

Evaluating the Impact of an Assessment Course on Preservice Teachers' Classroom Assessment Literacy and Self-Efficacy

Lance Kruse
North Carolina State University

Whitney Impellizeri
Drexel University

Claire E. Witherel
University of Pennsylvania

Toni A. Sondergeld
Drexel University

While educational policy has resulted in many teachers being responsible for developing and interpreting student growth measures, they generally lack adequate assessment literacy and feel ill-prepared to manage such tasks. This quasi-experimental study explores the impact of an applied assessment course on preservice teachers' assessment literacy and self-efficacy, while also examining the intersection between these variables. Results indicate a significant increase in assessment literacy and self-efficacy after completing the assessment course. Assessment literacy and self-efficacy held a significantly negative relationship at pre-test and a significantly positive relationship at post-test, suggesting meta-ignorance existed regarding classroom assessment skills among preservice teachers.

Introduction

Assessment of student growth and learning has been a key component of federal, state, and local educational policy initiatives in the United States for decades. The role of student assessment in educational policy has shifted over time and now has become a prominent metric in teacher evaluation systems across the U.S.A. (Garret & Steinberg, 2015). Many classroom teachers have subsequently become more responsible for the development, implementation, and interpretation of these student growth measures (Harris, Ingle, & Rutledge, 2014). Simultaneously, teachers report feeling ill-prepared to manage these types of high-stakes student growth assessments (Beziat & Coleman, 2015) that are often tied to their own evaluation and merit pay (Balch & Springer, 2015) without appropriate assessment literacy training. A significant body of research has demonstrated that both preservice and inservice teachers lack adequate assessment literacy skills (e.g., Coombs, DeLuca, LaPointe-McEwan, & Chalas, 2018; DeLuca & Klinger, 2012; Maclellan, 2004; Mertler, 2009; Mertler & Campbell, 2005; Odo, 2016; Volante & Beckett, 2011; Volante & Fazio, 2007). Further, many have reported that teachers do not feel confident in terms of their classroom assessment practices (e.g., Koloï-Keaikitse, 2016; Dial, 2015; DeLuca et al., 2013; Odo, 2016). Previous research has, however, suggested that both teacher assessment literacy, and of significant importance, self-efficacy, are improved when assessment workshops

and/or interventions are offered (e.g., Koh et al., 2017; Stiggins, 2014; DeLuca & Johnson, 2017). While a considerable amount of empirical research has been conducted on the above-mentioned topics, no study to date has investigated the intersection of teacher (preservice or inservice) assessment literacy skills with their classroom assessment confidence (or self-efficacy) as a result of assessment literacy training.

In this study, we explored the impact of a best-practices assessment course on the classroom assessment literacy and self-efficacy of preservice teachers. Additionally, we investigated the general relationship between classroom assessment literacy and classroom assessment self-efficacy among preservice teachers. The following three research questions were specifically investigated:

1. Is there a statistically significant difference in preservice teachers' assessment literacy after participating in a 17-week applied assessment course?
2. Is there a statistically significant difference in preservice teachers' classroom assessment self-efficacy after participating in a 17-week applied assessment course?
3. Is there a statistically significant relationship between attendance in a classroom assessment course and either classroom assessment literacy or classroom assessment self-efficacy among preservice teachers?

Literature Review

Importance of Assessment in U.S.A. K-12 Educational Policy

Historically, assessment has played a varied role in the U.S.A. K-12 educational landscape. Federal education policy has largely directed the shift in importance and magnitude of state standardized and classroom assessment practices. As a result of a decline in U.S.A. educational performance over time, both domestically and when compared to other countries, the landmark federal policy report *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education [NCEE], 1983) highlighted the need to increase classroom rigor and expectations; develop stronger academic standards; and improve teacher preparation programs (NCEE, 1983). Along with a push for increased educational standards and expectations, a conversation about evaluating the impact of educational reform through assessment of student learning was begun. This educational framework persisted until the passage of the *Goals 2000: Educate America Act* in 1994. Assessment of student learning was pushed to the forefront of U.S.A. educational policy as *Goals 2000* mandated states adopt a "...set of high-quality, yearly student assessments, including assessments in at least mathematics and reading or language arts..." (Senate Bill, 142 Congress, 1994, p. 7). To add "teeth" to the educational policy and growing culture of student assessment as a method of evaluating schools, the *No Child Left Behind Act (NCLB)* was introduced in 2002 and required all students to meet or exceed state standards in reading and math by the year 2014, or schools could face sanctions. Under *NCLB* states were mandated to maintain academic standards, develop state-wide assessment systems, and consistently make adequate yearly progress or their schools could lose federal funding (Senate Bill, 107 Congress, 2002).

Passage of the *Every Student Succeeds Act (ESSA)* in 2015 replaced *NCLB* as a result of federal government changes and the states' inability to meet or maintain student proficiency in reading and math by 2014. *ESSA* mandated that all students be taught to the highest academic standards in order to promote societal advancement, and required statewide assessments be developed and implemented to measure students' growth through on-going and varied assessments (U.S.A. Department of Education, 2018). Since then, states have systematically developed methods for measuring student academic performance (both achievement/proficiency and growth), and many states have further connected student growth measures to an evaluation of teacher effectiveness (U.S.A. Department of Education, 2018). As of 2017, 39 states, including Washington D.C., were using student learning growth measures as a direct indicator of teacher effectiveness, while only 12 states were not (National Council on Teacher Quality [NCTQ], 2018). Without a doubt assessment plays a defining role, not only in measuring student learning in both daily and high-stakes tasks (Suskie, 2018), but also in evaluating teacher effectiveness (Goldhaber, Lavery, & Theobald, 2015) and determining the impact of curriculum selections and student services within a school district on student growth (Bassok, Latham, & Rorem, 2016; Jimerson et al., 2016). Unfortunately, much of the work associated with developing assessments of student learning for non-state-tested subject areas to be used for high-stakes reasons (i.e., teacher evaluation systems) falls directly on an unprepared classroom teacher population (Prince et al., 2009).

States and school districts are attempting to navigate the demonstration of teacher effectiveness through student academic growth with a variety of methods. Within state standardized test subject areas, such as reading and mathematics in grades 4-8, this task is easier because student growth from year to year on these assessments is generally used as a proxy for teacher effectiveness (Prince et al., 2009). However, this method only accounts for 31% of teachers across the U.S.A. Teachers who do not teach in a state tested area, often including science and social studies, other high school content areas, physical education, health, and world languages, have been referred to as the "other 69 percent" (Prince et al., 2009). In order for these teachers to demonstrate evidence of effectiveness through measures of student growth, they are tasked with either developing their own pre- and post-assessments, with little evidence of quality, or selecting a vendor assessment which may or may not be well aligned with their curriculum. Consequences for teachers whose students do not demonstrate adequate performance include incurring loss in performance pay and possibly losing their jobs (Balch & Springer, 2015). This added pressure placed on teachers to demonstrate effectiveness, also creates greater stress for utilizing the most appropriate assessments particularly in high-stakes testing situations.

Teacher Assessment Literacy

If teachers are to be held responsible for assessing student learning through high- or low-stake situations, then they must possess assessment literacy skills and the confidence in their ability to develop, implement, and interpret assessments and their corresponding results. In 1990, collaborative efforts by the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) led to the identification of a set of standards for teacher competence in student assessment: *The Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA,

1990). According to AFT et. al. (1990), the following seven Standards outline key assessment areas in which teachers should be knowledgeable:

1. Choosing assessment methods;
2. Developing assessment methods;
3. Administering, scoring, and interpreting assessment results;
4. Using assessment results for decision making;
5. Developing grading procedures;
6. Communicating assessment results; and
7. Recognizing unethical assessment practices.

Taken together, these Standards can be seen as the essentials of assessment literacy. Assessment literacy refers to "...an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (Popham, 2011, p. 267). Mertler and Campbell (2005) further explain that an assessment literate educator is able to:

understand which assessment methods to use to gather dependable information and student achievement; communicate assessment results effectively, whether using report card grades, test scores, portfolios, or conferences; and can use assessment to maximize student motivation and learning by involving students as full partners in assessment, record keeping, and communication. (p.6)

Decades after the joint commission developed their assessment Standards for teachers, research continues to suggest that both preservice and inservice teachers consistently struggle with understanding and implementing appropriate assessment practices in their classroom (e.g., Coombs, DeLuca, LaPointe-McEwan, & Chalas, 2018; DeLuca & Klinger, 2012; Maclellan, 2004; Mertler, 2009; Mertler & Campbell, 2005; Odo, 2016; Volante & Beckett, 2011; Volante & Fazio, 2007). In general, studies have found that preservice teachers tend to lack understanding of assessment practices (Coombs et al., 2018; DeLuca & Klinger, 2012; Mertler & Campbell, 2005). Similarly, preservice teachers have often expressed "little explicit knowledge of assessment methods" (Maclellan, 2004, p.530). Instead, preservice teachers tend to possess a rather superficial understanding of assessment practices (DeLuca & Lam, 2014; Maclellan, 2004; Mertler, 2009; Mertler & Campbell, 2005), such that they are able to discuss basic assessment principles and methods but lack clear practical applications to classroom instruction (DeLuca & Lam, 2014; Maclellan, 2004; Volante & Fazio, 2007). Some specific assessment topics that have been found to be particularly problematic among preservice and inservice teachers alike are: quality assessment selection (DeLuca & Klinger, 2010; DeLuca & Lam, 2014; Gareis & Grant, 2015; Koh et al., 2017; Maclellan, 2004); issues with test fairness (validity and reliability) (DeLuca & Klinger, 2010; DeLuca & Lam, 2014; Mertler, 2009); standardized testing interpretation (DeLuca & Klinger, 2010; Maclellan, 2004; Mertler, 2009; Reeves & Honig, 2015); assessing higher-order thinking skills (Maclellan, 2004; Mertler, 2009); using assessments in diverse classrooms (DeLuca, Chavez, & Cao, 2013; DeLuca & Klinger, 2010; DeLuca & Lam, 2014); implementing varied types of assessments (DeLuca, Chavez, and Cao, 2013; DeLuca & Klinger, 2010; Mertler, 2009; Volante & Fazio, 2007); ability to provide formative feedback to students (DeLuca, Chavez, and Cao, 2013; Falter Thomas & Sondergeld, 2015; Reeves & Honig, 2015; Volante & Beckett, 2011); developing high-quality classroom

assessments (DeLuca, Chavez, and Cao, 2013; DeLuca & Klinger, 2010; DeLuca & Lam, 2014; Maclellan, 2004; Mertler, 2009; Sondergeld, Rychener, & Koskey, 2015) Volante & Fazio, 2007); creating appropriate learning objectives (Sondergeld, Rychener, & Koskey, 2015); and interpreting assessment data to drive classroom instruction (DeLuca, Chavez, and Cao, 2013; DeLuca & Klinger, 2010; Mertler, 2009; Reeves & Honig, 2015; Sondergeld, Rychener, & Koskey, 2015). Preservice and inservice teachers frequently appear to lack an adequate capacity to “articulate significant connections between assessment intentions, theories, and practices” (DeLuca & Lam, 2014, p. 18).

While both preservice and inservice teachers have been shown to lack assessment literacy skills, a study by Coombs et al. (2018) explored the variability of teachers’ assessment literacy skills across teaching experience. Both preservice teachers and experienced inservice teachers (i.e., more than five years of teaching experience) were administered a scenario-based assessment that was developed to evaluate assessment literacy. Results from the study suggested there was a continuum of assessment literacy that significantly increased with teaching experience. However, great variability existed amongst practicing teachers and therefore suggested a critical need for professional learning opportunities to advance inservice teacher assessment literacy (Coombs et al., 2018). These results were similar to other studies exploring preservice and inservice teachers’ assessment literacy using a similar instrument, *The Assessment Literacy Inventory*. Studies by Plake (1993), Campbell, Murphy, and Holt (2002), and Mertler (2003), all found that both preservice and inservice teachers scored quite poorly. While Mertler (2003) found the differences between preservice and inservice teacher’s assessment literacy to be statistically significant, he concluded that both preservice and inservice teachers did not appear to possess a high, or even adequate, degree of assessment literacy.

Intersection of Teacher Assessment Literacy and Beliefs

Similar to assessment literacy, a substantial body of evidence reveals novice classroom assessment perceptions and beliefs among preservice and inservice teachers in the absence of specific assessment literacy training (DeLuca et al., 2013; Howley et al. 2013; Lee & Son, 2015; Mertler, 2009; Reeves & Honig, 2015). More specifically, studies have suggested that preservice teachers’ lack of critical assessment literacy skills are correlated with deficient perceptions regarding assessment (DeLuca et al., 2013; Maclellan, 2004), but an empirical study of this intersection remains lacking. Interestingly, negative assessment skills perceptions have also been demonstrated in inservice teachers as well. In one statewide survey study, Mertler (1999) (as cited in Mertler, 2009) found that over 85% of inservice teachers who participated believed they were not well prepared to assess their students’ learning. Additionally, in a more recent study by Mertler (2009), teachers reflected on their assessment literacy learning through a two-week professional development workshop with daily journaling and self-reported limited ability related to classroom assessment at the onset of the program. Other qualitative studies have provided similar findings related to preservice and inservice teachers’ lack of assessment-related confidence levels before assessment literacy training (Koloi-Keaikitse, 2016; Dial, 2015; DeLuca et al., 2013; Odo, 2016).

Although many studies generally discuss teacher assessment beliefs, perceptions, and confidence, or they identify themes related to these concepts through qualitative research, we have found no studies to date that have quantitatively investigated the intersection of preservice or inservice teacher assessment literacy and specific assessment-related task self-perceptions or confidence. While self-confidence in general is considered a psychological trait, confidence in one's ability to learn or perform behaviors in a specific domain is referred to as self-efficacy (Tschannen-Moran & Hoy, 2007). Self-efficacy has been long studied broadly with preservice and inservice teachers through research on teacher efficacy, or a teacher's confidence in his or her capacity to support student learning (Bandura, 1986; Hoy, 2000; Pajares, 1996). Findings have overwhelmingly shown that teachers with greater teacher efficacy tend to demonstrate greater levels of planning and organization (Allinder, 1994); are willing to try out new instructional methods (Berman et al., 1977; Guskey, 1988; Stein & Wang, 1988); have higher levels of persistence and resilience in the classroom (Jerald, 2007); do not criticize student errors as much (Ashton & Webb, 1986); and are more supportive of challenging students (Gibson & Dembo, 1984). Preservice and inservice teacher self-efficacy in specific content fields such as math and science have also been studied extensively (e.g., Gerde et al., 2017; Knaggs & Sondergeld, 2015; Schoon & Boone, 1998). In general, findings in specific educational domains have shown that teachers with higher self-efficacy in a content area, have a better attitude regarding teaching that domain and are more effective at actually doing it (Knaggs & Sondergeld, 2015). Certainly, there has been significant research on examining teacher efficacy and specific domain teacher efficacy; however, there remains a dearth of study on teacher efficacy related to classroom assessment.

Teacher Preparation and Assessment Training

With the problem of insufficient teacher assessment literacy and perceptions noted above, practicing teachers have suggested this is largely due to inadequate assessment training during teacher preparation programs (e.g., Dial, 2015; Mertler, 2009; Sondergeld, 2014; Volante & Beckett, 2011; Yamtim & Wongwanich, 2014). Oftentimes, teacher candidates complete their programs with little to no focus on assessment development and evaluation (Dial, 2015; Sondergeld, 2014; Volante & Beckett, 2011). Rather, assessment literacy is indirectly supported or woven into other pedagogical coursework. Or when it is included, a narrow more theoretical scope of assessment (e.g., validity and reliability) is the primary focus over practical application of classroom assessment best practices (Sondergeld, 2014; Yamtim & Wongwanich, 2014).

Acknowledging the misalignment between preservice teacher assessment literacy training and teacher assessment skill requirements, many teacher preparation programs have begun to address this assessment literacy gap. Some programs have designed and implemented questionnaires geared toward understanding student perceptions of assessments, in an attempt to evaluate their assessment literacy, along with their understanding of the purpose and utilization of assessments (e.g., Pereria, Niklasson, & Flores, 2017; Koloï-Keaikitse, 2016; Volante & Fazio, 2007). Other programs have utilized individual and peer tutoring on the theory and application of assessment techniques (Odo, 2016). In addition, some programs have redesigned courses in order to embed assessment content. Such actions demonstrate a stronger focus on promoting assessment literacy to ensure preservice teachers will be better equipped for developing and discerning appropriate

assessments in their future classrooms (Lee & Son, 2015; Mandinach et al., 2015). While it is not the norm, some institutions do provide a stand-alone classroom assessment course designed to specifically teach best practices in classroom assessment at either the graduate level alone or both graduate and undergraduate levels (Sondergeld, 2014). These strategies all aim to provide preservice and inservice teachers with adequate assessment literacy training through university-level undergraduate or graduate-level coursework. Research suggests that such efforts have been successful in increasing preservice and inservice assessment literacy skills when implemented (Wang, Wang, & Huang, 2008; Mertler, 2004).

Context of the Study

Assessment Course

At a large public university in the Midwest, an undergraduate classroom assessment course is required of all middle and high school teacher candidates in their junior year. Preservice teachers have successfully finished nearly all of their content-area courses, educational foundation courses, and multiple domain-specific pedagogical courses prior to entering this 3-credit hour course. This assessment class is intentionally taken during the same semester preservice teachers are preparing for and teaching in a methods field placement. As such, preservice teachers learn about assessment best practices and develop multiple assessments (formative and summative), aligned with their methods teaching content. These assessments are delivered during their methods teaching experience and the teachers are encouraged to reflect on the outcomes during a methods classroom teaching experience. The course runs for 17 weeks, including 11 weeks of lessons delivered at the university (assessment learning/developing), 4 weeks of methods teaching in the field (assessment implementation), and 2 weeks of look-back and wrap-up at the university (assessment interpretation/reflection).

Preservice teachers in this course complete multiple homework assignments focused on the iterative process of developing and refining assessment tools used in their classroom during methods teaching. Homework assignments were completed and aligned with the content preservice teachers would be delivering in methods, and it was subsequently brought to class for peer-review. Preservice teachers were trained by the instructor in the process of peer-reviewing, and given rubrics to use for assessing each other's work. After peers provided feedback, preservice teachers revised their homework and then submitted it to the course instructor for feedback. Instructor feedback was used by preservice teachers to further refine assessment homework and then implemented in one of six Key Assessments (KAs). For example, in a lesson on objective items, preservice teachers were given a homework assignment of creating three multiple-choice items that aligned with specific learning objectives they would be covering during their methods teaching. Preservice teachers were explicitly taught how to assess sample multiple-choice items and aligned learning objectives using a rubric provided that focused on information learned in their assessment course lessons. Peers then assessed each other and gave formative feedback aligned with rubric criteria. Preservice teachers revised their multiple-choice items, submitted these for instructor feedback, revised again, and finally used them as a component of KA1: Traditional Assessment, which was later delivered to K-12 students during their methods teaching in the same semester. Table 1 describes the course Key Assessments.

Table 2 details course content taught throughout the 17 weeks aligned with Key Assessments completed.

Methods

Data Collection and Instrumentation

Two instruments were used in this pre-post quasi-experimental design without control

Table 1.
Assessment Course Key Assessments with Descriptions

Key Assessment	Description
KA1: Traditional Assessment	Preservice teachers developed a written test aligned with instructional objectives and content, and cognitive skills (e.g. Bloom’s Taxonomy). The test created included test and item directions, an answer key and/or rubrics, and incorporated at least 3 of the following question types: true-false (binary choice), matching, multiple-choice, completion, short answer, and essay.
KA2: Performance-based Assessment with Rubric	Preservice teachers developed a student task (e.g. performance, product, paper, presentation) that was assessed with a scoring rubric and aligned with instructional objectives and content, and cognitive skills. The description of the task was presented in a “student handout” that would provide students with information needed to successfully perform the task. A scoring rubric was appropriately aligned with the developed task.
KA3: Formative Assessment	Preservice teachers planned, implemented, and reflected on the delivery of one formative assessment strategy and technique conducted during their methods field placement.
KA4: Value-added Assessment	Preservice teachers interpreted a school district’s “grade card” in terms of student Achievement, Progress, and Gap Closing measures.
KA5: Assessment Implementation & Reflection (AIR) ¹	Preservice teachers wrote a reflection paper on a formal assessment they administered (KA1 or KA2) during their methods field placement and the results from the assessment. This reflection paper included a description of their lesson (learning objectives); student results from the assessment (graphic representing scores, response patterns, strengths, challenges); sample feedback given to students and how it aligns with best practices (actual student work with preservice teacher feedback on it); and reflection on the effectiveness of the assessment (description of how to modify the assessment for future use based on data).
KA6: Standardized Test Results Interpretation	Preservice teachers used a student’s standardized test report to write a letter to a parent, explaining the student’s standardized test results. The letter needed to be written in meaningful, non-technical language, and explain the student’s strengths and weaknesses as indicated by the test results. Educational recommendations specifically aligned with the student’s test results were made.

¹ All teacher candidates at this university were required to successfully complete the edTPA (Pearson, 2018) for graduation purposes. The AIR Key Assessment in this course served as a practice for edTPA’s Task 3. See <https://www.edtpa.com/> for additional information on the edTPA.

group study (Creswell, 2014): the *Assessment Literacy Inventory (ALI)* and the *Classroom Assessment Self-Efficacy Survey (CASES)*. Both measures were administered through an online survey platform (Qualtrics) to participants in pre-post fashion; pre-assessment during the first week of class, and then again in the final exam week (16 weeks later) as a post-assessment. Both assessments were completed anonymously with students reporting the last 4-digits of their phone

Table 2.

Assessment Course Schedule, Content, and Key Assessment Alignment

Weeks <i>Stage of Course</i>	Course Lessons/Learning Experiences	Key Assessments (KA) Developed/Completed
Weeks 1-9 <i>Learning/Developing</i>	Lesson 1: Classroom Decision Making	KA1: Traditional Assessment
	Lesson 2: Learning Goals/Objectives	
	Lesson 3: Taxonomies	
	Lesson 4: Validity & Reliability	KA2: Performance-based Assessment with Rubric
	Lesson 5: Formative Assessment	
	Lesson 6: Giving Quality Feedback	
	Lesson 7: Grading & Reporting Practices	KA3: Formative Assessment
	Lesson 8: Objective Items & Item Analysis	
	Lesson 9: Constructed Response Items	
	Lesson 10: Rubrics	KA4: Value-added Assessment
	Lesson 11: Planning Assessment & Instruction	
	Lesson 12: Educational Accountability	
Weeks 10-14 <i>Implementing</i>	Methods Teaching and Delivering Assessments Developed in Course	
Week 15 <i>Interpreting/Reflecting</i>	Writing Up Student Assessment Results and Interpretation	KA5: Assessment Implementation & Reflection (AIR)
Week 16 <i>Learning/Developing</i>	Lesson 13: Interpreting Standardized Tests	KA6: Standardized Test Results Interpretation
Week 17 <i>Interpreting/Reflecting</i>	Reflecting on Best Practices in Assessment	

number as a code to link pre-post data. Both verbal and written directions from the course instructor were given to participants discussing the importance of completing each assessment honestly and doing their best without looking up correct answers through additional resources. Students were informed their scores were to be used formatively and not towards course grades. Aggregate results were shared with preservice teacher participants who were further asked to view these measures as documentation of their assessment literacy and confidence growth from participating in the course.

Assessment Literacy Inventory. The *ALI* (Campbell & Mertler, 2005) is a 35-item scenario-based objective test. There are five scenarios that feature teachers addressing various assessment-related decisions, and each scenario contains seven multiple-choice items. The multiple-choice items were designed to align with the *Standards for Teacher Competence in Educational Assessment of Students* (AFT et. al., 1990) demonstrating strong content validity evidence.

Therefore, each scenario contains seven items that align with the following concepts: 1) choosing assessments, 2) developing assessments, 3) administering, scoring, and interpreting assessment results, 4) using assessment results for decision-making, 5) developing grading procedures, 6) communicating assessment results, and 7) recognizing inappropriate assessment practices. The *ALI* has undergone various pilot testing and revisions to determine the appropriateness of using the instrument for preservice teachers, contributing to the support of strong internal structure validity evidence across the assessment ($KR20=0.74$) (Mertler & Campbell, 2005). For scoring, participants receive seven subscale scores (5 points possible per subscale) and a total score (35 points possible).

Classroom Assessment Self-Efficacy Survey. *CASES* consists of 24 items and was designed by the authors to measure two aspects of preservice teachers' perceptions of their confidence towards classroom assessment: Evaluating Others Confidence (EOC – 12 items) and Personal Performance Confidence (PPC – 12 items). Both scales focus on 12 common classroom assessment tasks teachers face when designing, implementing, and using assessments in their daily instruction. For example, survey items were based on tasks such as designing high quality learning objectives and typical classroom assessment items (e.g., multiple choice items, essays), implementing formative and summative assessments, and using assessment data to make classroom-level and student-level decisions. First, EOC is measured by asking participants to indicate their classroom assessment confidence level for evaluating a teacher they were observing across those 12 tasks. A four-point Likert-type scale is used (Not Confident, Somewhat Confident, Confident, Extremely Confident) with a higher score indicating a higher level of assessment confidence (or efficacy). Second, PPC is measured by asking participants indicate their level of confidence in performing those same 12 classroom assessment tasks, using the same four-point Likert-type scale. Lastly, an aggregate measure of classroom assessment confidence is evaluated by combining the first and second sections for a Total Assessment Confidence (TAC) measure. EOC and PPC scores range from 12-48, whereas TAC scores range from 24-96. Overall and by subscales, *CASES* indicated strong support for internal structure validity evidence (TAC, $\alpha = 0.894$; EOC, $\alpha = 0.858$; PPC, $\alpha = 0.818$). Content validity evidence for *CASES* is also supported through its item alignment with the *Standards for Teacher Competence in Educational Assessment* (AFT et. al., 1990), *ALI* items (Campbell & Mertler, 2005), and assessment course lessons (see Appendix A).

Sample

A total of 96 preservice teachers completed both the *ALI* and *CASES* at pre- and post- during the last semester of their junior year in an undergraduate education program. All participants were enrolled in either a middle-years (grades 4-9; 69.8%) or secondary-years (grades 7-12; 30.2%) bachelor's degree program offered from a Midwestern public university's College of Education. The majority of participants (66.7%) were studying to become a teacher in an academic "core" area (Language Arts, 17.7%; Math, 17.7%; Science, 6.3%; Social Studies, 25.0%). While the remaining participants (13.6%) who identified their area of study were reported being in an academic "non-core" field (World Languages, 11.5%; Business Education, 2.1%). Approximately 20% of participants did not report their content area of study. In terms of prior assessment instruction, more than three-quarters (78.1%) indicated they had completed prior

courses that discussed assessment in some manner, but nearly all (95.8%) reported they had never taken a course focused specifically on assessment prior to this class. No other demographic data were collected from preservice teachers to protect participant identities.

Data Analysis

For research question one, data from the *ALI* were scored as either correct (1) or incorrect (0), such that a higher score on the *ALI* indicated a higher level of assessment literacy. Missing data were scored as incorrect. Pre- and post-test subscale and total scores were computed. Since the *ALI* was administered before the assessment course content began, and again after the assessment course content was completed, a series of dependent samples *t*-tests were conducted to determine if there was a statistically significant difference in preservice teachers' assessment literacy over time. To evaluate practical significance of sub-scales (individual Standards), logical assessment criteria were applied to average Standard scores. Each Standard assessed by the *ALI* is comprised of 5 items (ranging from 0 to 5 points). Average Standard scores were computed and the following grading criteria were applied: A = 5pts-4pts; B=3.99pts-3pts; C=2.99pts-2pts; D=1.99pts-1pt; F=1pt-0pts.

Research question two was answered with data from *CASES*. Rating scores were coded using the following scheme: Not Confident=1, Somewhat Confident=2, Confident=3, and Extremely Confident=4, such that a higher score on the survey indicate a higher perception of classroom assessment confidence. Subscale and overall classroom assessment confidence scores were computed for pre- and post-surveys. A series of dependent samples *t*-tests were conducted to determine if there was a statistically significant difference in preservice teachers' classroom assessment confidence over time. Effect sizes were computed for all statistical tests for research questions one and two using partial η^2 such that 0.01 is small, 0.06 is medium, and 0.14 is large (Cohen, 1988).

To answer research question three, multiple Pearson Correlations were conducted to determine if there was a statistically significant relationship between preservice teachers' assessment literacy and classroom assessment confidence before the assessment course and after. SPSS version 24 was used for all statistical analysis. Effect sizes were computed for all statistical tests in research question three using r^2 such that 0.01 is a small effect, 0.06 is a medium effect, and 0.14 is a large effect (Cohen, 1988).

Results

RQ1. Change in Assessment Literacy

Overall, there was a significant difference in total *ALI* scores from the pre-post ($p < 0.001$). This increase in overall assessment literacy had a large effect size ($\eta^2 = 0.40$) with 40% of the variance in total *ALI* score accounted for by participation in the assessment course. With the exception of Standard 3, preservice teachers showed a significant amount of growth from pre-post on all Standards ($p < .01$). Standard 1 and Standard 7 showed a significant increase over time and the effect size was considered medium (η^2 ranging from 0.07 – 0.13) with 7%-13% of the variance in

these Standard scores accounted for by participation in the assessment course. Effect sizes for the remaining Standards were considered large (η^2 ranging from 0.18 - 0.30) with 18%-30% of the variance in these Standard scores accounted for by participation in the assessment course. See Table 3 for descriptive and inferential statistics related to these analyses.

Table 3.

Dependent Samples t-test Descriptive and Inferential Results for Assessment Literacy Change over Time (n=96)

Standard (pts possible)	Pre-M(SD)	Post-M(SD)	t-statistic	η^2
1 – Choosing Assessments (5 pts)	2.78 (1.09)	3.28 (1.15)	3.71***	0.13
2 – Developing Assessments (5 pts)	1.93 (0.94)	2.69 (1.04)	5.94***	0.27
3 – Administering, Scoring, and Interpreting Assessment Results (5 pts)	2.94 (1.19)	2.98 (1.21)	0.31	0.01
4 – Using Assessment Results for Decision-Making (5 pts)	2.79 (1.06)	3.50 (1.21)	4.52***	0.18
5 – Developing Grading Procedures (5 pts)	2.35 (0.98)	3.13 (1.14)	6.34***	0.30
6 – Communicating Assessment Results (5 pts)	2.49 (0.97)	3.20 (1.02)	5.32***	0.23
7 – Recognizing Inappropriate Assessment Practices (5 pts)	2.77 (1.09)	3.10 (1.18)	2.74**	0.07
Total ALI Score (35 pts)	18.05 (3.76)	21.88 (4.83)	7.94***	0.40

*= $p < .05$, **= $p < .01$, ***= $p < .001$

All of the Standards (except Standard 3) increased one letter grade from an average D to a C or from an average C to a B. Therefore, preservice teachers demonstrated a practical improvement in their conceptual understanding of assessment literacy across six of the seven Standards after participating in the assessment course. Standard 3 did not increase over time because preservice teachers had a high baseline to begin (almost B average) and maintained that average over time. Figure 1 shows ALI score change in practical terms from pre- to post-assessment.

RQ2. Change in Classroom Assessment Confidence

Overall, there was a significant difference in TAC scores from the pre-post ($p < 0.001$). This increase had a very large effect size ($\eta^2 = 0.89$) with 89% of the variance in TAC score accounted for by the intervention. Preservice teachers showed a larger increase in EOC scores from pre-post than in PPC, although both significantly increased ($p < 0.001$) with large effect sizes (η^2 ranging from 0.80-0.90). These large effect sizes indicated that 90% of the variance in EOC was accounted for by participating in the assessment course, and 80% of the variance in PPC was accounted for by participating in the assessment course. Table 4 provides descriptive and inferential statistics for these analyses.

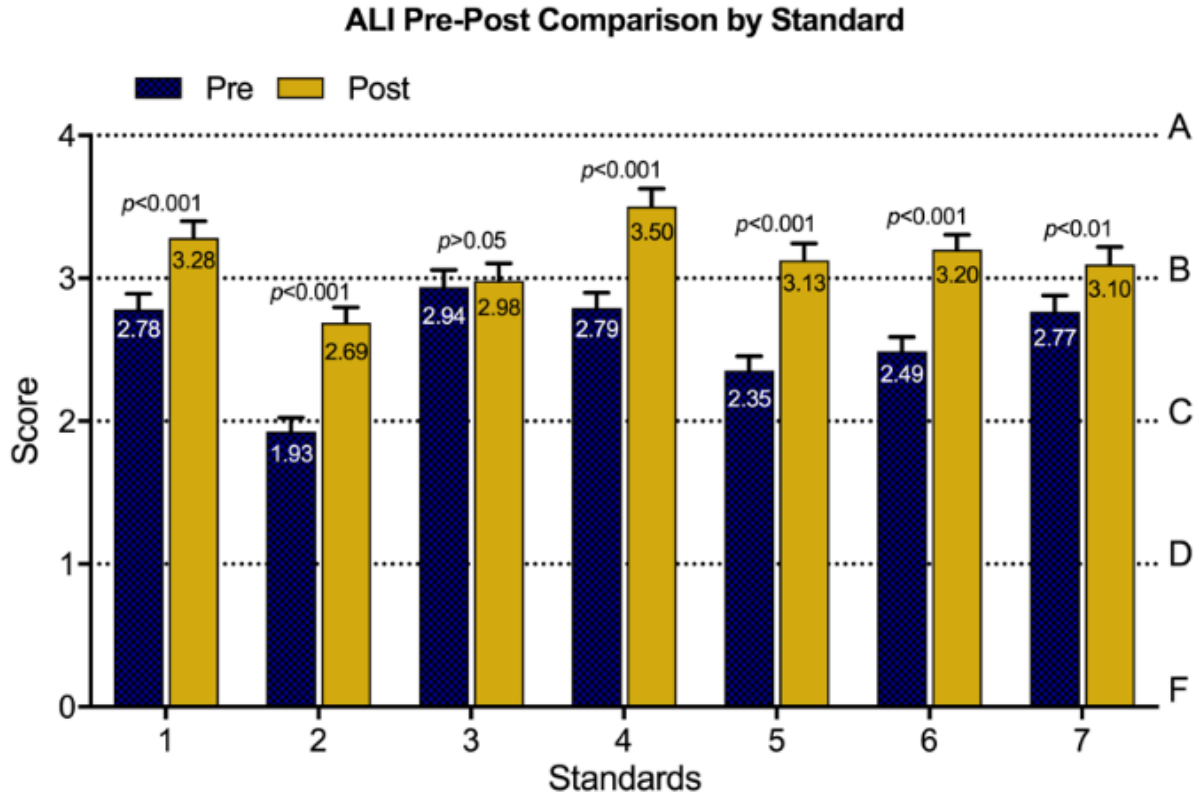


Figure 1. ALI pre-post comparison by Standard. Data are represented as the mean \pm standard error of the mean (SEM) with the mean value provided within the column and *p*-values directly reported. This figure shows average change from pre-post across each Standard including the change in letter grades.

Table 4.

Dependent Samples *t*-test Descriptive and Inferential Results for Confidence Change over Time (*n*=77)

Confidence Scale (pts possible)	Pre-M(SD)	Post-M(SD)	<i>t</i> -statistic	η^2
Evaluating Others Confidence (48)	32.39 (5.35)	44.31 (3.38)	28.84***	0.90
Personal Performance Confidence (48)	30.70 (4.78)	40.77 (4.62)	19.63***	0.80
Total Assessment Confidence (96)	63.09 (9.08)	85.08 (7.32)	27.04***	0.89

*=*p*<.05, **=*p*<.01, ***=*p*<.001

RQ3. Relationship between Assessment Literacy and Classroom Assessment Confidence

Prior to completing the assessment course, assessment literacy and classroom assessment confidence had a significantly negative weak relationship among preservice teachers; $r(75) = -0.247$, *p*<0.05, two-tailed. The effect size is considered medium ($r^2=0.061$) with 6.1% of the variance in preservice teachers' assessment literacy results explaining their classroom assessment confidence. Therefore, prior to the course, the preservice teachers with self-reported higher

assessment confidence demonstrated slightly lower levels of assessment literacy, and those with self-reported lower assessment confidence demonstrated slightly higher levels of assessment literacy.

After the assessment course had been completed, assessment literacy and classroom assessment confidence had a significantly positive weak relationship among preservice teachers; $r(75)=0.386$, $p<0.001$, two-tailed. The effect size is considered large ($r^2=0.1490$) with 14.90% of the variance in preservice teachers' assessment literacy results explaining their classroom assessment confidence. Therefore, preservice teachers with self-reported higher assessment confidence demonstrated higher levels of assessment literacy, and vice versa.

Discussion

Through our study, we sought to explore the impact of a preservice teacher assessment course that focuses specifically on best practices in classroom assessment on preservice teacher classroom assessment literacy and classroom assessment self-efficacy. We also investigated the relationship between classroom assessment literacy and classroom assessment self-efficacy among these preservice teachers. In the following sections, we explain how our results are well-aligned with prior research while providing some new insights.

Differential Assessment Literacy Growth

Findings from our study align with existing research suggesting that preservice teachers lack adequate assessment literacy prior to receiving specific assessment training (e.g., Coombs, DeLuca, LaPointe-McEwan, & Chalas, 2018; DeLuca & Klinger, 2012; Maclellan, 2004; Mertler, 2009; Mertler & Campbell, 2005; Odo, 2016; Volante & Beckett, 2011; Volante & Fazio, 2007). Additionally, the assessment course proved to be instrumental in improving preservice teachers' assessment literacy, which also supports earlier reported findings (e.g., Lee & Son, 2015; Mandinach et al., 2015; Mertler, 2009; Odo, 2016; Wang, Wang, & Huang, 2008). Although, growth was not equal across all seven of the teacher assessment competence Standards tested on the *ALI* as some Standards had considerably less or more room for growth than others.

While all *ALI* subsections had improved scores from pre-post among our preservice teachers, this growth was non-significant ($p>0.05$) for Standard 3 (Administering, Scoring, and Interpreting Scores). The non-significant growth in Standard 3 is due to the relatively higher than average pre-test score of 2.94 (practical grade = B). A plausible explanation for the higher than expected pre-assessment score on this Standard is prior learning of topics covered on this section of the *ALI* from previous courses completed. As mentioned earlier, to enroll in this assessment course, nearly all general content courses and multiple education courses had to be successfully completed. Items on the *ALI* assessing Standard 3 include concepts such as interpreting standardized test results, understanding percentile rank, and standard deviation. All College of Education students at this university were required to complete a certain number of quantitative (mathematics) courses. A popular choice for non-mathematics content area preservice teachers is a course that covers basic statistical concepts which would include topics such as percentile rank and standard deviation in an educational context. Further, those with a mathematics content area

specialization were required to have completed a statistics course from the mathematics department prior to this assessment course, which covers the same basic statistical topics, although not in an educational context. Thus, we suspect this is likely the reason why many preservice teachers were able to demonstrate higher than expected understanding related to this Standard prior to learning about it in the assessment course.

On the other hand, preservice teachers saw the most growth for Standard 2 (Developing Assessments) and Standard 5 (Developing Grading Procedures). Large growth across these two Standards seems reasonable as preservice teachers explicitly learned the skills of designing high-quality assessments and grading procedures from participation in the assessment course. Numerous Key Assessments (KAs) in the course required preservice teachers to design and implement traditional (KA1), performance-based (KA2), and formative assessments (KA3) during their practicum experience, which align with Standard 2 *ALI* items. For each of these KAs, preservice teachers also had to develop appropriate grading procedures, which aligns with Standard 5 *ALI* items. As discussed in other studies (DeLuca, Chavez, & Cao, 2013; DeLuca & Klinger, 2010; DeLuca & Lam, 2014; Maclellan, 2004; Mertler, 2009; Sondergeld, Rychener, & Koskey, 2015; Volante & Fazio, 2007), developing assessments and appropriate grading procedures are very challenging tasks for preservice and inservice teachers alike. Therefore, it stands to reason that the process of engaging in hands-on practical construction of high-quality assessments and grading procedures, along with an iterative feedback looping system to provide opportunities for learning through revision, would result in significant and practical growth in these two assessment literacy Standards.

Many teacher preparation programs do not require a specific classroom assessment course while some only address theoretical aspects of assessment (e.g., reliability and validity) within other educational courses (Dial, 2015; Sondergeld, 2014; Volante & Beckett, 2011; Yamtim & Wongwanich, 2014). Those preservice teachers are not provided the opportunity to engage in a practical hands-on learning experience of constructing and utilizing classroom assessments. As mentioned previously, a majority of the preservice teachers in this study had already addressed some aspect of assessment in previous education courses, and yet they demonstrated low pre-test assessment literacy scores. However, after participating in this specifically-designed assessment course, the preservice teachers demonstrated significantly higher post-test assessment literacy scores. Therefore, there is a need for a specific classroom assessment course, such as this one, to provide students the opportunity to practice and receive feedback on designing and implementing classroom assessments, which may lead to a significant improvement in assessment literacy.

Classroom Assessment Self-Efficacy

Prior to participating in the assessment course, preservice teachers had relatively low levels of confidence in their classroom assessment skills, which supports findings from prior research (e.g., DeLuca et al., 2013; Howley et al. 2013; Lee & Son, 2015; Mertler, 2009; Reeves & Honig 2015). After the assessment course, the preservice teachers reported a significantly higher level of classroom assessment confidence, which suggests the course positively impacted preservice teachers' classroom assessment efficacy. These findings are similar to those from other studies

which have implemented rigorous interventions focusing explicitly on the design and application of classroom assessments (e.g., Deluca et al., 2013; Mertler, 2009; Reeves & Honig, 2015).

Differing from other research on classroom assessment perceptions, our study investigated preservice teachers' confidence in two specific domains of classroom assessment: evaluating others (EOC) and personal performance (PPC). While both EOC and PPC significantly increased as a result of participating in the assessment course, preservice teachers reported higher levels of EOC than PPC at both pre- and post-testing, which suggests they are more confident in evaluating others' work than completing the task themselves. These findings are not surprising as they align with the revised Bloom's taxonomy (Airasian et. al., 2001), which suggests that the act of creating is a more complex and challenging task in comparison to evaluating the work of others. Furthermore, the lower level of PPC at both pre- and post-testing suggests a need for preservice teachers to have more applied experiences creating assessments and actually implementing assessment principles learned through some practical manner (e.g., during methods or student teaching), rather than abstractly learning about them. Our specific assessment course, like other classroom assessment interventions (e.g., DeLuca, 2012; Mertler, 2009; Odo, 2016), focused intentionally on promoting authentic learning experiences for these preservice teachers to improve their assessment skills and self-efficacy. The findings suggest that it is critical to ensure a scaffolded learning environment, with an iterative loop of feedback and revision opportunities, in order to properly allow preservice teachers time to safely engage with classroom assessment content – practice is a key factor in this learning process.

Intersection of Assessment Literacy and Self-Efficacy

With a lack of existing research exploring the intersection of preservice teachers' assessment literacy and their self-efficacy towards classroom assessments, our study sought to investigate this relationship in a pre-post manner. To our surprise, findings from our study demonstrated what is known in psychology as the Dunning-Kruger effect (Kruger & Dunning, 1999), where preservice teachers who were less skilled in classroom assessment literacy (as demonstrated in the pre-*ALI* objective test) had inflated (or overestimated) confidence in their classroom assessment skills (as demonstrated in pre-*CASES* self-reported assessment) prior to classroom assessment training. This phenomenon of "meta-ignorance" (or ignorance of one's own ignorance) (Dunning, 2011) has indeed been noted in many social and intellectual domains such as logical reasoning and grammar skills (Kruger & Dunning, 1999), emotional intelligence (Sheldon, Dunning, & Ames, 1999), and the medical field (Haun, Zerinque, Leach, & Foley, 2000) to mention a few. Although "meta-ignorance" is a prevalent condition, research has shown that helping participants to improve their skills in the domain under study allows them to better recognize their own ability limitations and thus increase their metacognitive skills (Griffin, Jee, & Wiley, 2009; Koriat, 2008; Kruger & Dunning, 1999). Our study demonstrated this same increase in preservice metacognitive skills related to classroom assessment with the shift from a significantly negative relationship between classroom assessment literacy and confidence at pre-testing (prior to the assessment course) to a significantly positive relationship at post-testing (after the assessment course). While not specifically addressed in this study, we believe that these findings imply that our preservice teachers grew in their understanding of the difficulty and complexity related to developing classroom assessments on their own, which aligns with

Mertler's (2009) findings where inservice teachers reported they did not realize how difficult it was to make high-quality assessments prior to participating in an assessment professional development workshop.

Practical Evaluation of Findings

Average overall *ALI* scores from preservice teachers in this study were similar to those found in other research (Mertler, 2003; Mertler & Campbell, 2005; Plake, 1993). However, interpretation of these comparable findings differ depending on the lens through which they are evaluated: traditional/mathematical vs. practical/logical. From a strictly mathematical perspective, Mertler (2009) interpreted overall *ALI* scores of 60% correct (21 out of 35 items) as representing insufficient understanding of assessment practices. Certainly, when viewing results through the lens of a traditional/mathematical percentage-based grading scale, a 60% seems lacking. Instead of using a percentage-based approach, we propose implementation of a more practical/logical approach to evaluating *ALI* findings. Consider that the *ALI* contains 35 items with five items for each of seven Assessment Standard subscales. If a test-taker misses one item on a subscale (or scores four out of five) this mathematically equates to an 80% (a B or C depending on the instructor). And missing two or more items on a subscale (40% or lower) will result in an F on any traditional grading scale. Utilizing a simple percentage-based model for evaluating *ALI* findings results in a restricted range of interpretation slanted towards test-taker deficiency. To allow for a more complete range of results interpretation, we chose to employ a practical/logical assessment scale rather than one that was purely traditional/mathematical. Considering the specifics of the *ALI* assessment, we developed a practical assessment scale where correctly answering four or more items = A, three items = B, two items = C, one item = D, and zero items = F. Multiplying these thresholds by seven results in the following overall grading scale for the *ALI*: 28+ = A, 21-27 = B, 14-20 = C, 7-13 = D, below 7 = F. Therefore, a practical/logical interpretation of a mean score of 21 would represent a B, or above average understanding of assessment concepts. While percentages play a historical role in testing and grading, the design and type of assessment needs to be considered when reviewing and interpreting scores. Thus, the practical assessment scale was reviewed by three subject matter experts with doctorates in assessment and measurement who provided face validity evidence for the appropriateness of the scale with this instrument. As such, we believe that using an adjusted logical interpretation of *ALI* findings over a traditional mathematical one provides a more accurate and practical depiction of preservice teachers' assessment literacy and growth.

Future Study

Given the findings of this study, it should be noted that correctly answering assessment-related items on an objective test or indicating high assessment confidence does not necessarily translate into effective assessment practices implemented in the classroom. This study demonstrated that preservice teachers significantly increased both their classroom assessment literacy and confidence levels as a result of participating in an assessment course that provided extensive supports. We did not, however, investigate the effectiveness of these individuals in terms of applying assessment best practices in actual classrooms. We therefore suggest that future studies explore how practicing teachers design, implement, and use classroom assessments by critically

evaluating (or observing) not only the quality of assessments delivered, but their actual delivery on a more routine basis. Some states, such as Ohio, currently have a Resident Educator Program in place for new teachers that focuses heavily on classroom assessment practices and their use of assessment data (formative and summative) (ODE, 2018). Studying programs like this would likely be beneficial to and expand knowledge in the field of assessment literacy since a proficient teacher should be able to demonstrate best practices in classroom assessment on their own. Additionally, the notion of determining proficiency levels or cut scores through more objective standard setting methods should be investigated further when utilizing tools to evaluate achievement and growth. Arbitrary or traditional mathematically-based approaches used to determine appropriate levels of proficiency or academic growth likely will not lead to decisions that are as well informed as more practical or criterion-based methods. Lastly, this study does not investigate the quality of the course design as compared to traditional assessment courses, but instead highlights the need for some type of specific classroom assessment course to be required in teacher preparation programs. Future studies should explore the impact of a scaffolded learning environment, such as the one described here, as compared to more traditional types of instruction.

Final Thoughts

While educational policy surrounding K-12 student assessment changes over time, it seems unlikely that teachers will be relieved of their duties to design, implement, and have both low- and high-stakes decisions made based on teacher-created classroom assessments. In order to reduce the prevailing gap between teacher classroom assessment skills and educational policy expectations, it is critical that preservice teachers complete rigorously-designed undergraduate assessment courses that intentionally foster assessment literacy and work to deliberately improve preservice teacher classroom assessment self-efficacy. Doing anything less is setting our future teachers up for potential failure and ultimately weakening our K-12 educational system due to an inability to effectively measure and evaluate students' academic learning and growth within the classroom.

Author Notes:

Lance Kruse, PhD is a Research Scholar in STEM Education at North Carolina State University. His research interest includes the intersection of research and measurement within STEM education focusing on assessment development, teacher education, and program evaluation.

Whitney Impellizeri, EdD is a Postdoctoral Research Scholar through the West Philly Promise Neighborhood grant in Drexel University's School of Education.

Claire E. Witherel, PhD is currently a Postdoctoral Research Fellow at the University of Pennsylvania in the Department of Bioengineering. Her research focuses on designing injectable therapeutic biomaterials for the treatment of osteoarthritis.

Toni A. Sondergeld, PhD is an Associate Professor of assessment, research, and statistics in Drexel University's School of Education. Her research focuses on evaluation and assessment of school reform efforts and STEM programs.

Correspondence concerning this article should be addressed to Lance Kruse at lmkruse2@ncsu.edu.

References

- Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education*, 17(2), 86-95. doi:10.1177/088840649401700203
- Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. L. W. Anderson & D. R. Krathwohl (Eds.). Boston, MA: Pearson
- American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), & National Education Association (NEA) (1990). *Standards for Teacher Competence in the Educational Assessment of Students*. Retrieved from <https://eric.ed.gov/?id=ED323186>
- Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*: New York City, NY: Addison-Wesley Longman Limited.
- Balch, R., & Springer, M. G. (2015). Performance pay, test scores, and student learning objectives. *Economics of Education Review*, 44, 114-125. doi:10.1016/j.econedurev.2014.11.002
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, 2(1). doi:10.1177/2332858415616358
- Berman, P., McLaughlin, M. W., Bass-Golod, G.V., Pauly, E., & Zellman, G. L. (1977) *Federal programs supporting educational change: Vol. VII: Factors affecting implementation and continuation*. Santa Monica, CA: RAND Corporation. Retrieved from <https://www.rand.org/pubs/reports/R1589z7.html>
- Beziat, T. L., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, 27(1), 25-30. Retrieved from <http://www.nrmera.org/educational-research-theory-practice/ertp-past-issues/vol-27-issue-1-2015/>
- Campbell, C., Murphy, J. A. & Holt, J. K. (2002, October) *Psychometric analysis of an assessment literacy instrument: applicability to preservice teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.
- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education, 71*, 134-144. doi:10.1016/j.tate.2017.12.010
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed). Thousand Oaks: SAGE Publications.
- DeLuca C. & Lam, C. (2014). Preparing teachers for assessment within diverse classrooms: An analysis of teacher candidates' conceptualizations. *Teacher Education Quarterly, 3*, 3. Retrieved from <https://eric.ed.gov/?id=EJ1078652>
- DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education, 34*(5-6), 576-591. doi:10.1080/01626620.2012.730347
- DeLuca, C. & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability, *Assessment in Education: Principles, Policy & Practice, 24*(2), 121-126, doi:10.1080/0969594X.2017.1297010
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*(4), 419-438. doi:10.1080/0969594X.2010.516643
- DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgement on student learning: The role of pre-service assessment education. *Assessment in Education: Principles, Policy & Practice, 20*(1), 107-126. doi:10.1080/0969594X.2012.668870
- DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *Teacher Educator, 48*(2), 128-142. doi:10.1080/08878730.2012.760024
- Dial, E. (2015). An experienced practical reflection: Creating confident, effective teachers by changing pre-service programs to promote resilience, assessment literacy, and collaboration. *National Teacher Education Journal, 8*(1), 33-39. Retrieved from <http://0-search.ebscohost.com.carlson.utoledo.edu/login.aspx?direct=true&db=ehh&AN=110268126&site=eds-live>
- Dunning, D. (2011). The Dunning–Kruger Effect. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 247–296). Elsevier. doi:10.1016/B978-0-12-385522-0.00005-6

- Falter Thomas, A. R., & Sondergeld, T. A. (2015). Investigating the impact of feedback instruction: Partnering preservice teachers with middle school students to provide digital scaffolded feedback. *Journal of Scholarship of Teaching and Learning*, 15(4), 83-109.
- Garrett, R., & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, 37(2), 224-242. doi:10.3102/0162373714537551
- Gerde, H. K., Pierce, S. J., Lee, K., & Van Egeren, L. A. (2017). Early childhood educators' self-efficacy in science, math, and literacy instruction and science practice in the classroom. *Early Education and Development*, 1-21. doi:10.1080/10409289.2017.1360127
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76(4), 569-582. Retrieved from <http://dx.doi.org/10.1037/0022-0663.76.4.569>
- Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational researcher*, 44(5), 293-307. doi:10.3102/0013189X15592622
- Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, 37, 1001-1013.
- Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, 4(1), 63-69. doi:10.1016/0742-051X(88)90025-X
- Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principals and teacher value-added measures. *American Educational Research Journal*, 51(1), 73-112. doi:10.3102/0002831213517130
- Haun, D. E., Zeringue, A., Leach, A., & Foley, A. (2000). Assessing the competence of specimen-processing personnel. *Laboratory Medicine*, 31, 633-637.
- Howley, M.D., Howley, A., Henning, J.E., Gilla, M.B. & Weade, G. (2013) Intersecting domains of assessment knowledge: School typologies based on interviews with secondary teachers. *Educational Assessment*, 18(1), 2648. doi:10.1080/10627197.2013.761527
- Hoy, A. W. (2000, April). *Changes in teacher efficacy during the early years of teaching*. In annual meeting of the American Educational Research Association, New Orleans, LA. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.2527&rep=rep1&type=pdf>

- Jerald, C. D. (2007). *Believing and achieving* (Issue Brief). Washington, DC: Center for Comprehensive School Reform and Improvement.
- Jimerson S., Stein R., Haddock A., Shahroozi R. (2016) Common core state standards and response to intervention: The importance of assessment, intervention, and progress monitoring. In Jimerson S., Burns M., VanDerHeyden A. (eds) *Handbook of Response to Intervention* (165 – 184). Springer, Boston, MA. doi:10.1007/978-1-4899-7568-3_11
- Knaggs, C. M., & Sondergeld, T. A. (2015). Science as a learner and as a teacher: Measuring Science self-efficacy of elementary pre-service teachers. *School Science and Mathematics Journal*, 115(3), 117-128.
- Koh, K., Carol-Ann Burke, L. E., Luke, A., Gong, W., & Tan, C. (2017). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*. doi:10.1177/1362168816684366
- Koloi-Keaikitse, S. (2016). Assessment training: A precondition for teachers' competencies and use of classroom assessment practices. *International Journal of Training and Development*, 20(2), 107-123. doi:10.1111/ijttd.12072
- Koriat, A. (2008). When confidence in a choice is independent of which choice is made. *Psychonomic Bulletin & Review*, 15, 997–1001.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-34.
- Lee, J.-E., & Son, J.-W. (2015). Two teacher educators' approaches to developing preservice elementary teachers' mathematics assessment literacy: Intentions, outcomes, and new learning. *Teaching and Learning Inquiry: The ISSOTL Journal*, 3(1), 47-62. doi: 10.2979/teachlearninqu.3.1.47
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching & Teacher Education: An International Journal of Research and Studies*, 20(5), 523-535. Retrieved from <https://eric.ed.gov/?id=EJ731221>
- Mandinach, E., Friedman, J. M., & Gunner, E. (2015). How can schools of education help to build educators' capacity to use data? A systemic view of the issue. *Teachers College Record*, 117(4), 1-50. Retrieved from <https://eric.ed.gov/?id=EJ1056728>
- Mertler, C. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49-64. Retrieved from <http://www.jstor.org/stable/41064623>

- Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education, 120*(2), 285-296. Retrieved from <https://www.questia.com/library/journal/1G1-59644154/assessing-student-performance-a-descriptive-study>
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101-113. doi:10.1177/1365480209105575
- Mertler, C.A. & Campbell, C. S. (2005, April) Measuring teachers' knowledge and application of classroom assessment concepts: development of the *Assessment Literacy Inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada. Retrieved from <https://eric.ed.gov/?id=ED490355>
- National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal, 84*(2), 113-130. doi:10.1086/461348
- National Council on Teacher Quality. (2018). NCTQ: Yearbook. Retrieved from <https://www.nctq.org/yearbook/>
- Odo, D. M. (2016). An investigation of the development of pre-service teacher assessment literacy through individualized tutoring and peer debriefing. *Journal of Inquiry and Action in Education, 7*(2), 31-61. Retrieved from <https://eric.ed.gov/?id=EJ1133529>
- Ohio Department of Education. (2018). *Resident educator program*. Retrieved from <http://education.ohio.gov/Topics/Teaching/Resident-Educator-Program>
- One Hundred Seventh Congress of the United States of America. (2002, January 8). *H.R.1 - 107th Congress (2001-2002): No Child Left Behind Act of 2001*. Retrieved from <https://www.congress.gov/bill/107th-congress/house-bill/1>
- One Hundred Third Congress of the United States of America. (1994). *Improving America's Schools Act of 1994* (H.R.6). Retrieved from <https://www.gpo.gov/fdsys/pkg/BILLS-103hr6enr/pdf/BILLS-103hr6enr.pdf>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578. doi:10.3102/00346543066004543
- Pereira, D., Niklasson, L., & Flores, M. A. (2017). Students' perceptions of assessment: a comparative analysis between Portugal and Sweden. *Higher education, 73*(1), 153-173. doi:10.1007/s10734-016-0005-0

- Plake, B. S. (1993) Teacher assessment literacy: teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(2), 21–7. Retrieved from <https://eric.ed.gov/?id=EJ460188>
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *Teacher Educator*, 46(4), 265-273. doi:10.1080/08878730.2011.605048
- Popham, W.J. (2008). *Classroom assessment: What teachers need to know* (5th ed.), New York, NY: Pearson/Allyn & Bacon.
- Prince, C. D., Schuermann, P. J., Guthrie, J. W., Witham, P. J., Milanowski, A. T., & Thorn, C. A. (2009). *The other 69 percent: Fairly rewarding the performance of teachers of nontested subjects and grades (revised ed.)*. Center for Educator Compensation Reform. Retrieved from: <http://www1.gcsnc.com/whatmatters/pdf/other69Percent.pdf>
- Reeves, T. D. & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, 50, 90-101. doi:10.1016/j.tate.2015.05.007
- Schoon, K. J., & Boone, W. J. (1998). Self-efficacy and alternative conceptions of science of preservice elementary teachers. *Science Education*, 82(5), 553-568. doi:10.1002/(SICI)1098-237X(199809)82:5<553::AID-SCE2>3.0.CO;2-8
- Self-Efficacy in Science, Math, and Literacy Instruction and Science Practice in the Classroom. *Early Education and Development*, 1-21. doi:10.1080/10409289.2017.1360127
- Sheldon, O. J., Dunning, D., & Ames, D. R. (2014). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology*, 99(1), 125–137. doi:10.1037/a0034138
- Sondergeld, T. A. (2014). Closing the gap between STEM teacher classroom assessment expectations and skills. *School Science and Mathematics Journal*, 114(4), 151-153.
- Sondergeld, T. A., Rychener, S. R., & Koskey, K. L. (2015). *2014-2015 Oregon City Schools Straight A Fund evaluation report*. Bowling Green, OH: BGSU Center of Assessment and Evaluation Services.
- Stein, M. K., & Wang, M. C. (1988). Teacher development and school improvement: The process of teacher change. *Teaching and Teacher Education*, 4(2), 171-187. doi:10.1016/0742-051X(88)90016-9
- Stiggins, R. (2014). Improve assessment literacy outside of schools too. *Phi Delta Kappan*, 96(2), 67-72. doi:10.1177/0031721714553413
- Suskie, L. (2018). *Assessing student learning: A common sense guide*. John Wiley & Sons.

- Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, 23(6), 944–956. doi:10.1016/j.tate.2006.05.003
- United States Department of Education. (2018). *Every Student Succeeds Act (ESSA)*. Retrieved from <https://www.ed.gov/essa?src=ft>
- Volante L. & Beckett D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education* (2), 239-255. Retrieved from <https://eric.ed.gov/?id=EJ936752>
- Volante, L. & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education* (3), 749. Retrieved from <https://eric.ed.gov/?id=EJ780818>
- Wang, T. H., Wang, K. H., & Huang, S. C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, 51(1), 448-462. doi:10.1016/j.compedu.2007.06.010
- Yamtin, V., & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioral Sciences*, 116, 2998–3004. doi:10.1016/j.sbspro.2014.01.696

Appendix A

Alignment between *Standards for Teacher Competence in Educational Assessment*, *ALI* items, *CASES* items, and Assessment Course Lessons

While *ALI* items were designed to align with a specific *Standard*, *CASES* items and Course Lessons often covered components addressed in multiple *Standards*. As such, *ALI* items are only listed once in their column showing their alignment with a single *Standard*. Yet *CASES* items and Course Lessons are aligned with multiple *Standards* (or pieces from *Standards*).

Standard	<i>ALI</i> Items	<i>CASES</i> Items	Course Lessons
1. Choosing Assessments	1, 8, 15, 22, 29	1, 4, 5, 13, 16, 17	1, 2, 3, 4, 11, 13
2. Developing Assessments	2, 9, 16, 23, 30	1, 2, 4, 6, 7, 13, 14, 16, 18, 19	2, 3, 5, 8, 9
3. Administering, scoring, and interpreting assessment results	3, 10, 17, 24, 31	3, 4, 9, 10, 12, 15, 16, 21, 22, 24	4, 5, 6, 7, 8, 10, 13
4. Using assessment results for decision-making	4, 11, 18, 25, 32	3, 11, 12, 15, 23, 24	1, 5, 7, 13
5. Developing Grading Procedures	5, 12, 19, 26, 33	9, 10, 21, 22	7, 8, 10, 11
6. Communicating Assessment Results	6, 13, 20, 27, 34	3, 10, 12, 15, 22, 24	5, 6, 7, 8, 10, 13
7. Recognizing Inappropriate Assessment	7, 14, 21, 28, 35	4, 10, 12, 16, 22, 24	1, 4, 7, 12, 13