.* ICF International.

Reform Support Network. (n.d.). *Targeting growth: Using student learning objectives as a measure of educator effectiveness.* ICF International.

Reform Support Network. (2014). *A toolkit for implementing high-quality student learning objectives 2.0.* ICF International.

Reilly, A. H. (2018). Using reflective practice to support management student learning: Three brief assignments. *Management Teaching Review, 3*(2), 139-147. https://doi.org/10.1177/2379298117719686

Rembach, L., & Dison, L. (2016). Transforming taxonomies into rubrics: Using SOLO in social science and inclusive education. *Perspectives in Education, 34*(1), 68-83. https://doi.org/10.18820/0258-2236/pie. v34i1.6

Rigby, J. G., Corriell, R., & Kuhl, K. J. (2018). Leading or instructional improvement in the context of accountability: Central office leadership. *Journal of Cases in Educational Leadership, 21*(1), 28-42. https://doi.org/10.1177/1555458917722183

Riordan, J., Lacireno-Paquet, N., Shakman, K., Bocala, C., & Chang, Q. (2015). *Redesigning teacher evaluations: Lessons from a pilot implementation* (REL 2015–030). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. http://ies.ed.gov/ncee/edlabs

Riordan, J., Lacireno-Paquet, N., Shakman, K., & Bocala, C., & Chang, Q. (2016). *Redesigning teacher evaluation: Lessons learned from a pilot implementation in New Hampshire.* Society for Research on Educational Effectiveness. https://eric.ed.gov/?id=ED567482

Ritter, G. W., & Barnett, J. H. (2016). Learning on the job: Teacher evaluation can foster

real growth. *Kappan, 97*(9), 48-52. https://doi.org/10.1177/0031721716641649

Robertson-Kraft, C., & Zhang, R. S. (2018). Keeping great teachers: A case study on the impact and implementation of a pilot teacher evaluation system. *Educational Policy, 32*(3), 363-394. https://doi.org/10.1177/0895904816637685

Rogers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104*(4), 842-866. http://c2l.mcnrc.org/wp-content/uploads/sites/8/2013/05/CarolRodgers-Article.pdf

Rosen, R., & Parise, L. M. (2017). *Using evaluation systems for teacher improvement: Are school districts ready to meet new federal goals?* MDRC. www.mdrc.org/sites/default/files/iPD_ESSA_Brief_2017.pdf

Roussin, J. L., & Zimmerman, D. P. (2014). Inspire learning, not dread: Create a feedback culture that leads to improved practice. *Journal of Staff Development, 35*(6), 36-47. https://learningforward.org/wp-content/uploads/2014/12/inspire-learning-not-dread.pdf

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* [3rd Ed.]. Sage.

Russell, T. (2018). A teacher educator's lessons learned from reflective practice. *European Journal of Teacher Education, 41*(1), 4-14. https://doi.org/10.1080/02619768.2017.1395852

Saldaña, J. (2016). *The coding manual for qualitative researchers.* Sage.

Sarama, J., Clements, D. H., & Spitler, M. E. (2017). Evidence of teacher change after participating in TRIAD's learning trajectories-based professional development and after implementing learning trajectory-based mathematics instruction. *Mathematics Teacher Education and Development*, *19*(3), 58-75. https://eric.ed.gov/?id=EJ1163880

Saultz, A., White, R. S., Meachin, A., Fusarelli, L. D., & Fusarelli, B. C. (2017) Teacher quality, distribution, and equity in ESSA. (2017). *Journal of School Leadership, 21*(5), 652-674. https://doi.org/10.1177/105268461702700503

Schmitt, L. N. T. & Hutchins, S. D. (2015). *Student Learning Objectives (SLOs): Analysis of student growth in 2013-2014, by type and source of Assessment*. Austin Independent School District.

Schön, D. A. (1983). *The reflective practitioner.* Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions.* John Wiley & Sons, Inc.

Scoular, C. (2018). *Equipping teachers with tools to assess and teach general capabilities.* Australian Council for Educational Research Conference 2018. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1343&context=research_conference

Shape America. (2018). *National standards.* https://www.shapeamerica.org/standards/

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22,* 63-75. https://doi.org/10.3233/EFI-2004-2201

Shepard, L. A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education, 31*(2), 165-174. https://doi.org/10.1080/08957347.2017.1408628

Shneyderman, A., & Froman, T. (2015). *Using student growth to evaluate teachers: A comparison of three methods.* Miami-Dade County Public Schools, Research Services. https://eric.ed.gov/?id=ED570129

Shudak, N. J. (2018). Phenomenology. In B. B. Frey (Ed.) *The Sage encyclopedia of educational research, measurement, and evaluation.* Sage. https://doi.org/10.4135/9781506326139

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14. https://doi.org/10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Slotnik, W. J., Bugler, D., & Liang, G. (2014). *Real progress in Maryland: Student learning objectives and teacher and principal evaluation.* (A report from the Mid-Atlantic Comprehensive Center at WestEd.) WestEd and Boston, MA: Community Training and Assistance Center. https://eric.ed.gov/?id=ED559701

Slotnik, W. J., Bugler, D., & Liang, G. (2015). *Change in practice in Maryland: Student learning objectives and teacher and principal evaluation.* Mid-Atlantic Comprehensive Center.

Slotnik, W. J., & Smith, M. D., with 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.

Smit, R., Bachmann, P., Blum, V., Birri, T., & Hess, K. (2017). Effects of a rubric for mathematical reasoning on teaching and learning in primary school. *Instructional Science, 45*(5), 603-622. https://doi.org/10.1007/s11251-017-9416-2

Smith, T. W., & Colby, S. A. (2007). Teaching for deep learning. *The Clearing House: A Journal of Educational Strategies, 80*(5), 205-210. https://doi.org/10.3200/TCHS.80.5.205-210

Smith, W. C. & Holloway, J. (2020) School testing culture and teacher satisfaction. *Educational Assessment, Evaluation and Accountability, 32,* 461–479. https://doi.org/10.1007/s11092-020-09342-8

Smith, J., & Their, M. (2017). Challenges to Common Core State Standards Implementation: Views form six states. *NASSP Bulletin, 101*(3), 169-187. https://doi.org/10.1177/0192636517712450

Soobard, R., & Rannikmae, M. (2015). Examining curriculum related progress using a context-based test instrument: A comparison of Estonian grade 10 and 11 students. *Science Education International, 26*(3), 263-283. https://files.eric.ed.gov/fulltext/EJ1074874.pdf

Stewart, M. (2012). Joined up thinking? Evaluating the use of concept-mapping to develop complex system learning. *Assessment & Evaluation in Higher Education, 37*(3), 349-368. https://doi.org/10.1080/02602938.2010.534764

Stosich, E. L. (2016). Joint inquiry: Teachers' collective learning about Common Core in high-poverty urban schools. *American Educational Research Journal, 53*(6), 1698-1731. https://doi.org/10.3102/0002831216675403

Sullivan, J. P. (2012). A collaborative effort: Peer review and the history of teacher evaluations in Montgomery County, Maryland. *Harvard Educational Review, 82*(1), 142-152. https://doi.org/10.17763/haer.82.1.dj01085408250244

Supovitz, J., Fink, R., & Newman, B. (2016). From the inside in: Common Core knowledge and communication within schools. *AREA Open, 3*(3), 1-18. https://doi.org/10.1177/2332858416653757

Supovitz, J., & McGuinn, P. (2017). Interest group activity in the context of Common Core implementation. *Educational Policy*, 1-33. https://doi.org/10.1177/0895904817719516

Swars, S. L. & Chestnut, C. (2016). Transitioning to the Common Core State Standards for mathematics: A mixed methods study of elementary teachers' experiences and perspectives. *School Science and Mathematics*, *116*(4), 212-224. https://doi.org/10.1111/ssm.12171

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? Evidence of systemic growth in the effectiveness of midcareer teachers. *Education Next*, *12*(4), 78–84. https://www.educationnext.org/can-teacher-evaluation-improve-teaching/

Thissen, D. (2015). Growth through levels. *Measurement*, *13*(3/4), 128-131. https://doi.org/10.1080/15366367.2015.1055146

Thompson, J., Lyons, S., Marion, S.F., Pace, L., & Williams, M. (2016). *Ensuring and Evaluating Assessment Quality for Innovative Assessment and Accountability Systems*. KnowledgeWorks and The National Center for the Improvement of Educational Assessment.

Timar, T. & Carter, A. (2017). *Surprising strengths and substantial needs: Rural district implementation of Common Core State Standards.* Policy Analysis for California Education. https://eric.ed.gov/?id=ED574815

Toma, J. D. (2011). Approaching rigor in applied qualitative research in C. F. Conrad & R. C. Serlin (Eds.) *The SAGE handbook for research in education: Pursuing ideas as the keystone of exemplary inquiry* (2nd ed., pp. 405=423). SAGE.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851. https://doi.org/10.1177/1077800410383121

Tripamer, A. J., Reeves, A. G., & Meinz, E. J. (2014). Teacher perceptions of teacher evaluations in the Fort Zumwalt School District. *NCPEA Education Leadership Review of Doctoral Research, 1*(1), 58-74. https://eric.ed.gov/?id=ED555087

Tuma, A. P., Hamilton, L. S., & Tsai, T. (2018). *A nationwide look at teacher perceptions of feedback and evaluation systems: Findings from the American Teacher Panel.* Rand Corporation. https://www.rand.org/pubs/research_reports/RR2558.html

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

United States Department of Education. (2012a). ESEA flexibility policy document. U.S. Department of Education. http://www2.ed.gov/policy/eseaflex/approved-requests/flexrequest.doc

United States Department of Education. (2012b, February 10). ESEA flexibility review guidance. U.S. Department of Education. www.ed.gov/sites/default/files/review-guidance.doc

United States Department of Education. (2017). Every Student Succeeds Act (ESSA). https://www.ed.gov/essa?src=ft

United States Department of Education. (2009). *Race to the top program executive summary.*

Urick, A., Wilson, A. S. P., Ford, T. G., Frick, W. C., & Wronowski, M. L. (2018). Testing a framework of math progress indicators for ESSA: How opportunity to learn and instructional leadership matter. *Educational Administration Quarterly, 54*(3), 396-438. https://doi.org/10.1177/0013161X18761343

Van den Heuvel-Panhuizen, M., Sangari, A. A., & Veldhuis, M. (2021). Teachers' use of descriptive assessment in primary school mathematics education in Iran. *Education Sciences, 11* (100), 1-23. https://doi.org/10:3390/educsci11030100

VanTassel-Baska, J., & Johnsen, S. K. (2016). From the classroom: Implementing the Common Core in English language arts and in mathematics: Practitioners' Perspectives. *Gifted Child Today, 39*(1), 51-62. https://doi.org/10.1177/1076217515597271

Walsh, K., Joseph, N., Lakis, K., & Lubell, S. (2017). *Running in place: How new teacher evaluations fail to live up to promises.* National Council on Teacher Quality (NCTQ). http://www.nctq.org/dmsView/Final_Evaluation_Pap

Watson, J.G., Kraemer, S.B., & Thorn, C.A. (2009). *The Other 69 Percent.* Center for Educator Compensation Reform. U.S. Department of Education, Office of Elementary and Secondary Education.

Weisberg, D., Sexton, S, Mulhern, J., & Keeling, D. (2009). *The Widget Effect*. The New Teacher Project. https://files.eric.ed.gov/fulltext/ED515656.pdf

Wells, C. (2015). The structure of observed learning outcomes (SOLO) taxonomy model: How effective is it? *Journal of Initial Teacher Inquiry, 1*, 37-39. http://dx.doi.org/10.26021/845

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice, 37*(1), 5-20. https://doi.org/10.1111/emip.12188

Wilson, P.H., & Downs, H. A. (2014). Supporting mathematics teachers in the Common Core implementation. *AASA Journal of Scholarship & Practice, 11*(1), 38-47. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.674.7480&rep=rep1&type=pdf#page=38

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10,* 1-14. https://doi.org/10.3389/fpsyg.2019.03087

Wong, K. K., & Reilly, M. (2014). *Education waivers as reform leverage in the Obama administration: State implementation of ESEA flexibility waiver request.* Paper presented to Annual Meeting of the American Political Science Association, August 28-31, 2014.

Worthington, M. (2010). *Differences between phenomenological research and a basic qualitative research design.* http://a1149861.sites.myregisteredsite.com/DifferencesBetweenPhenomenologicalResearchAndBas icQualitativeResearchDesign.pdf

Woulfin, S. L., Donaldson, M. L., & Gonzales, R. (2016). District leaders' framing of educator evaluation policy. *Educational Administration Quarterly*, *52*(1), 110-143. https://doi.org/10.1177/0013161X15616661

Woulfin, S. L., & Rigby, J. G. (2017). Coaching for coherence: How instructional coaches lead change in the evaluation era. *Educational Researcher, 46*(6), 323-328. https://doi.org/10.3102/0013189X17725525

Xu, X., Grant, L. W., & Ward, T. J. (2016). Validation of a statewide teacher evaluation system: Relationship between scores from evaluation and student achievement

progress. *NASSP Bulletin, 100*(4), 203-222. https://doi.org/10.1177/0192636516683247

Young, S., Range, B. G., Hvidston, D., & Mette, I. M. (2015). Teacher evaluation reform: Principals' beliefs about newly adopted teacher evaluation systems. *Planning and Changing, 46*(1/2), 158-174. https://eric.ed.gov/?id=EJ1145477

Zefran, M., Weber, G., & Fipaza, J. (2015). *Intensifying implementation support: An interactive guide to successful teacher evaluation for professional growth.* Center on Great Teachers & Leaders at American Institutes for Research. https://gtlcenter.org/sites/default/files/NEA_Guide_Evaluation_Implementation.PDF

Zenouzagh, Z. M. (2019). The effect of online summative and formative teacher assessment on teacher competences. *Asia Pacific Education Review, 20*, 343-359. https://doi.org/10.1007/s12564-018-9566-1

Zhang, S. (2014). New teachers' implementation of the Common Core State Standards. *Action in Teacher Education, 36*(5-6), 465-479. https://doi.org/10.1080/01626620.2014.977745

Zwozdiak-Myers, P. (2018). *The teacher's reflective practice handbook: Becoming an extended professional through capturing evidence-informed practice.* London, England: Routledge. https://doi.org/10.4324/9780203118733

Appendix A: Questionnaire for Educators

The purpose of this questionnaire is to outline the goals and structure for the Student Learning Objective (SLO).

1. Email address
2. How many years have you been in education?
3. Please describe your school setting:

    - Rural
    - Suburban
    - Urban

4. Please provide the learning standard(s) of your SLO. (SQ1)
5. For what grade(s) is this SLO being applied? (Check all that apply)

    - Kindergarten
    - 1st Grade
    - 2nd Grade
    - 3rd Grade
    - 4th Grade
    - 5th Grade
    - 6th Grade
    - 7th Grade
    - 8th Grade

6. How will baseline data be gathered? (SQ2)
7. With whom do you discuss your SLO process and rubric? (Check all that apply.) (RQ1, SQ3)

    - No one
    - My evaluator
    - My grade level team
    - My subject area team
    - Special education teacher or paraprofessional
    - English language learner teacher or paraprofessional

- Other: _____

8. Please describe “student growth” as it applies to an SLO. (RQ1, RQ2, SQ2)
9. Please describe your current beliefs about implementing the Common Core State Standards for Mathematics. (RQ1, RQ2, SQ1)
10. Please describe the current structure of the mathematics classroom. How do you decide what to teach, when, and to whom? (RQ1, RQ2)
11. What do you hope to gain from this SLO experience? (Check all that apply.) (RQ1, RQ2, SQ1, SQ2, SQ3)
    - Increase knowledge of standards
    - Increase knowledge of my students
    - Increase knowledge of assessment practices
    - Increase knowledge of instructional practices
    - Other: _____

A copy of your responses will be emailed to the address you provided.

Appendix B: Teacher Interview Guide

Introduction: Thank you for agreeing to participate in this study. The purpose of the study is to explore your perceptions of using standards-based rubrics to monitor student growth and how the rubrics support teachers' reflective practice and Pedagogical Content Knowledge. I appreciate you sharing your SLO plan. The purpose of this interview is to follow-up on your experience monitoring student growth for your mathematics SLO.

Informed Consent: I will be recording this interview so that I can create a transcript. I will share the transcript with you after it is created. If you have any corrections or additions, you can share them with me by email or by phone. I am asking that you sign the consent form to document that you give permission to record the interview and use your comments in this study.

Questions:

1. Tell me a bit about your background as a teacher.
    a. Possible follow-up question: How long have you been teaching?
    b. Possible follow-up question: How long have you been at this school?
    c. Possible follow-up question: What other grade levels have you taught?
    d. Possible follow-up question: How has your teaching been influenced by your experience as a mathematics learner?
2. Please describe the training you received for implementing the Common Core State Standards for mathematics (CCSS-M)? (RQ2, SQ1)
    a. Possible follow-up question: How has your implementation changed since using the standards-based rubric?

3. Tell me about how you used the standards-based rubric to begin implementation of your SLO. (RQ1, RQ2, SQ2, SQ3)
    a. Possible follow up question: How did the rubric influence how you gathered baseline data? (Refer to questionnaire response for question 6).
    b. Possible follow-up question: How did the rubric influence how you used the baseline data?
4. What role did the rubric play in your reflection before, during, and after assessment and instruction? (Refer to questionnaire response for questions 9 and 10.) (RQ1, SQ1, SQ2, SQ3)
    a. Possible follow-up question: How did the rubric influence your assessment planning?
    b. Possible follow-up question: How did the rubric influence your instructional planning?
    c. Possible follow-up question: If you used an SLO process before this, how did using the rubric-based process compare to your prior experience?
5. What role did the rubric play in sharing student growth data with your evaluator as part of your teacher evaluation? (Refer to questionnaire response for questions 7 and 8.) (RQ1, SQ2)
6. What did you learn about your students' abilities, progress, and learning tactics during the process of monitoring student growth with standards-based rubrics? (Refer to questionnaire response for question 10.) (RQ1, RQ2, SQ1, SQ2, SQ3)

7. How did the rubric influence the feedback you provided to students? (RQ1, RQ2, SQ1, SQ2)
    a. Follow-up question: Can you share examples of how any of your students used this feedback?
8. If you discussed your student assessment results with colleagues, can you describe how you and your colleagues used your collective information? (Refer to questionnaire response for question 7.) (RQ1, RQ2, SQ1, SQ2, SQ3)
9. What did you learn about the targeted mathematics standards during this process? (Refer to questionnaire response for questions 4 and 10.) (RQ1, RQ2, SQ1)
    a. Possible follow up question: What did you learn about implementing Common Core State Standards for Mathematics during this process? (Refer to questionnaire response for question 8.)
10. What changes would you make for the next time you teach this topic? (RQ1, RQ2, SQ1, SQ3)
    a. Possible follow up question: How has this experience influenced how you will plan for instruction of other mathematics topics in the future?
11. How do you think any of your assessment or instructional techniques have changed as a result of this process? (Refer to questionnaire response for questions 9 & 10.) (RQ1, RQ2, SQ2, SQ3)
12. Is there anything else you would like to share about your experience using rubrics to monitor student growth? (RQ1, RQ2)

Conclusion: Thank you so much for sharing your experience and reflections with me. I will be sharing the transcript of this interview with you once it is created. Once you receive it, you can contact me by email or phone to share any revisions or additional thoughts about your experience.

Grade 3 Fractions Standards-Based Rubric

Standards	Level 1 *Minimal* Student demonstrates a lack of understanding of prerequisite content for this unit.	Level 2 *Partial* Student demonstrates understanding of prerequisite content for this unit.	Level 3 *Moderate* Student demonstrates understanding of the simple grade level expectations.	Level 4 *Strong* Student demonstrates understanding of the complex grade level expectations.	Level 5 *Distinguished* Student demonstrates understanding of content that goes beyond graded level expectations.
3.NF.1 Understand a fraction 1/b as the quantity formed by 1 part when a whole is partitioned into b equal parts; understand a fraction a/b as the quantity formed by a parts of size 1/b.	Students cannot yet partition circles and rectangles into two, three, or four equal shares, describe the shares using the words *halves, thirds, half of, a third of,* etc., and describe the whole as two halves, three thirds, four fourths. Students cannot yet recognize that equal shares of identical wholes need not have the same shape.	Students can partition circles and rectangles into two, three, or four equal shares, describe the shares using the words *halves, thirds, half of, a third of,* etc., and describe the whole as two halves, three thirds, four fourths. Students recognize that equal shares of identical wholes need not have the same shape.	Students can label a fractional part of a shape.	Students can critique and explain the value of a fractional part using unit fractions.	Students can use addition, subtraction, or multiplication expressions to represent and justify the value of a fraction. Students can label a fractional part of a set.
3.NF.2 Understand fractions as numbers on a number line. Represent fractions on a number line diagram. a. Represent a fraction 1/*b* on a number line diagram by defining the interval from 0 to 1 as the whole and partitioning it into *b* equal parts. Recognize that each part has size 1/*b* and that the endpoint of the part based at 0 locates the number 1/*b* on the number line. b. Represent a fraction *a*/*b* on a number line diagram by marking off *a* lengths 1/*b* from 0. Recognize that the resulting interval has size *a*/*b* and that its endpoint locates the number *a*/*b* on the number line.			Students can represent a fraction 1/*b* on a number line diagram by defining the interval from 0 to 1 as the whole and partitioning it into *b* equal parts. Students can represent a fraction a/*b* on a number line diagram by defining the interval from 0 to 1 as the whole and partitioning it into *b* equal parts.	Students can explain the position of a unit fraction as a point on a number line diagram. Students show that each part has a size 1/*b* and explain the relationship to zero and 1. Students can explain the position of a unit fraction as a point on a number line diagram. Students show that each part has a size a/*b* and explain the relationship to zero and 1.	Students can partition a number line to locate fractions on the number line diagram and show that the location can be represented by a repeated addition or multiplication expression. *(For example, ¾ = 3 x ¼ or ¾ = ¼ + ¼ + ¼)*

Grade 3 Fractions Standards-Based Rubric

3.NF.3 Explain equivalence of fractions in special cases and compare fractions by reasoning about their size. d. Compare two fractions with the same numerator or the same denominator by reasoning about their size. Recognize that comparisons are valid only when the two fractions refer to the same whole. Record the results of comparison with the symbols >, =, or <, and justify the conclusions, e.g., by using a visual fraction model.	Students cannot yet partition circles and rectangles into two, three, or four equal shares, describe the shares using the words *halves, thirds, half of, a third of,* etc., and describe the whole as two halves, three thirds, four fourths. Students cannot yet recognize that equal shares of identical wholes need not have the same shape.	Students can partition circles and rectangles into two, three, or four equal shares, describe the shares using the words *halves, thirds, half of, a third of,* etc., and describe the whole as two halves, three thirds, four fourths. Recognize that equal shares of identical wholes need not have the same shape.	Students can record <, > or = symbols to compare two fractions with the same numerator or the same denominator.	Students can compare two fractions with the same numerator or the same denominator by reasoning about their size. Students can justify or critique the validity of comparisons based on whether the two fractions refer to the same whole. Students can justify the comparisons by using a visual fraction model.	Students can compare two fractions with different numerators and different denominators, e.g., by creating common denominators or numerators, or by comparing to a benchmark fraction such as ½. Students record the results of comparisons with symbols >, =, or <, and justify the conclusions by using a visual fraction model, written, or verbal explanation.
3.G.2 Partition shapes into parts with equal areas. Express the area of each part as a unit fraction of the whole.			Students can partition shapes into parts (halves, thirds, fourths, sixths, and eighths) with equal areas and express the area of each part as a unit fraction of the whole.	Students can critique and justify fractional labels (halves, thirds, fourths, sixths, and eighths) that result from partitioning of shapes into equal areas.	Students can critique and justify fractional labels beyond (halves, thirds, fourths, sixths, and eighths) that result from partitioning of shapes into equal areas. This may include mixed number values.
3.MD.4 Generate measurement data by measuring lengths using rulers marked with halves and fourths of an inch. Show the data by making a line plot, where the horizontal scale is marked off in appropriate units—whole numbers, halves, or quarters.	Students cannot yet generate measurement data by measuring lengths of several objects to the nearest whole unit, or by making repeated measurements of the same object. Students cannot yet make a line plot, where the horizontal scale is marked off in whole-number units.	Students can generate measurement data by measuring lengths of several objects to the nearest whole unit, or by making repeated measurements of the same object. Students can make a line plot, where the horizontal scale is marked off in whole-number units.	Students can measure objects to the nearest half or fourth of an inch. Students can make a line plot for measurement data, where the horizontal scale is marked off in whole numbers, halves, or quarters.	Students can critique and justify measurement to the nearest half and/or quarter inch. Students can critique or justify line plot representations of measurement data to the nearest half and quarter inch.	Students can measure and critique and justify measurements of objects to the nearest eighth of an inch. Students can solve addition or subtraction problems using measurement data (to the nearest half, quarter, or eighth of an inch) presented in line plots.

Grade 3 Fractions Standards-Based Rubric

3.NF.3 Explain equivalence of fractions in special cases and compare fractions by reasoning about their size. a. Understand two fractions as equivalent (equal) if they are the same size, or the same point on a number line. b. Recognize and generate simple equivalent fractions, e.g., ½ = 2/4, 4/6 = 2/3. Explain why the fractions are equivalent, e.g., by using a visual fraction model.	Students cannot yet partition circles and rectangles into two, three, or four equal shares, describe the shares using the words *halves, thirds, half of, a third of,* etc., and describe the whole as two halves, three thirds, four fourths. Students cannot yet recognize that equal shares of identical wholes need not have the same shape.	Students can partition circles and rectangles into two, three, or four equal shares, describe the shares using the words *halves, thirds, half of, a third of,* etc., and describe the whole as two halves, three thirds, four fourths. Students recognize that equal shares of identical wholes need not have the same shape.	Students can identify and model equivalent fractions on a number line diagram.	Students can critique and explain equivalence of fractions using a number line diagram.	Students can critique and explain equivalence of fractions using an addition or multiplication equation, For example, ½ = ¼ + ¼.
3.NF.3 Explain equivalence of fractions in special cases and compare fractions by reasoning about their size. c. Express whole numbers as fractions, and recognize fractions that are equivalent to whole numbers. *Examples: Express 3 in the form 3 = 3/1; recognize that 6/1 = 6; locate 4/4 and 1 at the same point of a number line diagram.*			Students can express whole numbers as fractions and recognize fractions equivalent to whole numbers.	Students can critique and justify equivalent representations of fractions and whole number values.	Students can express fractions as mixed numbers and mixed numbers as fractions. Students can critique and justify equivalent representations of fractions and mixed numbers.

Appendix D: Invitation to Participate

Dear colleague,

As you are aware, the PERA legislation requires that districts incorporate student growth data into the teacher evaluation system. Typically, districts comply with this legislation using two methods. The first method is to use standardized testing data in a statistical model, such as Value-Added Modeling (VAM) or student growth percentiles (SGP). These studies have found that statistical models do not promote teacher growth. The second method is a goal-based approach, such as student learning objectives (SLO). Although researchers have noted the potential for SLO in supporting teacher growth, studies of SLO have produced inconsistent interpretations of what constitutes an SLO. Several of these studies have identified the need for structure to support implementation.

Therefore, I am conducting a research study to examine the teacher perceptions of using standards-based rubrics as the structure for SLO. The purpose of this study is to learn about how the rubrics may influence teachers' reflective practice and pedagogical content knowledge. In order to study this model, I am seeking partner districts who use an SLO process to monitor student growth using standards-based rubrics for mathematics in K-8 classroom settings.

Once their formal consent is obtained, willing teacher participants will be asked to complete an initial survey to gather demographic and baseline data. Teachers will then be invited to participate in one-on-one interviews about the rubric based SLO process.