Examining the Instructional Priorities of a Secondary Education Teacher Preparation Program Utilizing an Integrated Approach to Assessment Education

Daniel M. Frederking
American Institutes for Research

This study examined the instructional priorities of a four-year undergraduate secondary education program that does not offer an assessment-specific course, but instead integrates assessment instruction throughout the broader curriculum. An alignment method was utilized to better understand the program's instructional priorities for teaching about assessment in comparison to the locally-used professional teaching standards and the locally-used teacher accreditation performance assessment. Areas of both alignment and misalignment were found in three dimensions: categorical concurrence (CC), depth of knowledge (DOK), and range of knowledge (ROK). Findings include course syllabi objectives that were written at a broader ROK level than the other two sources, syllabi and performance assessment objectives that were written at a higher DOK level than the standards, and a variety of approaches to teaching the concepts of formative assessment. Implications for teacher education are discussed.

Introduction

Student assessment serves as one of the central pillars of the educational process, yet it remains one of the most misused and misunderstood aspects in schools. There has been much debate and many research studies that examine the fundamental aspects of both preservice and in-service teacher education on the topic of assessment, yet the concerns over educators' assessment literacy remain prevalent in the current landscape (Campbell, 2013; Opre, 2015; Siegel & Wissehr, 2011). As employers increasingly desire assessment literate teachers to fill their schools, teacher education programs continue to search for effective ways to prepare their candidates for a field that consistently expects greater use of student data to guide everyday classroom instruction.

Generally, the term "assessment literacy" refers to the basic understanding of assessment practices and the ability to apply these skills to accurately collect evidence of student achievement (Stiggins, 1991). Though in the past it might have been enough to simply understand assessment for grading purposes, the field has grown to demand more advanced assessment skills including the frequent use of data to inform and individualize instructional practices. These advances have come on the heels of monumental publications such as Black and Wiliam's (1998) findings on the educational impacts of formative assessment and Hattie's (2009) meta-analysis of factors relating to student achievement, many of which were assessment-related.

The value of assessment literacy is unquestionable now as standardized assessments have become the norm in the United States and as new laws are passed requiring student learning data to be used in teacher evaluations. As Popham (2011), a teacher educator at UCLA said as of

several years ago, "…for teachers to remain ignorant about the nature of educational assessments these days surely constitutes the quintessence of professional suicide" (p. 269). This growing sense of urgency puts more pressure on the preservice teacher education programs to produce teachers who are capable of understanding and utilizing assessment data to their full potential.

Approaches to Including Instruction on Assessment

DeLuca and Klinger (2010) categorize teacher education programs' approaches to teaching assessment literacy into three classifications: explicit, integrated, and blended. Explicit assessment education involves requiring specific assessment-centered courses as part of the larger program. Integrated assessment education incorporates assessment instruction into the broader curriculum without offering specific courses on the topic. A blended approach is a combination of both explicit and integrated assessment education models. Regardless of the approach a program takes, the assessment curriculum has typically been unstandardized across institutions and the instruction methods controlled by the individual instructors (DeLuca, Klinger, Searle, & Shulha, 2010).

DeLuca and Bellara (2013) examined programs using an explicit approach to teaching about assessment by formulating a study that considered the education priorities on the topic of assessment in three different realms: university assessment course syllabi, teacher education policy documents, and standards for teacher practice. They analyzed course syllabi from 10 Florida teacher education programs, policy documents including Florida and national accreditation guidelines, and various teaching and student assessment standards. Using an alignment model, they examined the congruence between the three realms by exploring three dimensions including categorical concurrence (CC), depth of knowledge (DOK), and range of knowledge (ROK). Their study yielded many areas of alignment and misalignment within all three realms.

The DeLuca and Bellara (2013) study was unique because it focused on the priorities of various programs and organizations regarding assessment education. By examining the assessment course syllabi from 10 different programs, accreditation documents, and teaching standards, they put the emphasis on what Porter and Smithson (2001) define as the *intended curriculum*. They classify this term as being "such policy tools as curriculum standards, frameworks, or guidelines that outline the curriculum teachers are expected to deliver" (p. 2). This category is distinguished from their other two categories, the *assessed* curriculum, which is "the content on high-stakes tests" (p. 3), and the *enacted curriculum*, which is "the actual curricular content that students engage in the classroom" (p. 2).

In their study, DeLuca and Bellara (2013) found many areas of alignment and misalignment among the sources. They discussed the trend of course content being reasonably matched to standards by providing objectives that were specific and useful for daily practice. Policy documents, however, represented more global objectives that were broad and could require years to fully accomplish (e.g., teachers will be assessment literate). As these policy documents were used for accreditation purposes, the authors asserted that this misalignment must mean that assessment concepts were being addressed more fully in other areas of the program outside of

the assessment courses, which were short and not capable of covering the policy concepts in their entirety.

Greenberg and Walsh (2012) had previously used course syllabi as a source of data to examine the assessment instruction of 180 teacher preparation programs. They found that many programs were weak in some specific areas. Only 21% of the programs they studied adequately covered how to measure student performance using assessments, while less than 1% adequately covered how to analyze student performance data and less than 2% adequately covered how to use data from assessments to plan instruction. They also wrote of the tendency for most explicit assessment courses to be only one semester long and to focus mostly on introducing assessment policies and practices to the preservice teachers. Concurring with the findings of DeLuca and Bellara (2013), it would be very difficult for these courses to fully cover the breadth of the topic.

Though DeLuca and Bellara's (2013) study was commendable for its examination of programs utilizing the explicit approach, there is still a need for a similar look at a program using the integrated approach. Following the example set by DeLuca and Bellara, this current study adapted the alignment model to better understand the priorities of a program that does not require students to enroll in a stand-alone assessment course. It also introduced new realms to the alignment method, specifically the assessment standard and indicators of the locally-used professional teaching standards and the handbook and rubrics of the standardized teacher accreditation performance assessment. This study focused on secondary education teacher preparation programs, which introduces even more challenges due to the fragmented nature of educating multiple content areas within one program. The following research question guided the study: How closely does assessment instruction align with assessment practices as defined by the professional teaching standards and the standardized teacher accreditation performance assessment?

Theoretical Framework

A common concern in the literature is that instruction on educational assessment tends to be far too theory laden and teachers are entering the field without the practical skills to effectively assess (DeLuca, 2012). Siegel and Wissehr (2011), for example, found that though teachers were able to demonstrate strong assessment knowledge through conversations and surveys, they were not applying these skills during their practical classroom experiences. Wallace and White (2015) studied preservice teachers and found they generally learned how to assess before they learned why to assess. This finding, which corresponds with the alignment findings of DeLuca and Bellara (2013), adds to the concern that teachers have not progressed to the point where they are prepared to successfully use assessments in their own classrooms.

When looking specifically at secondary education teacher preparation, assessment instruction can become even murkier. Because many of these programs are segmented into different content areas, it can be difficult to teach preservice teachers about assessment in a way that applies to all students and subjects. Greenberg and Walsh (2012) found that 58% of secondary education programs either did not have a subject-specific methods course, or they did have one, but it did not address assessment. Talanquer, Bolger, and Tomanek (2015) qualitatively researched the

grading practices of preservice secondary teachers and found that they tended to focus more on the basics of assessment and simple description and less on a full understanding of student ideas.

Grainger and Adie (2014) surveyed secondary preservice teachers in an Australian education program and concluded that a single assessment course was not enough to properly prepare future assessors. The individuals in this study struggled with consistency in grading, using rubrics, and providing feedback. DeLuca (2013), on the other hand, surveyed preservice teachers and found that an assessment methods course did make a big difference in both assessment understanding and confidence levels. Hill, Gunn, Cowie, Smith, and Gilmore (2014) reported similar findings in their study of several teacher education programs in New Zealand.

For assessment education to be effective, a theoretical framework must first define its necessary aspects. Multiple frameworks have been developed, including Xu and Brown (2016)'s teaching assessment literacy in practice framework and Gottheiner and Siegel (2012)'s theory of assessment literacy, but none of these models met the specific needs and intricacies of secondary education. To inform the current study, an original theoretical framework known as framework of assessment education for secondary teachers (FAST) was developed. The framework is organized into three building blocks: what teachers should know, where it should be learned, and how it should be learned. In Figure 1, the framework is presented as an architectural structure where the cornerstone (what teachers should know) provides the foundation on which the remaining building blocks can be taught. As the arrows indicate, the decisions on assessment education must be made in this order: what, where, how. Once all building blocks are firmly in place, the structure of teacher understanding can be fully realized.

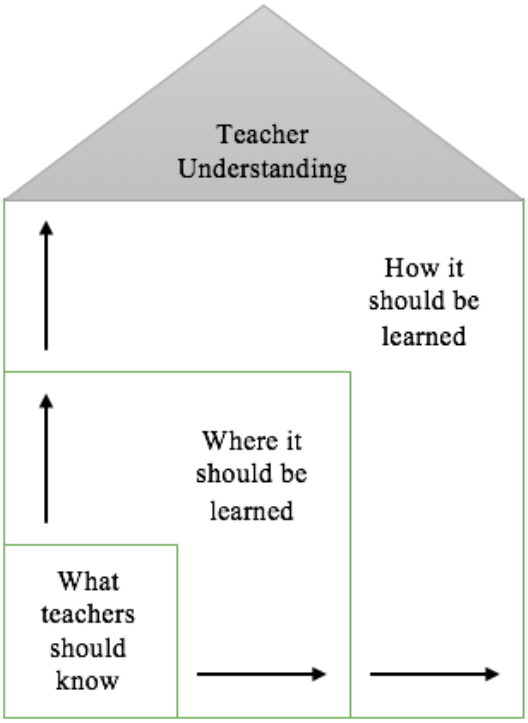

Figure 1. The Framework of Assessment Education for Secondary Teachers

In Table 1, the specific components of each building block are listed. The content components of what teachers should know draw from the literature on the topic, both practical and theoretical. Primarily, the content components were developed using Standard 6 of the Interstate New Teacher Assessment and Support Consortium (InTASC) professional teaching standards created by the Council of Chief State School Officers (2011), the assessment aspects from Charlotte Danielson's *Enhancing Professional Practice: A Framework for Teaching* (2007), and the ideas established by Shepard, Hammerness, Darling-Hammond, and Rust (2005) in their chapter on assessment in the influential and commonly-used book, *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*.

Table 1
The Framework of Assessment Education for Secondary Teachers: Content Components

What teachers should know	→	Where it should be learned	→	How it should be learned
Alignment to learning goals Diverse learners Feedback and motivation Formative and summative Language and literacy Results to guide instruction Statistical literacy Student self-assessment Tools and types		University • School of education • Content area Employer		Instruction Practical experience

Using these three resources as the focal point of this framework was appropriate for two reasons: Firstly, these three resources are heavily used within the profession, resulting in a large number of preservice and in-service teachers being exposed to them. Secondly, considering these three works combined allows for an examination of assessment in both the preservice and the in-service realms. Additionally, the framework drew on recommendations from organizations like the National Council for Accreditation of Teacher Education (2008), the Teacher Education Accreditation Council (2014), the Council for the Accreditation of Educator Preparation (2013), and the Joint Committee on Standards for Educational Evaluation (Klinger, McDivitt, Howard, Munoz, Roger, & Wylie, 2015) in conjunction with recommendations from experts in the field (Black & Wiliam, 1998; Hattie, 2009; Mandinach & Gummer, 2013; 2016; Marzano, 2010; Pierce & Chick, 2011; 2013; 2014).

Methodology

Following the example of DeLuca and Bellara (2013), this examination took the form of an alignment study, as the researcher developed a content analysis with a desire to present frequencies. The study centered on five undergraduate secondary education programs at a large

state university in the Midwestern United States. In these programs, undergraduate students must take a combination of courses within the School of Curriculum and Instruction (CI) and their chosen content area department (i.e. English, mathematics, theater, health, etc.). The university does not offer an educational assessment-specific course, so assessment instruction is incorporated throughout the other courses.

The education programs are guided by the professional teaching standards. Within this set of standards is the following concerning assessment: "The teacher understands and uses multiple methods of assessment to engage learners in their own growth, to monitor learner progress, and to guide the teacher's and learner's decision making" (Council of Chief State School Officers, 2011, n.p.). Under this standard there are 22 indicators that outline the specific assessment skills needed to be proficient in educational assessment.

To receive their teaching license in this state, teacher candidates must pass a performance-based assessment that requires them to submit various teaching artifacts to demonstrate their teaching abilities. Preservice teachers submit their portfolio during their student teaching experience. It is then scored by trained scorers (not employed by the university) using rubrics that are specific to the preservice teacher's teaching field.

While DeLuca and Bellara (2013) focused their study on the *intended curriculum*, this study reached a little more broadly. The professional teaching standards were used as *intended curriculum* while an examination of the performance assessment constituted as the *assessed curriculum* and the course syllabi stood in as the *enacted curriculum*. DeLuca and Bellara (2013) classified the syllabi they used as part of the *intended curriculum*, but the author of this study viewed the syllabi's use differently. The syllabi exist to provide a specific synopsis of course topics, objectives, and assignments. Though it could be argued that this falls under the Porter and Smithton (2001) definition of *intended curriculum*, which reads "such policy tools as curriculum standards, frameworks, or guidelines that outline the curriculum teachers are expected to deliver" (p. 2), the author of this study classified them as *enacted curriculum*, or "the actual curricular content that students engage in the classroom" (p. 2). There are issues with this classification as well since *enacted curriculum* is usually captured using tools like surveys and observations, but the author made this decision based on the syllabi being a record of what the university requires their instructors to teach. Syllabi might leave some content out, but the content that is present in the syllabi is required to be covered.

The researcher utilized a form of document analysis to examine data from a sampling of 17 course syllabi, the performance assessment handbook (which contains its 15 rubrics), and the teaching standard and its indicators. A convenience sample of syllabi was obtained by asking the directors of all secondary subject areas to participate. Five programs granted permission to examine their methods course syllabi: biology, chemistry, communication, physics, and theater. Biology offered one methods course, chemistry offered four, communication offered three, physics offered four, and theater offered three. Additionally, the School of Curriculum and Instruction supplied the three required secondary education courses within the department.

This study used an alignment model developed by Webb (1997, 1999, 2005) and modified by DeLuca and Bellara (2013). All standards, syllabi, and rubrics were classified and coded in three

different dimensions: categorical concurrence (CC), depth of knowledge (DOK), and range of knowledge (ROK) (Webb, 1997, 1999, 2005). The CC coding used a deductive approach based on the content components of the FAST framework: *alignment to learning goals, diverse learners, feedback and motivation, formative and summative, language and literacy, results to guide instruction, statistical literacy, student self-assessment,* and *tools and types.* Codes were also discovered during the study and retroactively included in the framework to provide a more explicit view of the research findings. The code criteria for these themes can be found in Appendix A. The DOK codes were adapted from Webb's DOK by DeLuca and Bellara (2013). The ROK levels were again defined by DeLuca & Bellara (2013) and adapted from Russell and Airasian (2011) and their typology of global standards, educational standards, and instructional standards, which in turn was adapted from the work done by Krathwohl and Payne (1971).

Because it would be impossible to assign a DOK and ROK level to every mention of assessment within the sources, the decision was made to only code objectives (or content resembling objectives) in all three dimensions. Additional content was only coded in the CC dimension and considered separately (for example, mentions of assessment within the syllabi were often brief like, "Introduction to Standard Assessments" or "Student Performance Assessments: Scoring Rubrics").

After the initial coding, the researcher obtained IRB approval and met with each of the content area department contacts to give them a chance to confirm or dispute any of the findings and to inquire about any other assessment instruction that occurred but was not listed in the syllabi. The findings were updated based on these member-checks and interviews.

Results and Discussion

Frequency tables were constructed to display the representation of each code relative to each data source. The frequencies are listed as percentages displaying the proportion of each individual code in relation to the overall number of codes within that dimension. Table 2 displays the frequencies across the three dimensions. Table 3 displays the frequencies across the individual syllabi. As can be seen in Table 3, all syllabi contained assessment-related objectives except for theater.

Contrary to DeLuca and Bellara (2013)'s findings, the dominant ROK level for this program's syllabi was *global* (though the other two levels were represented as well). The *global* level was the least commonly coded level on the syllabi studied by DeLuca and Bellara (2013). This does not align with the ROK levels of the standard and the performance assessment, which were both heavily coded as *educational*. This was a very interesting finding as it indicated syllabi objectives are often written in a manner in which they most likely cannot be accomplished within the course itself. This can be traced mostly to curriculum and instruction, communication, and chemistry courses. The standard and the performance assessment, on the other hand, were written in a way that they could be accomplished as a direct result of instruction. This can be considered a significant area of misalignment that could result in the teaching candidates not walking away from this program with the skills that are dictated by the standard and evaluated by the assessment.

Table 2
Frequency Table of Codes Across Three Major Coding Sources (Expressed as Percentages)

Category	Code	Syllabi	Standard	Assessment
Categorical Concurrence	Alignment to learning goals	7	12	0
	Diverse learners	14	15	13
	Feedback and motivation	0	12	25
	Formative and summative	4	6	0
	Language and literacy	11	0	13
	Results to guide instruction	21	6	25
	Statistical literacy	21	18	13
	Student self-assessment	0	12	0
	Tools and types	21	18	13
Depth of Knowledge	Level 1: Low	13	32	0
	Level 2: Moderate	42	50	71
	Level 3: High	46	18	29
Range of Knowledge	Level 1: Instructional	33	0	0
	Level 2: Educational	21	91	100
	Level 3: Global	46	9	0

Table 3
Frequency Table of Codes Across Each Individual Syllabus (Expressed as Percentages)

		Content Area					
Category	Code	BIO	CHE	COM	PHY	THT	CI
Categorical Concurrence	Alignment to learning goals	25	0	0	20	-	0
	Diverse learners	0	0	25	20	-	13
	Feedback and motivation	0	0	0	0	-	0
	Formative and summative	0	33	0	0	-	0
	Language and literacy	0	0	13	0	-	25
	Results to guide instruction	0	33	38	20	-	13
	Statistical literacy	50	0	13	40	-	13
	Student self-assessment	0	0	0	0	-	0
	Tools and types	25	33	13	0	-	38
Depth of Knowledge	Level 1: Low	17	0	17	0	-	20
	Level 2: Moderate	33	0	50	40	-	60
	Level 3: High	50	100	33	60	-	20
Range of Knowledge	Level 1: Instructional	33	50	17	80	-	0
	Level 2: Educational	33	0	33	0	-	20
	Level 3: Global	33	50	50	20	-	80

Note: BIO (biology), CHE (chemistry), COM (communication), PHY (physics), THT (theater), CI (curriculum and instruction)

Regarding DOK, the syllabi put most of their emphasis on the *moderate* (42%) and *high* (46%) levels, as does the performance assessment (71% and 29% respectively). The standard, on the other hand, is written mostly at a *low* (32%) and *moderate* (50%) DOK level. This indicated that the syllabi and performance assessment were asking teacher candidates to engage in higher order thinking skills, yet the standard did not require them to think at the same level. Because of its structure, it is not surprising that the performance assessment engaged teachers at a higher level, but this study highlights the encouraging trend that the university is pushing their students to synthesize, evaluate, and create, instead of simple recall and memorization.

It is important to remember that these tables only include the assessment-related objectives written in each syllabi which, though required by the university, do not always fully capture everything that happens in the program. One specific omission is the student teaching experience, which would most likely contain a high DOK level of assessment learning. It could be assumed that these objectives do include elements of this, as many of these courses do require clinical experiences, but the full clinical component cannot be assumed to be present.

To the point made by several researchers that assessment education is often too theory laden (DeLuca, 2012; Siegel & Wissehr, 2011), this study can offer some interesting perspectives. Because the syllabi objectives tend to be written at a moderate to high DOK level, there is evidence to suggest that teacher candidates are being frequently asked to apply their skills in practical situations. However, it is not within the scope of this study to indicate whether these skills are being translated into the classroom.

Diverse learners is the CC code with the closest alignment across sources. This suggests the equal importance that is placed on preparedness to assess the wide range of backgrounds, abilities, and learning styles within a classroom. Considering educational priorities like Response to Intervention (RtI), Individualized Education Programs (IEPs), and equity initiatives, this discovery shows the emphasis all content areas place on the topic.

Of the several areas of misalignment, *feedback and motivation* stands out as one of the most prevalent. The performance assessment emphasizes this skill heavily, while the syllabi objectives do not (and the standard falls in the middle). This code was applied to the supplementary content of four different syllabi, however, so it is not a topic that is completely ignored.

Student self-assessment, on the other hand, was not coded anywhere in the syllabi's course objectives nor was it found anywhere in the supplementary pieces of content. Students self-reporting their grades is listed as one of the top factors on Hattie (2009)'s list of the biggest factors related to student achievement. Its effect size cannot be ignored. It was found in four different indicators in the standard (12%) but was not found in the performance assessment either.

On a related note, the most common code applied through the interview process was *formative and summative*, as four different content area professors indicated this was taught. This code only appeared in one syllabi objective (CHE 303), however. This code was intended to capture the explicit coverage of the differences between the two types of assessment. The professors often volunteered that an investigation of formative assessment and its differences from

summative assessment was being taught, yet this blatant lesson did not often find its way into the objectives. This topic might be considered a "building block" for a deeper level of instruction on assessment because understanding the basic types of assessment is necessary to understand other aspects of the topic. This might partially explain why it was the most commonly inserted code during the interview process. Additionally, this could suggest that the concepts are certainly taught, but in more subtle ways. The findings of this study might indicate the need to combine this code with others, such as the *results to guide instruction* or *feedback and motivation*, which both share traits with the idea of formative assessment.

The *language and literacy* code was not originally included in the framework but was discovered during this study. The code was included to show the opportunities students had to express themselves through language. An effective assessor must consider students' literacy when designing assessments, regardless of the subject area. This code was initially discovered when elements of it appeared in the performance assessment handbook. One rubric, for example, features the question, "How does the candidate analyze students' use of language to develop content understanding?" It was determined that this did not fit into any of the established categories and, when similar objectives were observed in the syllabi, it was clear a new category had been discovered.

Conclusion

This alignment study focused on a teacher education program that uses an integrated approach to teaching assessment. Integrated approaches tend to be less common, yet this method can still be frequently found in American teacher education programs. DeLuca and Bellara (2013) formulated a study to examine the priorities of programs with an explicit approach in comparison to other policy and guidance sources, but this left a literature gap for alignment study centered on a program with an integrated approach. This study chose to focus on secondary education due to the additional complications this area poses.

Similar to DeLuca and Bellara's (2013) study, this study found areas of both alignment and misalignment. One key finding was the tendency for course syllabi to write their assessment objectives at a *global* ROK level. This does not align with the other two sources, which are dominantly written at the *educational* level. This also does not align with the syllabi ROK coding from DeLuca and Bellara (2013)'s study, where *global* was the least coded level. Another key finding is the moderate to high DOK levels that are present within the syllabi and the performance assessment, but not quite as much in the standard. A third key finding is the variety of ways in which formative assessment is taught (using the codes *formative and summative*, *results to guide instruction*, and *feedback and motivation*). This made examining and coding the topic difficult.

This examination lends itself to new questions and several areas of recommended research. To strengthen the data used as the *enacted* curriculum, future researchers could use techniques like observations or the Surveys of Enacted Curriculum (SEC). The SEC is a commonly used method that utilizes a survey tool to gather data from instructors on how much time and emphasis they place on various aspects of instruction (Blank, 2002; Porter, 2002). This method could contribute to a stronger discussion that more closely mirrors what is actually happening in the course.

This alignment study could also bring in new sources of data from the educational field by interviewing practicing teachers or principals. Additionally, aligning the priorities of a program with an integrated approach to a program with an explicit approach could offer some insight on the pros and cons of each method.

There would also be value in conducting a similar alignment study that includes a program using a blended approach to teaching assessment. This would provide new perspectives on the field of assessment education and it could highlight some of the strengths and weaknesses of each approach and provide a stronger knowledge base of the factors that can best prepare classroom teachers for the challenges of student assessment.

Author Notes

Daniel M. Frederking is a technical assistance consultant at American Institutes for Research.

Correspondence regarding this article should be addressed to Daniel M. Frederking at dfrederking@air.org.

References

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Blank, R. (2002). Using surveys of enacted curriculum to advance evaluation of instruction in relation to standards. *Peabody Journal of Education, 77*(4), 86-121.

Campbell, C. (2013). Research on teacher competence in classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 71-84). Los Angeles, CA: SAGE.

Council for the Accreditation of Educator Preparation. (2013). 2013 CAEP standards. Retrieved from http://www.caepnet.org.

Council of Chief State School Officers. (April 2011). *InTASC model core teaching standards: A Resource for state dialogue*. Retrieved from https://ccsso.org/sites/default/files/2017-11/InTASC_Model_Core_Teaching_Standards_2011.pdf

Danielson, C. (2007). *Enhancing professional practice: A Framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

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.

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education, 64*(4), 356-372.

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.

DeLuca, C., Klinger, D., Searle, M., & Shulha, L. (2010). Developing a curriculum for assessment education. *Assessment Matters, 2*, 133-155.

Grainger, P. R., & Adie, L. (2014). How do preservice teacher education students move from novice to expert assessors?. *Australian Journal of Teacher Education, 39*(7), 88-105.

Greenberg, J., & Walsh, K. (2012). *What teacher preparation programs teach about K-12 assessment: A review.* Retrieved from https://files.eric.ed.gov/fulltext/ED532766.pdf

Gottheiner, D. M. & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education, 23*(1), 531-557.

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. New York, NY: Routledge.

Hill, M. F., Gunn, A., Cowie, B., Smith, L. F., & Gilmore, A. (2014). Preparing primary and early childhood initial teacher education students to use assessment in teaching. [Special Issue]. *Assessment Matters, 7*, 4-23.

Klinger, D. A., McDivitt, P. J., Howard, B. B., Munoz, M. A., Roger, W. T., & Wylie, E. C. (2015). *Classroom assessment standards for PreK-12 teachers: Joint committee on standards for educational evaluation* [Kindle Edition].

Krathwohl, D. R., & Payne, D. A. (1971). Defining and assessing educational objectives. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 17-41). Washington, DC: American Council on Education.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(30), 30-37.

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education, 60,* 366-376.

Marzano, R. J. (2010). *Formative assessment & standards-based grading.* Bloomington, IN: Solution Tree Press.

National Council for Accreditation of Teacher Education. (2008). *Unit accreditation standards*. Retrieved from http://www.ncate.org/~/media/Files/caep/accreditation-resources/ncate-standards-2008.pdf?la=en

Opre, D. (2015). Teachers' conceptions of assessment. *Procedia - Social and Behavioral Sciences, 209*, 229-233.

Pierce, R., & Chick, H. (2011). Reacting to quantitative data: Teachers' perceptions of student achievement reports. In J. Clark, B. Kissane, J. Mousley, T. Spencer, & S. Thornton. (Eds.), *Mathematics: Traditions and [New] Practices: Proceedings of the 34th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 631-639). Adelaide, SA: AAMT.

Pierce, R., & Chick, H. (2013). Workplace statistical literacy for teachers: Interpreting box plots. *Mathematics Education Research Journal, 25*(2), 189-205.

Pierce, R., & Chick, H. (2014). Improving teachers' professional statistical literacy. *Topics From Australian Conferences on Teaching Statistics* (pp. 295-309). doi:10.1007/978-1-4939-0603-1_16.

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *Teacher Educator, 46*(4), 265-273.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3-14.

Porter, A. C. & Smithson, J. L. (2001). *Defining, developing, and using curriculum indicators.* (Research Report Series No. RR-048). Consortium for Policy Research in Education. Retrieved from https://pdfs.semanticscholar.org/85b5/c38554308ca52dedabb07eb3ab004f995477.pdf.

Russell, M. K., & Airasian, P. W. (2011). *Classroom assessment: Concepts and applications* (7th ed.). New York, NY: McGraw-Hill.

Shepard, L., Hammerness, K., Darling-Hammond, L., & Rust, F. (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275-326). San Francisco, CA: Josey-Bass.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education, 22*(4), 371-391.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72*, 534-539.

Talanquer, V., Bolger, M., & Tomanek, D. (2015). Exploring prospective teachers' assessment practices: Noticing and interpreting student understanding in the assessment of written work. *Journal of Research in Science Teaching, 52*(5), 585-609.

Teacher Education Accreditation Council. (2014). *Goals and principles*. Retrieved from http://www.teac.org/accreditation/goals-principles/.

Wallace, M., & White, T. (2015). Secondary mathematics preservice teachers' assessment perspectives and practices: An evolutionary portrait. *Mathematics Teacher Education and Development, 16*(2), 25-45.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states.* Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2005). *Webb alignment tool: Training manual*. Madison: Wisconsin Center for Education Research.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149-162.

Appendix A

Codes and Coding Criteria

Categorical Concurrence Code	Coding Criteria
Alignment to learning goals	The document mentions learning goals in conjunction with assessment.
Formative and summative	The document refers to the two assessment types and teaches students to distinguish between them.
Statistical literacy	The document refers to grading practices, assigning grades, validity, reliability, or data interpretation.
Feedback and motivation	The document refers to feedback and/or student motivation in relation to assessment.
Student self-assessment	The document refers to the theory and/or purposes of student self-assessment and its inclusion in the classroom.
Tools and types	The document mentions understanding or creating specific assessment tools and types.
Diverse learners	The document mentions an awareness of diversity and differentiation in relation to assessment.
Results to guide instruction	The document contains specific mentions of assessment data informing instruction.
Language and literacy	The document mentions language use or literacy skills in relation to assessment.

Depth of Knowledge Level	Coding Criteria
Level 1: Low cognitive level	Ability to identify, define, recognize and recall assessment knowledge.
Level 2: Moderate cognitive level	Ability to apply and analyze assessment knowledge. Establish connections between assessment knowledge, teaching practice, and person experiences.
Level 3: High cognitive level	Ability to evaluate, synthesize, and create assessment knowledge. Includes judging the quality and limitations of assessments as well as articulation of the linkage between assessment and other educational constructs.

Note. (DeLuca & Bellara, 2013, p. 369).

Range of Knowledge Level	Coding Criteria
Level 1: Instructional objective	Specific objectives used to plan assessment in daily teacher practice (e.g., test design, questioning approaches, etc.)
Level 2: Educational objective	Statements that describe teacher accomplishments that will result from instruction – specifically the behavior the teacher candidate will learn to perform and the content on which it will be performed. (e.g., teachers use assessment information to differentiate instruction and planning)
Level 3: Global objective	Very broad statements of intended learning that require years to accomplish (e.g., teachers will be assessment literate)

Note. (DeLuca & Bellara, 2013, p. 369).