

## Illinois State University ISU ReD: Research and eData

---

Theses and Dissertations

---


8-20-2018

# Teaching to Assess: An Evaluation of Assessment Education for Secondary Teacher Candidates

Daniel M. Frederking

*Illinois State University*, [danfrederking@gmail.com](mailto:danfrederking@gmail.com)

Follow this and additional works at: <https://ir.library.illinoisstate.edu/etd>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Elementary and Middle and Secondary Education Administration Commons](#), and the [Secondary Education and Teaching Commons](#)

---

### Recommended Citation

Frederking, Daniel M., "Teaching to Assess: An Evaluation of Assessment Education for Secondary Teacher Candidates" (2018). *Theses and Dissertations*. 1003.

<https://ir.library.illinoisstate.edu/etd/1003>

This Thesis and Dissertation is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ISU ReD: Research and eData. For more information, please contact [ISURed@ilstu.edu](mailto:ISURed@ilstu.edu).

TEACHING TO ASSESS: AN EVALUATION OF ASSESSMENT  
EDUCATION FOR SECONDARY  
TEACHER CANDIDATES

DANIEL M. FREDERKING

173 Pages

This study examined the assessment instruction of a four-year undergraduate secondary education program and its alignment to the teaching standards and the summative teacher licensure assessment. Document analyses were conducted on the program's syllabi as well as the assessment portions of the InTASC standards and the handbook of the edTPA. Results highlighted several areas of misalignment in the areas of categorical concurrence, depth of knowledge, and range of knowledge. Also of note, preservice teachers in different secondary subject areas are receiving different amounts of assessment instruction. Additionally, local edTPA data was examined to highlight areas of strength and weakness. Students did not perform significantly higher or lower on the assessment task as opposed to the other two tasks (planning and instruction), but there were significant differences among rubrics within the assessment task. A new theoretical framework is introduced to conceptualize the assessment education needed for secondary education teachers to be considered assessment literate.

**KEYWORDS:** assessment; assessment literacy; teacher education; secondary education

TEACHING TO ASSESS: AN EVALUATION OF ASSESSMENT  
EDUCATION FOR SECONDARY  
TEACHER CANDIDATES

DANIEL M. FREDERKING

A Dissertation Submitted in Partial  
Fulfillment of the Requirements  
for the Degree of

DOCTOR OF EDUCATION

School of Teaching and Learning

ILLINOIS STATE UNIVERSITY

2018

© 2018 Daniel M. Frederking

TEACHING TO ASSESS: AN EVALUATION OF ASSESSMENT  
EDUCATION FOR SECONDARY  
TEACHER CANDIDATES

DANIEL M. FREDERKING

COMMITTEE MEMBERS

Barbara Meyer, Chair

Kyle Miller

Mary O'Brian

## ACKNOWLEDGMENTS

Getting through this project would not have been possible without the strong support system I have behind me. This includes my stellar dissertation committee of Dr. Mary O'Brian, Dr. Kyle Miller, and Dr. Barbara Meyer. When I made the decision to leave the classroom, Dr. O'Brian hired me in my first education consulting job, which set me on my current career path and sparked my interest in assessment. Dr. Miller taught me everything I know about qualitative research techniques and empowered me to take on a study like this. And I cannot say enough about the impact Dr. Meyer has had on my life. After first serving as a professor during my undergraduate experience, Dr. Meyer remained a career and educational advisor to me over the next 10 years until she accepted the role as the best dissertation chair I could possibly ask for. Additionally, I'm incredibly thankful for the support of my parents who taught me the value of education and hard work, my whole family who have loved me unconditionally, my coworkers who put up with me constantly talking about my studies, my friends who were always there to grab a cold drink, and my soon-to-be wife Liliana for loving me and cheering me on every step of the way.

D. M. F.

## CONTENTS

	Page
ACKNOWLEDGMENTS	i
CONTENTS	ii
TABLES	viii
FIGURES	ix
CHAPTER I: INTRODUCTION	1
Assessment Literacy	2
Statement of the Problem	4
Purpose of the Study	5
Research Questions	6
Significance of the Study	6
CHAPTER II: REVIEW OF THE LITERATURE	8
Measuring Assessment Literacy	8
Practicing Teachers' Assessment Literacy	10
Preservice Teachers' Assessment Literacy	14
Assessment Education	17
Priorities	18
Curriculum	19
Theory vs. Practice	21
Formative Assessment	23
Strategies	25
Growth and Future Development	27

Summary	28
<b>CHAPTER III: THEORETICAL FRAMEWORK</b>	<b>30</b>
Existing Theoretical Frameworks on Assessment Education	32
The Framework of Assessment Education for Secondary Teachers	36
What Should Teachers Know?	38
Alignment to learning goals	41
Diverse learners	42
Feedback and motivation	43
Formative and summative	45
Language and literacy	46
Results to guide instruction	47
Statistical literacy	48
Student self-assessment	50
Tools and types	51
Where Should it be Learned?	51
How Should it be Learned?	53
Summary	54
<b>CHAPTER IV: RESEARCH DESIGN AND METHODOLOGY</b>	<b>56</b>
Research Questions	56
Setting	56
Instrumentation	58
Document Analysis	59
edTPA Data Analysis	60



Overall and Sample Populations	61
Methodology	63
Procedures	65
Categorical concurrence	65
Depth of knowledge	66
Range of knowledge	67
Coding process	68
Organization of codes	70
Member-checking	70
Quantitative data	71
Reliability and Validity	72
Researcher Bias	73
Ethical Considerations	73
CHAPTER V: RESULTS	74
Document Analysis Coding	74
Biology	81
Chemistry	82
Communications	84
Physics	87
Theater	89
Curriculum and Instruction	90
InTASC Standard 6	93
edTPA Science Assessment Handbook	94

edTPA Data	95
CHAPTER VI: DISCUSSION	100
Restatement of the Problem	100
Review of the Methodology	100
Study Discussion	101
Research Question 1: Where is Assessment Being Taught to Undergraduate Secondary Education Preservice Teachers?	102
Research Question 2: How Closely Does Assessment Instruction Align with Assessment Practices as Defined by InTASC and the edTPA?	104
Categorical concurrence	104
Depth of knowledge	107
Range of knowledge	107
Alignment matrix	108
Research Question 3: How Do Secondary Education Preservice Teachers Perform in the Realm of Student Assessment as Measured by the edTPA?	109
Cross-Question Discussion	112
CHAPTER VII: CONCLUSIONS	117
Implications for Practice	118
Limitations	119
Recommendations for Further Research	120
Final Thoughts	122

REFERENCES	124
APPENDIX A: ILLINOIS PROFESSIONAL TEACHING STANDARDS ASSESSMENT STANDARD AND INDICATORS	147
APPENDIX B: INTASC ASSESSMENT STANDARDS AND INDICATORS	149
APPENDIX C: THE FRAMEWORK FOR TEACHING (DANIELSON FRAMEWORK)	151
APPENDIX D: EDTPA SCIENCE RUBRIC 5: PLANNING ASSESSMENTS TO MONITOR AND SUPPORT STUDENT LEARNING	154
APPENDIX E: EDTPA SCIENCE RUBRIC 10: ANALYZING TEACHING EFFECTIVENESS	155
APPENDIX F: EDTPA SCIENCE RUBRIC 11: ANALYSIS OF LEARNING	156
APPENDIX G: EDTPA SCIENCE RUBRIC 12: PROVIDING FEEDBACK TO GUIDE LEARNING	157
APPENDIX H: EDTPA SCIENCE RUBRIC 13: STUDENT UNDERSTANDING AND USE OF FEEDBACK	158
APPENDIX I: EDTPA SCIENCE RUBRIC 14: ANALYZING STUDENTS’ LANGUAGE USE AND SCIENCE LEARNING	159
APPENDIX J: EDTPA SCIENCE RUBRIC 15: USING ASSESSMENT TO INFORM INSTRUCTION	160
APPENDIX K: MEMBER-CHECKING RECRUITMENT LETTER	161
APPENDIX L: INFORMED CONSENT FORM	162
APPENDIX M: WORKING CODEBOOK: CATEGORICAL CONCURRENCE	164
APPENDIX N: WORKING CODEBOOK: DEPTH OF KNOWLEDGE	168
APPENDIX O: WORKING CODEBOOK: RANGE OF KNOWLEDGE	170



## TABLES

Table		Page
1.	The Framework of Assessment Education for Secondary Teachers: Content Components	38
2.	Depth of Knowledge Levels	66
3.	Crosswalk from Webb’s DOK Levels to DeLuca and Bellara’s DOK Levels	67
4.	Range of Knowledge Levels	68
5.	Frequency Table of Codes Across Three Major Coding Sources (Expressed as Percentages)	77
6.	Frequency Table of Codes Across Each Individual Syllabus (Expressed as Percentages)	78
7.	Alignment Matrix Based on Code Frequency (Expressed as Percentages)	79
8.	Document Analysis Findings for BIO	82
9.	Document Analysis Findings for CHE	84
10.	Document Analysis Findings for COM	86
11.	Document Analysis Findings for PHY	88
12.	Document Analysis Findings for THT	90
13.	Document Analysis Findings for CI	92
14.	Document Analysis Findings for InTASC Standard 6	93
15.	Document Analysis Findings for edTPA	95
16.	edTPA Rubric and Content Being Assessed	96
17.	Descriptive Statistics for the edTPA Assessment-Related Rubrics	97
18.	Descriptive Statistics for the edTPA Tasks	99

## FIGURES

Figure	Page
1. Teaching Assessment Literacy in Practice (TALiP) by Xu and Brown (2016). A conceptual framework for teaching assessment literacy.	33
2. Theory of Assessment Literacy by Gottheiner and Siegel (2012). A visualization of teachers' understanding of assessment.	34
3. Framework for Professional Statistical Literacy by Pierce and Chick (2013). A visualization of the statistical knowledge needed to teach.	34
4. The Framework of Assessment Education for Secondary Teachers	37

## CHAPTER I: INTRODUCTION

Student assessment is an essential part of the educational process, yet it serves as one of the most misunderstood and misused aspects in schools. Many researchers have written of the need for assessment knowledge in the development of a competent teacher (Abell & Siegel, 2011; Brookhart, 2011; Darling-Hammond & Bransford, 2005; Elliott, 2010; Mandinach & Gummer, 2013; Popham, 2009) and university teacher education programs serve as the place to start this development. This is not a new phenomenon as, even prior to the last decade, research has been indicating there is a strong concern surrounding the assessment education of preservice teachers and whether it is improving (Brookhart, 2001; Crooks, 1988; Gullickson & Ellwein, 1985; Impara, Plake, & Fager, 1993; Lomax, 1996; Maclellan, 2004; Marso & Pigge, 1993; McMillan, 2001b, McMillan, Myran, & Workman, 2002; Mertler, 2004; Plake & Impara, 1997, Schafer & Lizzitz, 1987; Shulha, 1999). Though assessment education does hold a place within many teacher education programs, many argue that it is either not enough or not effective.

Past reviews of the literature consistently show teachers' assessment literacy to be less than desired (Brookhart, 2001; Campbell, 2013; Crooks, 1988; Marso & Pigge, 1993; Opre, 2015). In the age of teacher, school, and district accountability, this is a major problem. As Popham (2011) said, "...for teachers to remain ignorant about the nature of educational assessments these days surely constitutes the quintessence of professional suicide. If not suicidal, it is at least sublimely stupid. Teachers, and there are no exceptions, need to understand the nature of the instruments being employed to judge them as professionals" (p. 269).

Though grading, sorting, and ranking students is required in most schools, the power of assessment has been enhanced with new research in the past 20 years. Black and Wiliam's (1998) landmark article *Assessment and Classroom Learning* brought an incredible amount of

attention to the ideas of formative assessment, or assessment *for* learning instead of the traditional assessment *of* learning. They found that summative test scores go up when formative assessments are used throughout the learning process. Formative practices would include providing students with useable feedback on their learning and using assessment data to make instructional decisions that will benefit the students. There is concern over whether these formative practices are making their way into teacher education programs, however. As Mandinach and Gummer (2011) said, “The field of data-driven decision making falls between the cracks of traditional courses. It is not introductory statistics or measurement; it is not instruction, pedagogy, or methods. Data-driven decision making lies at the intersection of those courses” (p. 34).

Teaching assessment is a difficult task. In addition to determining methods for assessment education and finding the correct location within the education program, it is critical to determine exactly what skills and knowledge teacher candidates must possess in order to be effective assessors when they reach the classroom. This alone is a daunting endeavor and a troublesome task for professors and program organizers who, themselves, might not have had ample instruction on assessing students and using these data to inform instruction.

### **Assessment Literacy**

The term “assessment literacy” typically refers to the basic understanding of assessment practices and the ability to apply these skills to collect evidence of student achievement (Stiggins, 1991). Popham (2011) described the term by saying, “Assessment literacy consists of an individual’s *understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions* [emphasis in original]” (p. 267). He wrote of the tendency for some educators to equate assessment literacy with things like standard deviations,



correlation coefficients, and difficulty indices. He stressed that this is not the point of assessment literacy at all. Instead, assessment literacy requires educators to only understand the concepts that are “likely to influence educational decisions” (p. 268). For the typical teacher, this is the information they need to be successful. Willis, Adie, and Klenowski (2013) conceptualized assessment literacy in this way:

Assessment literacy is a dynamic context dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students (p. 2).

They argue that it is necessary to include this sociocultural piece to fully understand the concept. Conceptualizing it this way is important when creating systems for preservice teachers to improve.

Assessment literacy is discussed heavily by Lyon (2013). After an initial discussion, he began utilizing a different, yet related, term he called “assessment expertise” within his study. He used this term to “capture teacher *growth* toward more sophisticated and responsive assessment practices [emphasis in original]” (p. 443). He then dug even deeper by using terms like “assessment understanding” and “assessment facility” (p. 444). Defining the terms and understanding their importance was central to his discussion. Pierce and Chick (2011; 2013; 2014) frequently explored teachers’ “statistical literacy,” which is the ability to interpret statistical information and evaluate their significance. There is an increasing need for this skill in interpreting the large-scale assessment data that are common in schools today. Likewise, many favor the term “data literacy” to describe educators’ ability to interpret and use various pieces of

data from various sources to make decisions (Data Quality Campaign, 2014; Mandinach, 2012; Mandinach & Gummer, 2011; Mandinach & Gummer, 2013; Mandinach & Gummer, 2016).

Engelsen and Smith (2014) stressed that, for a successful assessment culture, everyone must be considered assessment literate in a school, including teachers, students, and administrators. For the purpose of this study, assessment literacy will encompass the general understandings of *why to assess, how to assess, evaluating assessment results, and understanding the purpose* of the assessments being used.

### **Statement of the Problem**

There is a large body of research showing teacher assessment literacy to be at a low level (as will be discussed more in chapter 2). Teachers are not capable of fully utilizing the power of assessment in their classrooms. This includes designing and administering formative assessments, designing efficient summative assessments that capture the right information, and analyzing assessment data that they receive from their own assessments or standardized assessments.

The difficulties of teaching these skills to education students are also well-documented in the research. Assessment looks different throughout each individual subject, making it a difficult subject to teach using a blanket approach. Data collection could take the form of a test, an essay, a project, a speech, or even a simple discussion. Though the purpose is to check and evaluate the progress of a student, the methods are varied.

Often, evaluations of preservice teachers' classroom readiness show assessment literacy to be one of the most troublesome areas for future educators. Both qualitative and quantitative measures paint a disturbing picture of assessment literacy. This results in the continuous churning out of teachers who struggle to interpret assessment data correctly and use assessments

to further student learning. In an age of increased importance placed on accountability data, this poses a big problem.

### **Purpose of the Study**

The purpose of this study grew from the need to identify the biggest problems in assessment education and to compare them to the recommendations for best practice. These recommendations are pulled from established teaching standards like those created by the Interstate New Teacher Assessment and Support Consortium (InTASC) (Council of Chief State School Officers, 2011) and established teacher evaluation tools like the framework created by Charlotte Danielson (2007).

Many states are now using the edTPA (Teacher Performance Assessment), a nationally-available, performance-based assessment, as a final step toward teacher licensure. Preservice teachers submit their edTPA portfolio during their final semester of college and are assessed on their abilities in three areas:

1. Planning for Instruction and Assessment
2. Instructing and Engaging Students in Learning
3. Assessing Student Learning

The emphasis the edTPA puts on assessment provides a worthy measure of the skills of preservice teachers who are approaching the end of their program. It is aligned to the InTASC standards and the Danielson Framework (Stanford Center for Assessment, Learning, & Equity, 2016).

After defining best practice and assessing preservice teachers on their ability, the next step would be to identify what they are specifically being taught and where any gaps in learning might exist. Most universities require the existence of syllabi to overview the content of each

course within their programs. An analysis of these documents will provide a glimpse into the instruction students receive.

One specific university was chosen for this study and its secondary education program and students will be examined. The hope is that many of the findings can be generalized to the larger population and considered when making curriculum decisions.

### **Research Questions**

To guide the study, the following research questions will be used:

1. Where is assessment being taught to undergraduate secondary education preservice teachers?
2. How closely does assessment instruction align with assessment practices as defined by InTASC and the edTPA?
3. How do secondary education preservice teachers perform in the realm of student assessment as measured by the edTPA?

The first question analyzes the approach to teaching assessment skills at the specific university used in this study. The second question considers the university, the standards, and the licensure examination in conjunction with each other and identifies the coverage and gaps of assessment instruction. The third question serves to assess the assessment literacy of the preservice teachers at the university.

### **Significance of the Study**

It is very common for studies to display the low levels of assessment literacy among preservice and inservice teachers, but it is not as common for them to dig into the specifics of where and why assessment literacy is low (Howley, Howley, Henning, Gillam, & Weade, 2013). This study will attempt to do just that. By defining best practice and comparing it to reality, this

study will add to the discussion of the specific strengths and weaknesses of assessment instruction for the secondary education level.

Through this, the conversation can shift from “there is a problem” to “this is how to address the problem.” This is the step toward seeing results in this field. As Black and Wiliam’s (1998) research shows, student performance improves with the proper use of formative assessment practices. And, as Popham (2011) pointed out, student assessment is currently being used for teacher evaluation purposes, so not being literate in this area would be a step toward professional suicide. Specifically identifying the causes of the issue is the first necessary stage in developing an improvement plan.

## CHAPTER II: REVIEW OF THE LITERATURE

The literature on this topic takes many different approaches to the issue. The common thread that links them all is that teachers should possess assessment skills that can be used when educating students. This literature review will overview the tools that are used to measure assessment literacy, the assessment skills of practicing teachers, the assessment skills of preservice teachers, and, finally, a deep look at how assessment is being taught in university education programs.

### **Measuring Assessment Literacy**

Many individuals have developed tools to measure the level of assessment literacy certain educators might possess. Plake and Impara (1992) created the *Teacher Assessment Literacy Questionnaire* (TALQ) that is still being used by many researchers today to measure assessment literacy. They have used it several times to show issues in teachers' understanding of assessment (Impara et al., 1993; Plake & Impara, 1997). Alkharusi (2011) studied its psychometric properties and determined the tool to be valid and reliable and useful for its intended purpose. Findings from this survey often showed the overall assessment literacy scores to be lower than desired. The highest scores tended to be in the area of "administering, scoring, and interpreting the results of assessment," and the lowest in "communicating assessment results" (Campbell, 2013).

Using a different tool, teachers' conceptions of assessment were studied extensively by Brown and his colleagues in various countries (Brown, 2004; 2006; 2007; 2008a; 2008b; 2009; Brown, Chaundhry, & Dhamija, 2015; Brown & Harris, 2006; Brown, Kennedy, Fok, Chan, & Yu, 2009; Brown & Lake, 2006; Brown & Lingbiao, 2015; Brown & Michaelides, 2011; Brown & Remesal, 2012). They used a tool called *Conceptions of Assessment* (COA) that measures four

major conceptions of assessment: improvement, school accountability, student accountability, and irrelevance. When exploring the diverse countries, they found that the countries' educational priorities certainly influence teachers' conceptions of assessment. Societal focuses like accountability or formative assessment tend to result in these things being focuses for the teachers as well.

Barnes, Fives, and Dacey (2015; 2017) brought the COA instrument to the U.S. and developed it further with the *Conceptions of Assessment Continuum*. They fleshed out the relationships between each conception and placed them on a scale that can be used to more specifically identify teacher beliefs and understandings. This continuum extends from "improvement," through "student accountability" and "school accountability," and then to "irrelevance." Their studies found that teachers fall into all of these domains but they also can have conflicting beliefs about assessment.

Using many of the same indicators, Mertler and Campbell (2005) developed the Assessment Literacy Inventory (ALI) which they began using to gather information on preservice teachers. Hailaya, Alagumalai, & Ben (2014) analyzed the psychometric qualities of this tool and found it has some limitations, but it can still be used to provide useful data.

Wayman, Wilkerson, Cho, Mandinach, and Supovitz (2016) explained the Teacher Data Use Survey that can be used to study the data literacy of teachers, administrators, and support staff. This can be used by school districts to learn how their educators use data, perceive data use, and to better support their educators with the necessary resources that exist to improve their data use for instructional decision-making.

Though all of these tools have been used to measure assessment literacy, there are some who question their accuracy. Gotch & French (2014) reviewed teacher assessment literacy

measures from 1991 to 2012. They collected various rubrics and objective tests of assessment knowledge to review the psychometric qualities of these measures. Across the 36 measures they reviewed, they found that the support for the measures' claims was weak and that the measures might not necessarily be accurately measuring assessment literacy.

All of these tools were designed to capture the strengths and weaknesses in the realm of assessing student performance. Though they all attempt to measure similar skills, they are not all aligned to the same standards of assessment literacy. It would be interesting for a future study to apply each tool to the same situation to highlight the similarities and differences of the findings. This would aid in creating a clearer picture of teachers' assessment literacy.

### **Practicing Teachers' Assessment Literacy**

Practicing teachers are expected to embed assessment into their teaching process and they are often evaluated based on their ability to do this. Howley et al. (2013) discussed how most literature focuses on "teachers' assessment knowledge and practices in an effort to gauge how far removed they are from what assessment experts prescribe" (p. 27). It is very rare that literature will describe teachers as being fully assessment literate. Before examining the way teachers learn (or don't learn) to be assessment literate, it is important to explore the abilities they currently do (or don't) possess.

Gunn and Gilmore (2014) found that preservice early childhood teachers in New Zealand have strong conceptions of assessment based on their own educational experiences. This usually results in them focusing more on evaluating students than assessment *for* learning. Buyukkarci (2014) saw the same findings in primary language teachers in Turkey, who did not apply many formative assessment techniques even though they held positive beliefs of them.



The push for more formative assessment has led to other studies, as well. Yao (2015) held a focus group interview with a group of teachers about their perceptions of classroom assessment and their uses of formative assessment. The majority of those interviewed spoke enthusiastically about the notion of using assessment *for* learning, but when discussing the challenges of assessment, they typically reverted back to summative assessment *of* learning mindsets.

Frey and Schmitt (2010) studied classroom assessment practices of 3<sup>rd</sup> through 12<sup>th</sup> grade teachers in a Midwestern U.S. state. They found that formative assessment is not commonly used by teachers and that 3 out of every 4 assessments are given after learning has occurred. They also found that traditional paper-and-pencil testing is the most popular form of classroom assessment and that teachers routinely rely on assessments that have been written by others, such as textbook companies.

In another study that paints a negative picture of practicing teachers' assessment literacy, DinanThompson and Penney (2015) used a qualitative study to examine the assessment practices of primary physical education teachers in Australia. They used surveys, interviews, and document analyses to try to better understand how 18 teachers were using assessment in their PE classrooms. They found that assessment was embedded into the class, but most of it was superficial. The students were not fully engaged in the assessment and the data collected were not a true reflection of student learning. They also were able to highlight many gaps in teacher knowledge about assessment use, including a lack of understanding of formative assessment practices.

Allal (2013) showcased a study that displayed the more nuanced ways in which teachers use assessment. He studied Swiss 6<sup>th</sup> grade teachers by interviewing them at two points

throughout the year and analyzing assessment documents that they used to determine student grades. The purpose of the study was to better understand the role professional judgement plays in awarding final grades. Findings indicated teachers use professional judgement as both an individual cognitive act and as a socially situated practice. They used assessments to gather reliable evidence of student abilities, but they also catered to individual student needs. Wyatt-Smith, Klenowski, and Gunn (2010) showed many of the same findings in a study of Australian teachers. These studies provide a more positive view of the teacher assessment use and the intricacies of their assessment knowledge.

Continuing the conversation on the culture of assessment, Howley, et al. (2013) interviewed teachers from three different high schools to better understand how they use assessment and the state of assessment culture within their schools. They found that the school environments and requirements affected teacher perceptions of assessment. Teachers tended to use the same terminology to describe assessment practices, but the actual execution of these ideas was different from school to school. They also found that most teachers regarded themselves as understanding assessment more than other stakeholders, especially those who are not in the classroom daily.

Pierce and Chick (2011; 2013) conducted multiple studies on the statistical literacy of teachers, specifically focusing on their ability to interpret various forms of student data reports. They found a wide range of abilities among the teachers observed. Frequently, graphical representations like box plots were interpreted incorrectly or completely misunderstood by the teachers. Considering the amount of data that are used in the education system, this finding is troublesome.

Likewise, in a report prepared for the U.S. Department of Education, Means, Chen, DeBarger, and Padilla (2011) studied teachers' statistical literacy and found that it was lacking. They wrote, "The most difficult data literacy concepts and skills appeared to be reasoning about data when multiple calculations were required, interpreting a contingency table, distinguishing a histogram from a bar graph, and recognizing differences between longitudinal and cross-sectional data" (p. 61). They also wrote of teachers' inability to use these data to make instructional decisions in the classroom. Phanchalaem, Sujiva, & Tangdhanakanond (2016) discussed this same finding in their study of teachers in Thailand.

Many local, state, and national laws and policies result in a large amount of data being collected, so the inability of teachers to use these data is problematic. It raises the question of why the system collects the data at all. Pierce, Chick, and Gordon (2013) found that, in both primary and secondary schools in Australia, male teachers were more positive about using national assessment data than female teachers were. They also found that teachers who had studied statistics in college were more comfortable using the data. Leighton, Gokier, Cor, & Heffernan (2010) found that secondary teachers believe that classroom assessments provide much more value to their teaching and student learning than large-scale tests. Kiomrs, Abdolmehdi, and Rashidi (2011) found that, regardless of their level of assessment literacy, teachers tended to tailor their teaching to the demands of standardized tests.

The investment in large-scale standardized assessments makes the need for assessment literate teachers even more important. Zhang and Burry-Stock (2003) found that, regardless of how long they have been teaching, inservice teachers with assessment and measurement training self-report assessment skills being higher than those without the training. This corresponds with Pierce et al. (2013). Quilter and Gallini (2000) surveyed inservice teachers and found that their

attitudes toward assessment are related to their knowledge about assessment. If teachers do not understand how to use assessments correctly, they will have a more negative attitude toward using them.

Adding to this negative attitude, Howley et al. (2013) highlighted teachers' beliefs that administrators, parents, and students are not assessment literate, which causes problems for the teachers. These groups usually put much more weight on the ideas of summative assessment and do not value formative assessments in the way that experts say they should. This puts pressure on teachers to emphasize summative assessments as well.

The literature identifies a lack of teacher emphasis on formative assessment, which is troublesome based on the earlier findings of Black and Wiliam (1998) indicating the importance of formative assessment in improving student performance (Shepard, Hammerness, Darling-Hammond, & Rust, 2005; Stiggins, 2004).

The literature also highlights the impact of teacher confidence in assessment, which research shows is improved by assessment training (DeLuca, Chavez, & Cao, 2013; Hill, Gunn, Cowie, Smith, & Gilmore, 2014). A strong confidence level in assessment can result in more accurate results and better usage for impacting student learning.

### **Preservice Teachers' Assessment Literacy**

Many researchers question whether teachers are prepared to assess when they first enter the field. How much of an impact does personal experience have on assessment literacy as opposed to preservice education? Stobaugh, Tassell, and Norman (2010) examined one university's Renaissance Teacher Work Sample (TWS) scores to better understand their preservice teachers' assessment knowledge. The TWS is a tool that measures the growth of teacher candidates. They found that their preservice teachers generally scored low in the areas of

assessment and analysis. Mertler (2004) used parts of Plake and Impara's (1992) TALQ tool and found that inservice secondary teachers generally score higher than preservice secondary teachers on assessment literacy.

Alkharusi, Kazem, and Al-Musawai (2011) formed a quantitative study to compare the assessment knowledge, skills, and attitudes of a group of preservice teachers and a group of inservice teacher from Oman. When comparing the results of a questionnaire, they found that inservice teachers showed a lower level of knowledge, a higher level of skillfulness, and a more favorable attitude toward assessment than preservice teachers. They credited the low level of inservice teachers' assessment knowledge to the fact that they have been out of coursework for a while and recommended continuous professional development to strengthen their knowledge. The more favorable attitude toward assessment shown by inservice teachers is interesting and somewhat contradictory to the findings of Quilter and Gallini (2000) and Howley et al. (2013) discussed earlier which found that teachers did not always have a positive view of assessment practices and formative practices in particular.

Armstrong (2011) conducted a quantitative study comparing the assessment knowledge and practices of Belizean primary school teachers who have at least an associate's degree in primary education to those who do not hold such a degree. She found that the level of understanding was significantly higher for the teachers with the degree. This corresponds with the findings of Zhang and Burry-Stock (2003) discussed earlier in which those who experienced the training self-reported higher assessment literacy levels. However, Armstrong did note that no differences were found in teachers' practices or their use of data to make instructional decisions.

Siegel and Wissehr (2011) studied preservice science teachers enrolled in an assessment-centered methods course. Their key finding was that the preservice teachers left the class with a

good theoretical understanding of assessment practices, yet this often did not translate to their practical experiences of assessing students. This aligns with the findings of Armstrong (2011). Siegel and Wissehr concluded, “The results of this study indicate that preservice teachers think quite differently about assessment, even when only planning lessons but not yet teaching, than they do when reflecting on assessment” (p. 388). Wang, Kao, and Lin (2010) also had similar findings when studying Taiwanese preservice elementary teachers. They generally understood a traditional view of assessment and learning but their ability to perform was not well developed.

Lorente-Catalán and Kirk (2016) found slightly different results when studying preservice teachers in England. Their qualitative study found that most preservice teachers understood the ideas of formative assessment, not just as a theoretical framework but also in practice. These preservice teachers also acknowledged that they needed to still learn more about the concepts before fully grasping them.

Dayal and Lingam (2015) qualitatively studied the assessment views of preservice and inservice teachers in Fiji. They found that preservice teachers tended to think of assessment more summatively (*assessment of learning*) while inservice teachers tended to think of assessment more formatively (*assessment for learning*). In another study, Volante and Fazio (2007) looked at preservice teachers at a Canadian university. They surveyed these teachers every year on their assessment literacy and found that it was consistently low all the way through the program. They also found that when preservice teachers discuss assessments, most typically discuss summative measures and very few discuss formative measures. This aligns with the Dayal and Lingam’s (2015) findings.

Talanquer, Bolger, and Tomanek (2015) qualitatively analyzed the grading practices of preservice secondary teachers, specifically on written student work. They found that preservice

teachers tended to focus more on the description and basics of the assessment and less on a full understanding of student ideas. This limited understanding corresponds with other findings such as those by Volante and Fazio (2007) and Talanquer et al. (2015).

Opposing many of the other studies, Levy-Vered and Alhija (2015) studied Israeli beginning teachers and found a moderate level of assessment literacy. They indicated that teachers are entering the classroom with a moderate ability to collect information of their students' learning. They also found that training in assessment both directly and indirectly affects assessment self-efficacy.

In summary, most preservice teachers do not have the knowledge of assessment that provides them with the foundations for good assessment practice in their classrooms. In the next section, literature on the teacher preparation programs relating to assessment education is reviewed.

### **Assessment Education**

Teacher education programs often look for ways to increase their assessment education in light of studies that show assessment literacy to be at a low level. DeLuca (2012) discussed the lack of studies that focus on ways to improve assessment education at the university level. He wrote of the tendency of studies to focus on whether programs and courses are effective by measuring preservice teachers' assessment literacy.

Though DeLuca (2012) does have a point about the body of research, there is not a complete dearth of studies in this area. Several researchers have discussed their findings and proposed solutions to the problem. These findings and proposed solutions will be discussed in the following sections. It is recommended that much more is done in this realm, however.

## Priorities

Greenberg and Walsh (2012) reviewed the literature on assessment education and formulated three domains of knowledge that teacher candidates need to have in order to be prepared to assess:

1. “How to measure student performance using assessments: ‘Assessment Literacy’
2. How to analyze student performance data from such assessments: ‘Analytical Skills’
3. How to use student performance data from assessments to plan instruction: ‘Instructional Decision Making’” (p. 7).

They then used these three domains to examine the assessment instruction of 180 teacher preparation programs. They found that only 21% of the programs they studied adequately cover “Assessment Literacy,” less than 1% adequately cover “Analytical Skills,” and less than 2% adequately cover “Instructional Decision Making.” The researchers also found that most programs tended to at least expose students to the concept of formative assessment, and that 58% of secondary programs either did not have a subject-specific methods course or they did have one but it did not address assessment.

DeLuca and Bellara (2013) examined the assessment education priorities of three different realms: teacher education university syllabi, teacher education policy documents, and standards for teacher practice. They analyzed course syllabi from 10 Florida teacher education programs that were certified by the National Council for Accreditation of Teacher Education (NCATE), policy documents like the NCATE Unit Standards and the Florida Department of Education Accomplished Practices, and various standards that exist for teacher practice in student assessment and evaluation. They found strong alignment in points like “assessment processes,” “measurement theory,” and “assessment fairness.” They also identified points of



misalignment in areas like “assessment for learning,” “assessment purposes,” “classroom environment and assessment,” and “communication of assessment results.”

In another study that focused on the priorities of assessment instruction, Campbell and Collins (2007) looked at the five top-selling assessment textbooks on the market in both general and special education to identify what leading textbook publishers regarded as important points to teach. They grouped the topics into 13 categories but found that sometimes the general education textbooks and the special education textbooks disagreed on what topics were important. The authors concluded, “we do not believe they sufficiently represent the broader range of skills needed by all teachers to make data-based decisions about their students and instruction” (p. 17).

Mandinach, Gummer, and Friedman (2013) surveyed 208 United States teacher preparation programs and found that 62.4% of them indicated that they offer at least one course that is primarily focused on teaching data use to inform instructional decisions. Additionally, 92% said that data use is integrated into at least one course that is offered. These claims would indicate that these programs heavily value data-driven decision-making as an important part of teacher education.

Other individuals have been vocal about assessment literacy in very specific areas. For example, Inbar-Lourie (2008; 2013) called for more work in language assessment literacy, the knowledge base that teachers hold around the assessment of student language skills. To navigate all these needs, the research on the most beneficial and impactful factors must be considered.

## **Curriculum**

Several publications have presented assessment curricula that teacher education programs have developed (DeLuca, Klingler, Searle, & Shulha, 2010; Fan, Wang, & Wang, 2011; Gareis &

Grant, 2015; Hill, Ell, Grudnoff, & Limbrick, 2014; Jie-Qi & McNamee, 2006; Jones, 2014; Lukin, Bandalos, Eckhout, & Mickelson, 2004; Mayor, 2005; Munroe, Foran, MacLeod, Graham, Lunney-Borden, & Curry, 2012; Schaffer, 2014). DeLuca, Chavez, Bellara, and Cao (2013) surveyed preservice teachers to identify the pedagogies that made a big difference in their assessment education. They identified four pedagogical constructs in this study:

1. Perspective-building conversations
2. Praxis activities
3. Modeling
4. Critical reflection and planning for learning.

Creating a solid curriculum with focused priorities is crucial, but it is still necessary to consider the research on whether assessment courses can make a difference. DeLuca et al. (2013) surveyed teacher candidates and found that they increased their understanding of assessment throughout the methods course in which they were enrolled. They moved from basic assessment knowledge to a better understanding of multiple forms and purposes of assessments. They also increased their confidence in assessments throughout the course. Hill et al. (2014) reported the same findings in their study of several New Zealand teacher education programs.

Conversely, Grainger and Adie (2014) surveyed secondary preservice teachers in an education program at an Australian university after they had taken their assessment course. They concluded that this single assessment course was not enough to prepare the preservice teachers for their future as assessors. The preservice teachers had difficulty with consistency in grading and using rubrics to grade and provide feedback. They also expressed difficulty in assessing in a standards-referenced assessment system.

Jones (2014) pointed out that assessment courses are often included toward the end of an education program, which is problematic because it becomes an aspect that is simply tacked on to their already-learned perception of the teaching process. She studied a year-long program for foreign language preservice teachers that embeds assessment instruction throughout. She found that this strategy allowed the preservice teachers to take their assessments beyond what others were capable of, specifically in the realm of formative assessments used for informing future instruction.

Examining the issue even further, Wallace and White (2015) examined the specific stages through which preservice teachers passed when learning how to assess. They formed a study around preservice mathematics teachers from three different California teaching programs. The preservice teachers were interviewed at various times on assessment practice throughout their apprenticeship while the researchers also collected numerous artifacts and other pieces of data as well. The key finding was that the preservice teachers tended to evolve through three general stages of assessment knowledge: the test-oriented stage, the task-oriented stage, and the tool-oriented stage. They also found that the preservice teachers typically focus on *how to assess* before they learn *why to assess*. Similarly, Mooi, Periasamy, Ming, and Osman (2014) examined students at a Malaysian university and suggested that fundamental assessment knowledge should be taught to preservice teachers before they are exposed to specifics of assessment in a specified area such as special education.

### **Theory vs. Practice**

Many studies have shown that learning assessment in theory and putting it into practice are two very separate things. DeLuca and Lam (2014) surveyed seniors in the school of education at their university. The university offered an assessment course that all education

majors were required to take, typically in their final year. From the surveys, they found that preservice teachers could articulate many of the concepts of effective assessment, but most did not list concrete examples or have a plan for how to assess. This suggests that preservice teachers are taught assessment in theory, but they still struggle to put it into practice.

Lian, Yew, and Meng (2014) also spoke of the need for more practice experiences in assessment instead of just theoretical lessons. They presented five skill targets on which teacher education programs should focus: validity of assessment, reliability of assessment, transparency of assessment, fairness of assessment, and using assessment information.

Siegel (2013) studied the growth of preservice science teachers and their ability to assess linguistically diverse populations. She indicated that participants grew from simply viewing equitable assessments as fair to more complex understandings, like using the assessments for learning and the importance of challenging students. Like many other studies, though, she indicated that the participants' understanding did not always line up with their assessment plans. There was a gap between theoretical and practical understandings. This concurs with the earlier study by Siegel and Wissehr (2011).

Kelting-Gibson, Karsted, and Weikert (2013) also discussed the tendency for assessment instruction to be based more on direct instruction than on experiences. The programs adequately teach assessment, but they fail to make the connections to practice by providing the necessary experience to the undergrads. Teachers are entering their first year without the ability to directly apply proper assessment strategies. To drive home their point and emphasize a better way of teaching, the researchers studied preservice teachers as they learned about assessment during one of their courses. These individuals accompanied a group of children to a museum and were given the task of assessing them on a variety of topics. The reflections with the preservice teachers

afterwards showed this to be a valuable experience. They were able to learn the intricacies of the assessment process much better through this interactive experience.

Though many studies agree with this need for more experience-based learning, DeLuca and Klinger (2010) studied teacher candidates in an education program in Canada and concluded that direct instruction on assessment is necessary to become a competent educator. These studies combined show the importance of both instruction and practical experiences. Neither method is sufficient by itself.

Lee and Yoon (2008) described a science methods course at the University of Texas at Arlington that teaches early childhood teachers how to assess student inquiry skills. The first hour of class each week is based around lecture and discussion but the second hour is a laboratory session where preservice teachers can explore and experiment with rubrics and other assessment tools.

Whether they are experiencing direct instruction, practical experiences, or a combination of the two, Cornish and Jenkins (2012) promoted more reflection of assessment practices to guide preservice teachers through their individual learning. Preservice teachers should explicitly examine their methods to maximize their understanding of the skill. They need to be conscious of every action and decision made during the assessment process. Likewise, Graham (2005) wrote that teacher candidates reported being strongly influenced by the professional dialogue both in their classes and during their field experiences, so a supportive, professional environment for this should be created.

### **Formative Assessment**

After Black and Wiliam's (1998) landmark article, many researchers have covered the learning of formative assessment techniques. Buck, Trauth-Nare, and Kaftan (2010) looked at

preservice science teachers in a methods course and examined their understanding of formative assessment methods. They found that the course did substantially improve their understanding of formative assessment, but they joined the chorus of many other researchers by calling for more field experiences and ongoing reflection to solidify the understanding. Jones (2014) also saw formative assessment knowledge increase in her study of a year-long course with assessment embedded throughout (instead of simply tacked on at the end, which tends to be common in teacher education courses).

Even though there has been a movement in the education world toward formative assessment practices, many universities have not made the necessary changes to keep up. Poth (2013) analyzed 57 syllabi from 14 Canadian teacher education programs and discussed how the shift toward formative assessment techniques has fundamentally changed the way teachers use assessment, yet the teacher education programs have not completely adjusted to the shift. She recommended three major changes: a realignment between what is taught in assessment courses with the roles and responsibilities of beginning teachers, more modelling of assessment techniques by instructors, and more inservice professional development once teachers are in the field.

To form a deeper understanding of the issue, Smith and Galvin (2014) studied preservice primary teachers to learn how their beliefs about assessment change during their three-year education program. They found that, during the first two years, preservice teachers thought of assessment in only a summative fashion and focused mostly on sort and rank strategies. By the third year, they saw assessment as being an essential part of the educational process and were able to differentiate between different types of assessments.

## **Strategies**

Though more is needed, a number of researchers have shared some strategies they use to improve the assessment education at their schools. Morrison (2005) used notebooks in her science methods courses to promote the ideas of formative assessment to the preservice teachers. Science notebooks are a research-based method where students are asked to keep a journal of their work, allowing the teacher to better assess for prior knowledge, conceptual understandings, and current learning. She found that the teachers in the study saw this method as valuable because they saw themselves being formatively assessed in their own class. They were encouraged to use them with their future students.

Adding to the literature on practical strategies that preservice teachers can use in their own future classrooms, Bennett and Cunningham (2009) formed a study where preservice teachers were provided a handheld computer with data-collection software that they could use to assess special needs students during tutoring sessions. They found that, though the data-collection software caused challenges, the participants in the study saw value in their use and the power they can bring to the formative assessment process.

As another method, Bangert and Kelting-Gibson (2006) drew on past research to promote the use of problem-based learning (PBL) to teach preservice teachers about assessment. This method applies the concepts in a more practical situation. They specifically promote the use of teacher work sample methodology (TWSM), a form of PBL that requires learners to create assessments and evaluative activities for a classroom unit. They create teacher work sample portfolios that contain various assessment artifacts that they produce for this unit of study. The researchers conducted a study of a group of undergraduate education students and found that the

majority indicated the TWSM helped them better understand the role assessment plays in the learning process.

Taking the idea out of the coursework and into the clinical experiences, Dorfman, Galluzzo, and Meisels (2006) looked at a curriculum-embedded performance assessment and interviewed preservice teachers who were exposed to this idea. They found that when student teachers were exposed to placements where the assessments were embedded in the curriculum and ongoing, their assessment literacy grew.

Additionally, Falter Thomas and Sondergeld (2015) studied a strategy that paired preservice teachers with middle school students. The teachers provided digital feedback on students' research projects while learning about scaffolded feedback in their methods course. This combination of clinical experiences and classroom instruction proved effective as they saw the teachers' ability and confidence levels improve.

Several researchers have also looked at ideas and methods for developing statistical literacy. Odom and Bell (2017) described how empirical experiences can develop hypothetical concepts in statistical literacy of preservice teachers. They discussed the need to incorporate statistics into nonstatistic classes in order to prepare teachers for the skills needed in the classroom. Jin (2010) surveyed 86 instructors of foreign language teacher education assessment courses in China and found that aspects of theory and practice were adequately taught, but statistical measurement was not covered quite as thoroughly. Pierce and Chick (2014) also formed a study showing that teachers did learn aspects of statistical literacy in a course designed for this purpose, yet they tended to not retain this information. They concluded that, for statistical literacy to fully develop, teachers must focus on these concepts more than a couple times a year.



This concurs with the previously-discussed findings of Jones (2014), who heavily promoted assessment concepts being embedded throughout a teacher education program.

Literature on effective strategies is incredibly important because it bridges the gap between the theoretical concepts of assessment education and the practical implementation of proven methods. Students must be put in an environment where their learning is maximized and research-based strategies are utilized. Their experiences must be valuable and thorough and they must be given time to reflect on their own learning.

### **Growth and Future Development**

As Falter Thomas and Sondergeld (2015) showed, continuous growth is necessary for teachers to reach their full potential as assessors. Lyon (2013) studied the growth of preservice secondary science teachers' assessment skills, especially in relation to the assessment of linguistically diverse populations. His study focused on one year of their education program. He conducted interviews and collected artifacts and concluded that they showed the most growth in the area of using assessment to support student learning. He also found that changes in their ability to consider equity and design assessments were not statistically significant.

Gunn and Gilmore (2014) also reported that the early childhood preservice teachers they studied were poised for future assessment growth when they entered the field. This is promising, but in order for it to be fulfilled, they need to be supported once they have left the university.

Some hold the belief that assessment education should mainly be the responsibility of the schools since so much of the teachers' knowledge is contextual. Many studies have focused on the best ways to professionally develop practicing teachers' assessment literacy (Adie, 2013; Bailey, Little, Rigney, Thaler, Weiderman, & Yorkovich, 2010; Dillon, Erkens, Sanna, & Savastano, 2015; Fulcher, 2012; Herrington, Herrington, & Glazer, 2002; Highland, 2015; Huai,

Braden, White, & Elliott, 2006; Koh, 2011; Lai & McNaughton, 2016; Lee & Son, 2015; Livingston & Hutchinson, 2016; Matthews, Trimble, & Gay, 2007; Mertler, 2009; Rorsyth, Cullen, Ringan, & Stubbs, 2015; Sato, Chung, & Darling-Hammond, 2008; Schneider & Randel, 2010; Sondergeld, Bell, & Leusner, 2010). It is true that, once graduated, teachers are out of the hands of the initial teacher education programs. If they are to continue their assessment education, they will need to continue their professional learning in other ways.

### **Summary**

A large body of research exists showing the assessment literacy of both inservice and preservice teachers to be below the level needed to be effective in the classroom. Many articles have theorized and studied ways of solving this problem, often focusing on the university teacher education programs. As DeLuca (2012) stated, though, the research on ways to improve assessment instruction is minimal. Most of the research is intended to identify that there is a problem. Further studies on strategies and methods for improvement are recommended.

A common feature of many articles on best practice is the need for more practical experience in assessment. Assessment education must connect theoretical and practical understandings of how to design, collect, and interpret data on student performance. This is an important point to consider in future studies and when crafting strategies for teaching assessment skills.

It is also interesting to note that studies in this field are not United States-specific. Other countries are experiencing the same issues, which provides an opportunity for learning and collaboration. Literature in this review include studies from Australia (DinanThompson & Penney, 2015; Grainger & Adie, 2014; Pierce et al., 2013; Wyatt-Smith et al., 2010), England (Lorente-Catalán & Kirk, 2016; Rorsyth et al., 2015), Switzerland (Allal, 2013), Taiwan (Wang

et al., 2010), Iran (Kiomrs et al., 2011), China (Jin, 2010), Belize, (Armstrong, 2011), Norway (Leirhaug, MacPhail, & Annerstedt, 2016), New Zealand (Gunn & Gilmore, 2014), Canada (DeLuca & Klinger, 2010; Poth, 2013; Shulha, 1999; Volante & Fazio, 2007), Turkey (Buyukkarci, 2014), Malaysia (Lian et al., 2014), Israel (Levy-Vered & Alhija, 2015), Oman (Alkharusi, 2011), Thailand (Phanchalaem et al., 2016), and Fiji (Dayal & Lingam, 2015).

There is a lack of hard data showcasing the effect low teacher assessment literacy can have on the students themselves. How does it affect their development? Collecting these data would be very difficult due to the need to isolate this factor's individual contribution to their development, but, if researchers can find a way to add to the literature, it would be greatly valued.

### CHAPTER III: THEORETICAL FRAMEWORK

As seen in the literature review, assessment education has been a focus for many researchers, but often the studies have consisted of evaluations of certain aspects of the programs. The research on how assessment is being taught or how well assessment education is aligned to standards and policy is minimal. Most of the literature that exists is focused on what teachers should know and measures of whether or not they know it. This theoretical framework will examine the issue in this way, placing a focus on the work that has already been done, but also define a framework for how assessment should be learned, where assessment should be learned, and what assessment topics teachers specifically need to know.

In an older article, Crooks (1988) discussed teachers' understanding of measurement techniques. "A substantial proportion of teachers have little or no formal training in educational measurement techniques, and many of those who do have such training find it of little relevance to their classroom evaluation activities" (p. 440). This was back in a time when it was considered sufficient for teachers to simply give tests that matched learning objectives. Today, research consistently shows that formative assessment is a necessary tool for student cognitive development (Black & Wiliam, 1998; Shepard et al., 2005). As the necessity for assessment skills grows, many call for an increase in training. Stiggins (2004) wrote "few teachers are prepared to face the challenges of classroom assessment because they have not been given the opportunity to learn to do so" (p. 762).

Two studies from the early nineties (Impara et al., 1993; Marso & Pigge, 1993) examined the impact of measurement courses for teachers and found that they did not make as much of an impact as many had hoped. Impara et al. (1993) administered a survey that consisted of two parts: a test of the *Standards for Teacher Competence in Educational Assessment of Students* and

a survey of teachers' perceptions and beliefs of assessment. They found that those who had taken a measurement class scored higher on the test of the *Standards*, but there was not much of a difference on the perceptions and beliefs survey between those who had taken a measurement class and those who had not.

Bearing a similarity to the Impara et al. (1993) findings, Siegel & Wissehr (2011) found that, though the teachers demonstrate a strong assessment literacy through surveys and conversations, they are not applying these methods during their practical classroom experiences. A large part of what they learned in their university classes was left behind when they entered the field. They posited that teacher actions are not only dependent on their assessment literacy, but on their views of the learning process in general.

DeLuca (2012) also discussed the criticism of assessment education courses for being too theory laden and furthering the disconnect between theory and practice. This is obviously a huge issue because the theory's only purpose is to prepare candidates for the practice. If it fails in doing that, then it was a waste of time. This points to the importance of practical experience in assessment education.

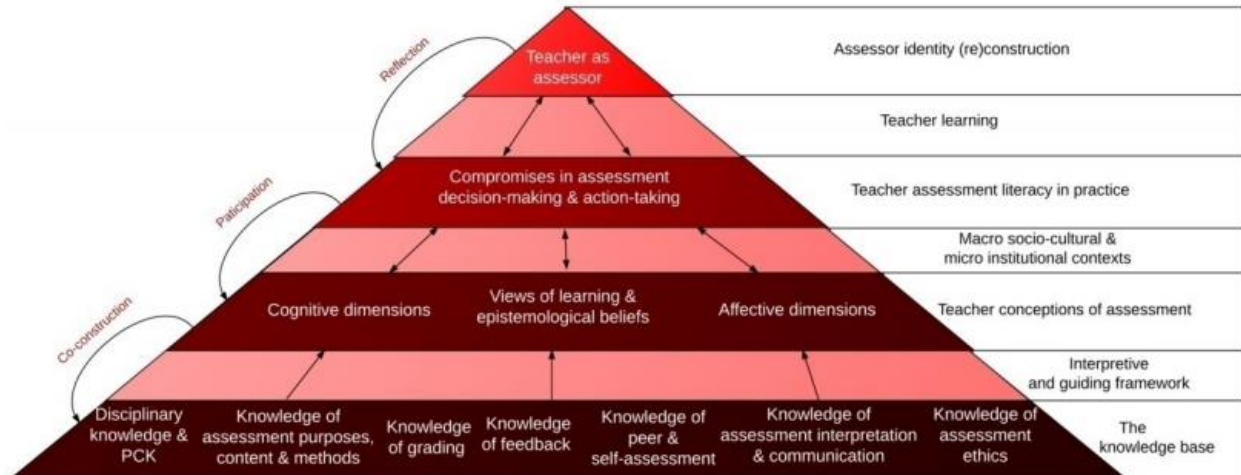
Kelting-Gibson et al. (2013) made a strong case for this when they studied their teacher candidates in a practical experience. After the teacher candidates completed an Assessment, Curriculum, and Instruction course, the researchers took a sample of 28 preservice teachers and qualitatively studied them as they assessed students during a trip to a children's discovery museum. In their findings, they concluded that preservice teachers must develop their own philosophy about assessment in order to become fully assessment literate. This comes as a result of practical experience and cannot strictly be learned in a classroom setting.

Wallace and White (2015) studied preservice mathematics teachers as they progressed through their apprenticeship. By interviewing them and collecting artifacts and data at various points throughout their apprenticeship, they found that the preservice teachers showed gradual growth and generally focused on *how to assess* before they learned *why to assess*. Only after learning the fundamental tools and types of assessments do their minds begin to connect to why it is being done in the first place.

### **Existing Theoretical Frameworks on Assessment Education**

Hill, Cowie, Gilmore, and Smith (2010) and Smith, Hill, Cowie, and Gilmore (2014) both stressed the importance of understanding preservice teachers' assessment beliefs while building a teacher education program. This, of course, is one of many things to consider when creating a framework. State laws and policies, research on best practice, and local factors are all crucial to the process as well.

Xu and Brown (2016) attempted to build a conceptual framework of teacher assessment literacy in practice (TALiP). To do this, they reviewed 100 studies on assessment literacy and arranged the information into three themes: knowledge and skills within assessment literacy, assessment education and its relationships with various mediating factors, and contextual considerations of assessment literacy. From this, they built a framework with six components, as can be seen in Figure 1. "From the bottom to the top respectively, they are the knowledge base, teacher conceptions of assessment, institutional and socio-cultural contexts, TALiP the core concept of the framework, teacher learning, and teacher identity (re)construction as assessors" (p. 18).



*Figure 1:* Teaching Assessment Literacy in Practice (TALiP) by Xu and Brown (2016). A conceptual framework for teaching assessment literacy.

Gottheiner and Siegel (2012) showcased a theory of assessment literacy that planted the teachers' views of learning in the middle and surrounded it with their assessment values and principles. Then, the knowledge of various assessment-related aspects encompasses these points. Figure 2 is their visual representation of the theory.

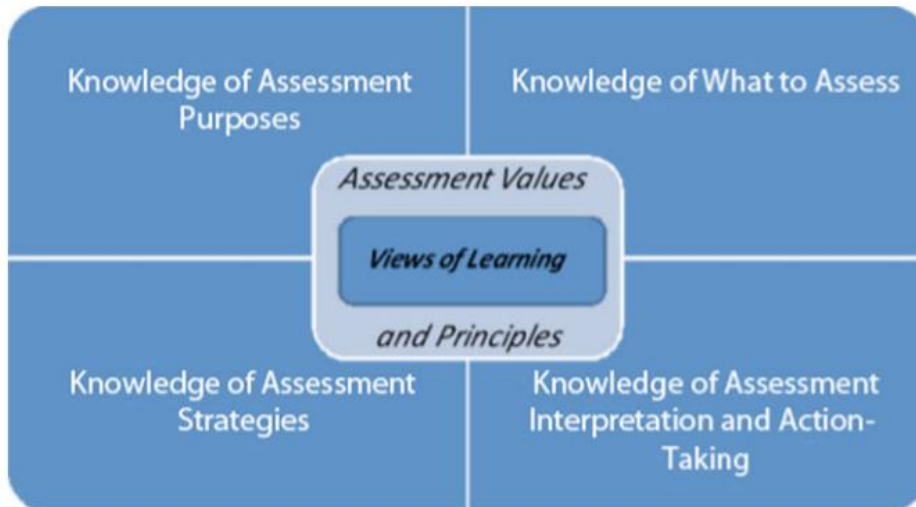


Figure 2: Theory of Assessment Literacy by Gottheiner and Siegel (2012). A visualization of teachers’ understanding of assessment.

Pierce and Chick (2013) formulated a “Framework for Professional Statistical Literacy” to show what teachers should be able to do. They stress that the hierarchy in this framework is not intended to promote a specific teaching sequence of these skills. Figure 3 shows a visual of their framework.

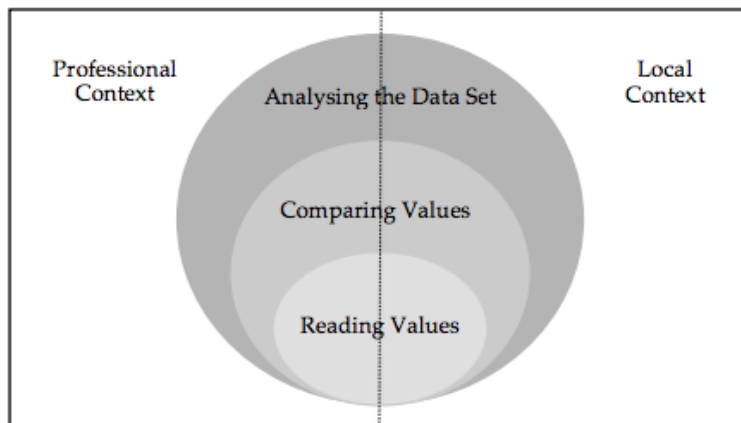


Figure 3: Framework for Professional Statistical Literacy by Pierce and Chick (2013). A visualization of the statistical knowledge needed to teach.



McMillan (2003) described a theory of assessment learning by saying, “What is needed is an understanding of how assessment and instruction are interwoven, with new conceptions about what assessment is and how it affects learning” (p. 39). He focused his theory on the importance of teaching teachers to embed assessment into every part of the learning process. “Assessment needs to be conceptualized as an ongoing activity that involves gathering, interpreting, and evaluating information, and action, based on results, rather than mere documentation of student performance (i.e., measurement)” (p. 39).

Mitton-Kukner, Munroe, & Graham (2015) laid out a framework that teaches grading as a “success for all” philosophy and not simply “sort and rank”. This framework aligns with many of the ideas of formative assessment as it teaches feedback, opportunities for improvement, and clear alignment to course objectives. They noted how this can become difficult because preservice teachers have often grown up in a “sort and rank” system and it is very difficult to break away from that mindset.

Elliott (2010) discussed the report by the National Academy of Sciences called *Knowing What Students Know: The Science and Design of Educational Assessment* (Pellegrino, Chudowsky, & Glaser, 2001). He posited that the recommendations in this report were not being followed and he wrote of a plan to prepare teachers for the assessment knowledge they will need. It consisted of the following:

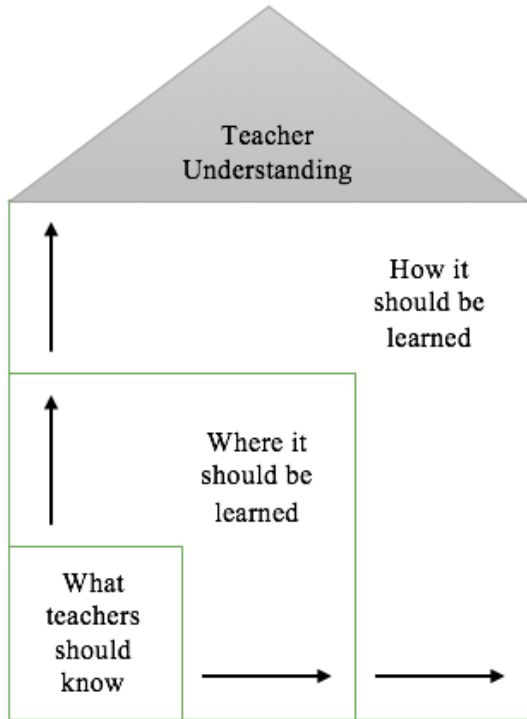
1. “That candidates be presented with multiple and rich course material in their preparation that will enable them to become assessment-literate and data-wise
2. That candidates participate in multiple and rich experiences with students so they can develop use of assessments in their teaching practice

3. That candidates' preparation experiences showcase faculty modeling that consistently demonstrates the application of good assessment practices for learning" (pp. 21-22)

### **The Framework of Assessment Education for Secondary Teachers**

To frame this study, a new theoretical framework was created by the researcher: The Framework of Assessment Education for Secondary Teachers (FAST). It was informed by the literature on the topic and organized into three building blocks: what teachers should know, where it should be learned, and how should it be learned. The study's methodology was created using the views of this framework and its results will be framed within this mindset.

FAST is conceptualized in Figure 4 and further explained in Table 1. In Figure 4, the theory of assessment education for secondary teachers is presented as an architectural structure. The cornerstone for this building is what should be taught to secondary teachers in the realm of assessment. Once this has been defined, it will inform the next building block of where assessment skills should be learned. In the same way, these two aspects will then provide a foundation for the decisions of how it should be learned. Once all these building blocks are firmly in place, the structure of teacher understanding will exist on a solid foundation.



*Figure 4: The Framework of Assessment Education for Secondary Teachers*

As the arrows indicate, the decisions on assessment education must be made in this order: what, where, how. Determining what should be learned is the essential cornerstone of this process. This corresponds with the bottom row and first step of Xu and Brown’s (2016) TALiP framework (Figure 1), which they refer to as “The Knowledge Base.” Teacher educators cannot decide where and how to teach the content if the content itself has not been defined. Likewise, they must answer the question of “where?” before they can tackle the question of “how?”. Knowing where it will be taught will provide information on available resources that can be used when deciding how it will be taught.

In Table 1, the specific components of each building block are listed. These components are further defined and justified in the sections that follow.

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	University	Instruction
Diverse learners	<ul style="list-style-type: none"> <li>• School of education</li> <li>• Content area</li> </ul>	Practical experience
Feedback and motivation	Employer	
Formative and summative		
Language and literacy		
Results to guide instruction		
Statistical literacy		
Student self-assessment		
Tools and types		

**What Should Teachers Know?**

When considering a theoretical framework, it is important to identify the target for what assessment knowledge all teachers must possess. Many independent organizations and government entities have developed recommendations of the assessment knowledge teachers should hold. For example, The National Council for Accreditation of Teacher Education (2008) (NCATE), in their *Unit Accreditation Standards*, recommended that teacher candidates should be able to “analyze student, classroom, and school performance data and make data-driven decisions about strategies for teaching and learning so that all students learn” (p. 15). Likewise, the Teacher Education Accreditation Council (2014) (TEAC), as part of their *Goals and Principles*, stated, “the program must provide evidence regarding the trustworthiness, reliability,

and validity of the evidence produced from the assessment method” (n.p.). When NCATE and TEAC merged in 2010, they formed the Council for the Accreditation of Educator Preparation (2013) (CAEP). CAEP addressed the issue in more broad terms, referring generally to all the skills covered in the national teaching standards which includes a section on assessment. Outside of the various standards for teacher education programs, other organizations like the Joint Committee on Standards for Educational Evaluation have created a complete set of standards for assessment use within the classroom (Klinger, McDivitt, Howard, Munoz, Roger, & Wylie, 2015). It can also be found in most teacher evaluation programs. For example, the commonly-used framework created by Charlotte Danielson (2007) contains component 3d, “using assessment in instruction.” Teachers are evaluated based on their mastery of this component within their classrooms.

In Illinois, the State Board of Education created the Illinois Professional Teaching Standards (IPTS) which outline the skills necessary to teach in the state (Illinois State Board of Education, 2013). Universities use these standards to guide the courses of their teacher preparation programs. Aside from many mentions of assessment throughout the standards, it is specifically addressed in standard 7: “The competent teacher understands and uses appropriate formative and summative assessments for determining student needs, monitoring student progress, measuring student growth, and evaluating student outcomes. The teacher makes decisions driven by data about curricular and instructional effectiveness and adjusts practices to meet the needs of each student” (p. 5-6). Under this there are 18 indicators that further detail these skills. These can be found in Appendix A.

To bring this concept to a more national scale, the IPTS are generally aligned to the Interstate New Teacher Assessment and Support Consortium (InTASC) teaching standards

created by the Council of Chief State School Officers (2011). REL West and WestEd (2009) reported on the similarities and differences between the sets of standards and found very few:

- “The IPTS are slightly different in wording and sequence, though the content of the two sets of standards is similar.
- Illinois added an 11<sup>th</sup> standard addressing professional conduct, and
- The INTASC standards include ‘dispositions’ under each principle that describe desired teacher attitudes” (p. 15).

InTASC standard 6, which addresses assessment, can be found in Appendix B.

To fully inform a theoretical framework for this study, this standard was considered in conjunction with the ideas established by Shepard et al. (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*. In this chapter, the authors grounded their “ideal assessment practices” (p. 276) in years of research on the topic. This research, as well as research that has surfaced in the years since the book’s publication, is discussed and synthesized in this framework.

The researcher will also consider the assessment aspects from Danielson’s Framework (2007) as this is how many teachers are evaluated once they enter the field. Charlotte Danielson created this framework in an attempt to define effective teaching practice. It was adopted and continues to be used as an approved model for teaching in over 20 states (Danielson Group, 2017). The Danielson Framework can be found in Appendix C.

The themes below were developed using the InTASC standard 6, Shepard et al. (2005) and Danielson (2007) along with other relevant resources and the study at hand. These are the themes determined to be the most crucial in assessment instruction. Using the InTASC standard

6, Shepard et al. (2005) and Danielson (2007) as the focal point of this framework is appropriate for two reasons. First, these three works combined cover assessment (and other teaching skills) in both the preservice and the inservice realms. Secondly, all of these resources are heavily used within the profession and a huge number of teachers and teacher candidates are exposed to them.

The study at hand confirmed these themes as being crucial to assessment education. One theme (language and literacy) was discovered during this study and then retroactively added to the framework. The merit of this additional theme will be discussed below.

**Alignment to learning goals.** A learning goal is the objective that the teacher intends for the student to meet at the end of a given period of time. It is only logical that all assessments within this given term should align to the specific goal or goals that are at hand. Teacher candidates should hold the ability to craft assessments that capture information on their students in accordance with these specific goals.

In regard to learning goals, Shepard et al. (2005) said,

Assessment cannot promote learning if it is based on tasks or questions that divert attention from the real goals of instruction. Historically, traditional tests have often misdirected instruction, if they focused on what was easiest to measure instead of what was important to learn. (p. 280)

This points to the necessity of teaching teacher candidates the importance of aligning their assessments to learning goals. Shepard (2003) even replaces the word “align” with the word “embody” to stress the complete importance of learning goals in the assessment process.

Standard 6(b) of the InTASC standards states, “The teacher designs assessments that match learning objectives with assessment methods and minimizes sources of bias that can distort assessment results” (p. 15). Standard 6(k) states, “The teacher understands the range of

types and multiple purposes of assessment and how to design, adapt, or select appropriate assessments to address specific learning goals and individual differences and to minimize sources of bias” (p. 15). Standard 6(r) states, “The teacher takes responsibility for aligning instruction and assessment with learning goals” (p. 15). These standards all stress the importance of aligning assessments to learning goals.

In the Danielson Framework (2007), component 1c is “Setting Instructional Outcomes” with the element “Value, sequence, and alignment” under it. Additionally, component 1f is “Designing Student Assessments” with the elements “Congruence with outcomes” and “Criteria and standards” under it. Like the InTASC standard, these include alignment as a crucial function of designing assessments and outcomes.

**Diverse learners.** Quality assessors must be aware of the cultural and educational differences present within their classrooms. In any given classroom, there might be diversity of race, gender, religion, language, sexual identity, socioeconomic status, or cultural upbringing, in addition to various learning styles and developmental levels. When assessing, these factors must be considered so no student is put at a disadvantage due to a factor that has nothing to do with what is being assessed.

As DeLuca and Lam (2014) suggested, if a teacher education program does not incorporate the connection into the curriculum, preservice teachers will leave their coursework without a firm grasp on the role diversity plays in the assessment process. This factor will become critical when they enter the classroom without the knowledge or ability to adjust. Aside from putting certain students at a disadvantage, there is even potential for legal ramifications if the teacher does not consider these issues of diversity.



Standard 6(h) says, “The teacher prepares all learners for the demands of particular assessment formats and makes appropriate accommodations in assessments or testing conditions, especially for learners with disabilities and language learning needs” (p. 15). Standard 6(p) states, “The teacher understands how to prepare learners for assessments and how to make accommodations in assessments and testing conditions, especially for learners with disabilities and language learning needs” (p. 15). Standard 6(u) says, “The teacher is committed to making accommodations in assessments and testing conditions, especially for learners with disabilities and language learning needs” (p. 15).

Diversity in relation to assessments is mentioned in Danielson’s Framework (2007) in component 1c “Setting Instructional Outcomes”. The outcomes as measured by the assessments should consider the element “Suitability for diverse learners”. Preservice teachers should walk away from their teacher education programs with an understanding of how to assess these diverse learners.

**Feedback and motivation.** Perhaps the greatest impact of an effective assessor is the emphasis they put on feedback. Shepard et al. (2005) wrote,

Novice teachers should be able to analyze student work and identify patterns of errors and gaps that most need to be addressed. [...] Teachers must also understand the theory of how feedback enhances learning so that they can develop classroom routines that check for student understanding and ensure that students are not left alone to persist in bad habits or misconceptions. (p. 288)

Marzano (2010) stressed the importance of feedback by saying, “its most important and dominant characteristic is that it informs the student, the teacher, and all other interested parties about how to best enhance student learning” (p. 3). In his 2009 book *Visible Learning: A*

*Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Hattie identified the biggest factors related to student achievement by undertaking a huge meta-analysis. In terms of effect size, feedback ranked as one of the highest factors on the list. Even as he continues to update his research, it remains high on the list. Teacher candidates should absolutely be skilled in providing feedback to students on where they are and how they can improve.

Many studies have focused on student motivation as a result of the feedback they receive on their performance. Tomlinson (2005) wrote,

In brief, positive classroom environment, balance of challenge and safety, sense of community, student acceptance of responsibility for personal growth, shared responsibility for effective classroom routines, and certainly achievement are all threatened if student motivation is impaired - by grading practices or otherwise. (p. 266)

She went on to discuss students who believe that success is out of their reach and their tendency to either give up on themselves or on their school. These situations can result in students completely shutting down or dropping out, thus making student motivation a huge priority for educators to consider.

Thomas and Oldfather (1997) examined the way students are affected by the grades they receive. They found that students who are empowered by ways to improve their learning are more likely to take ownership of their abilities and are more likely to experience learning in a way that is connected to their own lives and identities. The way they are assessed can add to or take away from this empowerment. They also reported that challenges that are either too easy or too hard prevent students from feeling competent. All of these factors point to the necessity for teacher candidates to understand the factors that can enhance or impair student motivation.

The InTASC standards address feedback in multiple places. Standard 6(d) says, “The teacher engages learners in understanding and identifying quality work and provides them with effective descriptive feedback to guide their progress toward that work” (p. 15). Standard 6(n) says, “The teacher understands the positive impact of effective descriptive feedback for learners and knows a variety of strategies for communicating this feedback” (p. 15). Standard 6(s) says, “The teacher is committed to providing timely and effective descriptive feedback to learners on their progress” (p. 15).

The Danielson Framework (2007) directly addresses feedback in component 3d “Using Assessment in Instruction” where the element “Feedback to students” is contained. All of these sources emphasize the fact that feedback is an essential aspect of both the teaching and assessing process and should be taught to future teachers.

**Formative and summative.** When assessment comes to mind, most often think of summative techniques first. Teacher candidates must know the elements of good assessment strategies after learning has (or has not) occurred, but it is also essential for them to master the art of data collection *during* the learning process. This idea of formative assessment is often informal and sometimes spontaneous, but it is crucial to students’ education. In Black and Wiliam’s (1998) review of literature on formative assessment, they discussed primary school practices where “teachers’ records tend to emphasise [sic] the quantity of students’ work rather than its quality, and that whilst tasks are often framed in cognitive terms, the assessments are in affective terms, with emphasis on social and managerial functions” (p. 18). They also reported concerns of the learning function being under-emphasized while the grading function is over-emphasized. These problems contradict the idea that both formative and summative assessment

should be thought of as assessment *for* learning instead of merely assessment *of* learning (Chappuis, Stiggins, Chappuis, & Arter, 2012).

Atkin, Black, and Coffey (2001) discuss the learning assessment process with these three questions:

1. Where are you trying to go?
2. Where are you now?
3. How can you get there?

Formative assessment supports the learning process by answering the second question, which is crucial to the overall goal. Its focus is not often on evaluating for a grade, but instead it is designed as a checkpoint for the student, the parent, and the teachers themselves.

Formative assessment is incorporated throughout all of the InTASC assessment standard components. It can explicitly be seen in standard 6(a): “The teacher balances the use of formative and summative assessment as appropriate to support, verify, and document learning” (p. 15). Also, it is in 6(j): “The teacher understands the differences between formative and summative applications of assessment and knows how and when to use each” (p. 15).

The Danielson Framework (2007) likewise promotes formative assessment throughout. It can be seen explicitly in component 1f “Designing Student Assessments” which contains the element “formative assessments”. It’s presence in all of Domain 3 also cannot be understated. All these factors demonstrate the importance that formative assessment should have in the educational process.

**Language and literacy.** Students must be given the opportunity to express themselves through language and quality assessors must consider literacy when designing assessments. Assessment is a form of communication; it is a message being received by the teacher from the

student. Teacher candidates must understand this and keep it in mind during the assessment process.

This is a theme that is not clearly apparent in either the InTASC standard 6 nor the Danielson Framework. During the study described in later chapters, this theme was discovered in multiple other sources, proving it important enough to be included on this list. It is unique enough from the other themes to stand on its own.

**Results to guide instruction.** Assessment is only partially about evaluating students for grading purposes. The overemphasis put on this purpose (Black & Wiliam, 1998) takes away from the necessary idea of using the assessment results to guide instruction. Not all assessment is summative, which means that there is still time to adjust instruction to better educate the students. This is a skill that must be taught to teacher candidates.

Shepard et al. (2005) discussed assessment use for the evaluation of teaching, a flip on the common perception of assessment use for the evaluation of learning. Teachers must analyze assessment results to better understand their own practices and how they affect the students. They must be able to identify potential issues and adjust their teaching accordingly. This idea must be instilled in teacher candidates.

Using assessment to guide instruction is written throughout the InTASC assessment standards, but it is mentioned specifically in a few places. Standard 6(c) says, “The teacher works independently and collaboratively to examine test and other performance data to understand each learner’s progress and to guide planning” (p. 15). Standard 6(l) states, “The teacher knows how to analyze assessment data to understand patterns and gaps in learning, to guide planning and instruction, and to provide meaningful feedback to all learners” (p. 15).

In Danielson's Framework (2007), component 1f "Designing Student Assessments" contains the element "Use for planning". Also, component 3e "Demonstrating Flexibility and Responsiveness" contains the elements "Lesson adjustment" and "Response to students". Multiple mentions of this display its need for instruction in teacher education programs.

**Statistical literacy.** Pierce and Chick (2011; 2013; 2014) frequently stress the importance of statistical literacy as it relates to assessment literacy as a whole. Grading is a huge part of this. Though Black and Wiliam (1998) say it is overemphasized, grading is still an essential part of the assessment process and is required by school districts in order to produce evaluative marks for a student. In many classrooms, grading is distorted by outside factors that result in a blurred vision of a student's achievement. As classical test theory states, Observed Score = True Score + Measurement Error (Lord & Novick, 1968). It is necessary to remember that the measurement error is a variable that will change the outcome of the observed score, thus it must be minimized. Stiggins and Knight (1997) referred to these factors that corrupt the true score as "grade pollution." They said, "when the object is effective communication [of achievement] ... schools [should] adopt grading policies that permit teachers to indicate each student's current level of academic achievement with nothing else factored in to interfere with that message" (in O'Connor, 2002, p. 87). There are many factors that can get in the way of a true score like poorly written task directions, key goals incorrectly matched to assessment instruments, or a reading or language barrier for the student, but these factors should be minimized (Tomlinson, 2005).

It is hugely important for grades to be considered valid. McMillan (2008) wrote, "strong validity is demonstrated when evidence and logic suggest that the evaluation is accurate and reasonable" (p. 19). Another essential is that grades should be reliable, which McMillan

described as being “concerned with the consistency of scores obtained from assessment” (p. 35). Therefore, with both of these terms, the importance is the interpretation and inference of the information gathered from the assessments. The teacher must be able to come to a valid and reliable conclusion based on the information that is gathered. Various forms of grade pollution can negatively affect these conclusions.

As a method of accurate scoring and as a communicative device to inform students of what is expected of them, teacher candidates must possess the abilities to draft solid rubrics in age-appropriate language (Shepard et al., 2005). Frederiksen and Collins (1989) used the term “transparency” to describe the idea that teachers should ensure students have a clear understanding of how they will be assessed. Teachers must be transparent with their grading methods so that the student and all other stakeholders know exactly what to expect before it is graded, and exactly why they received the mark they did after it is graded.

Though the InTASC assessment standards put more emphasis on assessing *for* learning, instances of statistical literacy can be found throughout. One example is standard 6(c) which states, “The teacher works independently and collaboratively to examine test and other performance data to understand each learner’s progress and to guide planning” (p. 15). This obviously would not be possible without an understanding of how to examine the data. Also, grade reporting can be found in standard 6(o) which states, “The teacher knows when and how to evaluate and report learner progress against standards” (p. 15).

Statistical literacy can be found in the Danielson Framework (2007) throughout all of component 4b “Maintaining Accurate Records”. Aspects of it are also suggested in areas of Domain 1 and Domain 3. All of these factors point to the need for skills in data interpretation to be an effective assessor.

**Student self-assessment.** As mentioned previously, Hattie (2009) identified the biggest factors related to student achievement by undertaking a huge meta-analysis. At the top of his original list with the biggest effect size was students self-reporting their grades. He emphasized this crucial step in students understanding their levels of achievement and how to improve. With the clear necessity of this factor, it is imperative that teacher candidates should be taught its importance.

Shepard et al. (2005) stressed both the motivational and the cognitive purposes behind student-self assessment. Students become more interested in their performance and are able to apply the criteria of evaluation within the context of their own work. Teacher candidates should understand the theory behind both of these purposes and learn skills to incorporate them into classroom lessons and activities.

Standard 6(f) of the InTASC standards states, “The teacher models and structures processes that guide learners in examining their own thinking and learning as well as the performance of others” (p. 15). Standard 6(m) says, “The teacher knows when and how to engage learners in analyzing their own assessment results and in helping to set goals for their own learning” (p. 15). Standard 6(q) says, “The teacher is committed to engaging learners actively in assessment processes and to developing each learner’s capacity to review and communicate about their own progress and learning” (p. 15).

The Danielson Framework contains component 3d “Using Assessment in Instruction” which includes the element “Student self-assessment and monitoring”. Its inclusion reflects the research that indicates its importance and, therefore, it should not be ignored in teacher education programs.



**Tools and types.** Because assessment is not conducted in only one manner, teacher candidates must have an arsenal of assessment tools and types that can be utilized correctly in the appropriate situation. As Shepard et al. (2005) said, “Assessment tools include oral questioning of students, observation, written work products, oral presentations, interviews, projects, portfolios, tests, and quizzes” (p. 294). These tools could fall within the realm of formative or summative assessments. Without full knowledge of these tactics, teacher candidates will find themselves without the correct tool for the job.

Standard 6(e) of the InTASC standards says, “The teacher engages learners in multiple ways of demonstrating knowledge and skill as part of the assessment process” (p. 15). Standard 6(t) states, “The teacher is committed to using multiple types of assessment processes to support, verify, and document learning” (p. 15). Specific references to tools or types are not made in the Danielson Framework (2007) though it can be argued that none of the other assessment components or elements could be completed without the right tools. Understanding which tools and types are appropriate should be a necessary outcome of all teacher education programs.

### **Where Should it be Learned?**

The question of where teachers should gain their assessment knowledge is frequently debated among stakeholders in the field. Though few would disagree it should be covered in undergraduate programs, there is disagreement on the extent of its coverage. In specialized secondary programs, there is also the conundrum of deciding which courses in which departments should take the responsibility.

CAEP (2013) issued the standards that accredited programs must follow. The five standards cover the following areas:

1. Content and Pedagogical Knowledge

2. Clinical Partnerships and Practice
3. Candidate Quality, Recruitment, and Selectivity
4. Program Impact
5. Provider Quality Assurance and Continuous Improvement

Under the Content and Pedagogical Knowledge section, the standards say, “Providers ensure that candidates use research and evidence to develop an understanding of the teaching profession and use both to measure their P-12 students’ progress and their own professional practice” (n.p.).

Practical experience during student teaching or classroom observations must certainly play a role, as evidenced by the research above. It cannot be expected, however, for the teacher candidates to simply learn how to assess on the go. They must gain an understanding of what is happening so they understand what they are watching during observations. They also must have a strong arsenal of assessment tools, skills, and knowledge that they can use as they build their own assessments during their time student teaching.

There is a school of thought that the school districts themselves should bear the brunt of a teacher’s assessment knowledge through professional development (Highland, 2015). In this scenario, the teacher education program would teach the basics to lay the groundwork on which districts would later build. This argument builds on the idea that true assessment knowledge can only be gained through experience while practicing in the real world. Also, because districts often have varying assessment policies (for example, standards-based grading, no homework, etc.), it would be nearly impossible for the teacher education program to prepare all students in a way that will be useful to them in the future.

In a secondary education program, the teacher candidate usually splits his or her time between the school of education and the school of their individual content area, be it English,

mathematics, business, etc. This becomes problematic for assessment education because of the varied forms education can take among the various disciplines. English assessments and mathematics assessments can often look very different, resulting in difficult instruction within the school of education. The content areas, however, do not always provide any assessment instruction because not all their students are working toward an education degree. This raises an interesting debate between the colleges.

In this theoretical framework, the emphasis on assessment education is a shared responsibility of the school of education and the content areas. The school of education should be thorough in their assessment instruction, but the content areas will need to fill in the areas that are subject-specific. It cannot stop with the college diploma, however. Schools absolutely must continue to provide opportunities for professional learning for their teachers and keep them up-to-date on the newest trends and strategies in academic assessment.

### **How Should it be Learned?**

The next step in the process is defining how assessment should be learned. The above body of research suggests that programs should teach *how to assess* before students can fully understand *why to assess*. Teacher education programs also should make a conscious effort to explicitly incorporate student diversity into their teaching of assessment. Additionally, they should incorporate the changing landscape of assessment by addressing accountability and assessment reform ideas such as computer-based testing and other innovative data collection procedures. When paired heavily with practical experience, this will lead to preservice teachers who are truly assessment literate.

Incorporating more practical experience would directly address the disconnect between theoretical assessment knowledge and practical assessment knowledge as it presents more

opportunity to gain experience in the practical realm. Kelting-Gibson et al. (2013) also connected this to assessing diverse learners by saying, “all students do not learn in the same way; therefore, it is important to know your students to design and deliver effective instruction” (p. 539). This is something that would not be possible to learn without experience. The need for practical experience is supported by the *how to assess* before *why to assess* issue presented by Wallace and White (2015). Making the connection to why this is being done is not as easy when they are not engaged in a real experience where real instructional decisions are being made.

### **Summary**

Debates continue on what specific assessment skills should be taught, where to teach it, and how it should be rolled out to educators. The general consensus from the research is that the average teacher is not fully assessment literate, but there has not been a consensus on how to fix this issue. Limited literature exists on this specific issue of secondary education assessment instruction. This study will explore it but it is also recommended that future studies take a closer look at the interaction between the school of education and the schools of the various content areas to better capture their handling of assessment instruction.

The theoretical framework under which this study will be conducted was carefully crafted from the research and policy documents that do exist on the matter. In theory, assessment should be taught during preservice teachers’ undergraduate programs. Elements of assessment should be taught by both the College of Education and the individual content areas and continually developed by the school districts. It is important to incorporate a large amount of practical experiences that will specifically target assessment instruction. The dominating content themes for assessment instruction are 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. Based on the research and recommendations from policy documents, these themes should be learned by all preservice teachers.

The content themes described above were used or discovered in the document analysis process of this study. This study extensively examines the “what teachers should know” aspect of this theoretical framework. Only aspects of the other two realms are covered in this study. For example, this study examines the coverage of assessment instruction within a university’s college of education (“where it should be learned”) but it does not examine the role that school districts could play in this process. There is a beginning discussion of the “how it should be learned” realm through the examination of the depth of knowledge and range of knowledge levels of assessment instruction, but these examinations are not sufficient to fully discuss this realm. It is recommended that future studies examine this in much more detail.

This program follows the strategy promoted by Jones (2014) when she discussed the value of embedding assessment instruction throughout. Many programs often tack on an assessment course toward the end of the education program, which makes connecting assessment to the larger education context somewhat difficult. She found that embedding it throughout helps make these connections, especially in the realm of formative assessments used for informing future instruction.

## CHAPTER IV: RESEARCH DESIGN AND METHODOLOGY

There are not many studies that sufficiently examine the coverage and depth of assessment instruction in the university setting and that is what this one set out to do. To do this, the researcher used several existing resources to assess the knowledge of undergraduate students and to analyze the instruction they receive. Using a developed methodology, the assessment, the instruction, and the standards were examined for alignment.

### **Research Questions**

This study followed a mixed-methods research design and was guided by three main questions:

1. Where is assessment being taught to undergraduate secondary education preservice teachers?
2. How closely does assessment instruction align with assessment practices as defined by InTASC and the edTPA?
3. How do secondary education preservice teachers perform in the realm of student assessment as measured by the edTPA?

### **Setting**

The study focused on five undergraduate secondary education programs at a large state university in Illinois. Identifying information for the university and these programs has been changed to protect confidentiality. The programs are accredited by the National Council for Accreditation of Teacher Education (NCATE). In these programs, undergraduate students must take a combination of courses within the School of Curriculum and Instruction (CI), the Department of Educational Administration and Foundations (EAF), the Department of Psychology (PSY), and their chosen content area department (English, mathematics, theater,

health, etc.). The CI department teaches early childhood education, elementary education, and middle level education while providing service courses for other education majors (special education, secondary, and K-12). The EAF department only offers service courses to other programs at the undergraduate level and is not an undergraduate program itself. Clinical experiences happen throughout and the program is capped off with a student teaching experience.

Within the College of Education (which includes CI, EAF, and Special Education), students are required to pass several courses with a C or better. Those courses are CI 210 (Child Growth and Development), CI 214 (The Teaching Profession in Secondary Schools), CI 218 (Principles and Practices for Teaching and Learning in Secondary Schools), CI 221 (Integrating Multiple Literacies and Technology Across the Secondary Curriculum), and one of the following: EAF 228 (Social Foundations), EAF 231 (Introduction to Philosophy of Education), and EAF 235 (Historical Foundations). The three EAF courses cover much of the same material, but they approach them through different lenses (social, philosophical, historical). There are also other offerings within these departments that students can choose as elective courses. In addition, there are required courses within their content departments that are specific to each program. Some are focused on the subject itself while others are methods courses that prepare for teaching the subject. The program and all education syllabi are aligned to the IPTS.

To receive their teaching license, teacher candidates also must pass a performance-based assessment known as edTPA. This nationally-available assessment is subject-specific and requires preservice teachers to submit teaching artifacts that demonstrate their ability to effectively teach and assess their students. It is aligned to major subject content standards and is

strongly aligned to the InTASC standards (Stanford Center for Assessment, Learning, & Equity, 2014). It consists of three tasks:

1. Planning for Instruction and Assessment (Planning)
2. Instructing and Engaging Students in Learning (Instruction)
3. Assessing Student Learning (Assessment)

Preservice teachers submit their edTPA portfolio during their student teaching experience. It is scored by trained scorers (not employed by the university) using rubrics that are specific to the preservice teacher's teaching field. There are five rubrics in each task, adding up to 15 rubrics total (Stanford Center for Assessment, Learning, & Equity, 2016). Researchers have studied the assessment extensively to assure reliability and validity of results. A full breakdown of these studies can be found in the most recent edTPA Administrative Report (edTPA, 2017).

### **Instrumentation**

To answer the research questions, this mixed methods study was conducted through document analyses of syllabi, teaching standards, and the edTPA handbook, and also through analysis of edTPA data. Of Creswell's (2013) six mixed methods design strategies, this study falls under the Concurrent Nested design strategy. This design strategy is characterized by one approach guiding the project while the other is embedded into the project in a supporting role. The qualitative document analysis and alignment study is the central piece as it provided for the most analysis. The quantitative data are supplemental and are mainly used to provide context for the larger study.



## **Document Analysis**

This study utilized a form of document analysis to examine data from course syllabi, edTPA rubrics, and the InTASC standards. Bowen (2009), in his article “Document Analysis as a Qualitative Research Method,” said,

The analytic procedure entails finding, selecting, appraising (making sense of), and synthesizing data contained in documents. Document analysis yields data – excerpts, quotations, or entire passages – that are then organized into major themes, categories, and case examples specifically through content analysis. (p. 28)

Document analyses are non-reactive to the research process, which makes the method effective for this study. The syllabi, rubrics, and standards exist for other reasons than this study, which allows for an unaltered glimpse of reality. Often, document analyses are used as complementary sources to aid in triangulating data (Bowen, 2009). In this study, however, the analysis took a more central role.

The researcher analyzed seventeen course syllabi. These documents were used to define the intended instruction within the university. Two other studies (DeLuca & Bellara, 2013; Greenberg & Walsh, 2012) have used course syllabi as a valid source of capturing intended assessment instruction as well.

The researcher also examined the edTPA assessment handbook (which contains the rubrics) to identify common themes and categories pertaining to assessment. Each content area has a different handbook (though all sciences use the same one), but all are designed to assess the same skills, simply using more subject-appropriate language. Because of this, only one subject was chosen for analysis. The secondary science handbook was used for this study. This content area was chosen because three of the five methods courses in this study come from the sciences.

To find common themes and categories pertaining to assessment, the researcher analyzed standard 6 of the InTASC standards and all its indicators. The standard (“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”) contains 22 indicators that are split into three groups: performances, essential knowledge, and critical dispositions. The standard and all indicators can be found in Appendix B.

### **edTPA Data Analysis**

Scores from seven edTPA rubrics were analyzed for the university’s secondary education preservice teachers. These seven rubrics are rubrics 5, 10, 11, 12, 13, 14, and 15 (included as appendices D, E, F, G, H, I, and J). These seven were chosen because they were identified during the document analysis as containing assessment concepts. The three tasks (planning, instruction, and assessment) were also analyzed.

The study analyzed data from four separate test administrations to determine how preservice teachers perform in the area of assessment. The data were from the 2015 fall administration, the 2016 spring administration, the 2016 fall administration, and the 2017 spring administration. All four administrations were used to increase the sample size. The analysis was focused on students’ first attempts at passing the examination. The edTPA guidelines allow for students to resubmit all or part of the assessment if they did not pass on their first attempt, but, to better understand student ability, only the first attempts were examined.

Only edTPA data from five subject areas were analyzed: biology, chemistry, communications, physics, and theater. This decision was made to better align the assessment data with the syllabi that were analyzed. The five subjects are secondary-only and their selection is explained in the “Overall and Sample Populations” section. The researcher obtained the data as a

spreadsheet containing individual student scores on each rubric and also a comprehensive score for each student. Student identifiable information was not provided to the researcher. Due to privacy restrictions, it was not possible to break down the data to the individual content areas (i.e. chemistry, physics, theater, etc.). It can be assured, however, that only the five identified subject areas were included in the spreadsheet.

On each rubric, student work is given a score of 1 through 5. Half points may be awarded as well. The rubrics are designed as a continuum of professional practice where a 3 is considered proficient. In total, the composite score is out of a total of 75 points (a combination of 15 rubrics). In the test administrations used for this study, the minimum passing score in the state of Illinois was a 35.

### **Overall and Sample Populations**

To identify which syllabi to use, the researcher chose a convenience sampling technique. This method was appropriate due to the *convenient* availability of the syllabi (Fraenkel & Wallen, 2006). Convenience samples are occasionally referred to as “accidental samples” due to the fact that “elements may be selected in the sample simply as they just happen to be situated, spatially or administratively, near to where the researcher is conducting the data collection” (Etikan, Musa, & Alkassim, 2016, p. 2). The method is limited by the inability to make strong statements about the population due to the sample not being selected purposively. Because of this, it is crucial that the researcher describe the sample extensively and note the parts of the population that might be left out.

To acquire the syllabi, the researcher contacted the director of the teacher education center at the university. The director of the teacher education center then sent an email to the directors of all the secondary and K-12 subject areas. Nine subjects responded, eight of which

agreed to participate in the study. Of those eight subjects, five of them were specifically secondary programs (the others encompassing K-12), so the decision was made to use these five to keep the focus of the study on secondary education. These five subjects are biology (BIO), chemistry (CHE), communication (COM), physics (PHY), and theater (THT). Additionally, the researcher contacted the director of CI and requested the syllabi to the required secondary education courses within the department. These three syllabi were emailed to the researcher as well.

Either the director or a professor from these subject areas emailed their respective methods syllabi to the researcher. Some individuals provided the course schedules as a separate document from the syllabus while others provided syllabi that contained a schedule within it. For this study, all documents provided by the subject areas were considered part of the syllabi and were analyzed.

In total, 17 syllabi were collected for use in this study. They are as follows:

1. BIO 305: Methods of Teaching Biology
2. CHE 160: Introduction to Teaching Science Seminar
3. CHE 303: Teaching of Chemistry
4. CHE 304: Student Teaching and Professional Seminar
5. COM 291: Principles of Speech Education
6. COM 293: Communication in the Classroom
7. COM 391: Theory and Practice of Communication Instruction
8. PHY 215: Introduction to Teaching High School Physics
9. PHY 323: Teaching High School Physics
10. PHY 324: Inquiry and High School Physics

11. PHY 357: Student Teaching Seminar
12. THT 186: Continuing Orientation to Theatre Teacher Education
13. THT 286: Continuing Orientation to Theatre Teacher Education
14. THT 386: Principles of Theatre Education
15. CI 214: The Teaching Profession in Secondary Schools
16. CI 218: Principles and Practices for Teaching and Learning in Secondary Schools
17. CI 221: Integrating Multiple Literacies and Technology Across the Secondary Curriculum

This study used the edTPA data from 60 students. Of these 60, seven were from the fall of 2015, 21 from the spring of 2016, 14 from the fall of 2016, and 18 from the spring of 2017. These are the first four semesters that edTPA was a requirement within the state of Illinois. During this time, there were no major changes in the university's education curriculum nor the assessment itself, so the researcher generally assumed that each of these semesters' data are comparable.

All 60 students come from one of five different subjects: biology, chemistry, communications, physics, or theater. This decision was made to better align with the syllabi that were acquired for the study. These quantitative data will provide additional depth to the study by allowing for analysis of the resulting performance after experiencing the program. Due to privacy concerns by the individual subject areas, it was not possible to disaggregate the data by subject areas, so a more general analysis was required.

### **Methodology**

To answer the research questions, this study largely took an alignment methodology approach. In the realm of assessment, the course syllabi were studied for alignment with InTASC

standard 6 and also for alignment with the edTPA Assessment Handbook. Porter and Smithson (2001) define these three pieces as the *intended curriculum*, the *assessed curriculum*, and the *enacted curriculum*. The standards would be the *intended curriculum*, which they define as “such policy tools as curriculum standards, frameworks, or guidelines that outline [what] the curriculum teachers are expected to deliver” (p. 2). The assessment handbook would constitute as the *assessed curriculum*, or the tools that measure if students have learned the intended curriculum. The syllabi would fall into the category of *enacted curriculum*, defined as “the actual curricular content that students engage in the classroom” (p. 2). It is important to note that syllabi did not fall into the exact definition of enacted curriculum. Often, when analyzing the enacted curriculum, researchers use tools like surveys or observations that capture what is actually happening in the classroom. The syllabi exist to provide a specific overview of course topics and assignments, but they are usually created and made available *before* the actual course has taken place. They are not a record of what actually happened in the course. DeLuca and Bellara (2013) classify the syllabi as *intended curriculum*, but the researcher of this study disagrees with this statement because they are still the element that are guided by the standards and policies that define what should be taught. The syllabi represent the curriculum in action. Regardless, the researcher did not assume the syllabi represent how assessment is being taught, but instead only took the information on what is being taught, where it is being taught, and the degree of assessment coverage in the curriculum.

In their review of alignment methodologies, Bhola, Impara, and Buckendahl (2003) classify alignment methodologies into three categories of complexity: low, moderate and high. Low complexity models simply look for content coverage between parts. Moderate complexity models add in extra dimensions to the study beyond simple coverage. These models often

examine additional realms like cognitive complexity or range of knowledge, which make it more comprehensive than a simple methodology. High complexity models bring in even more information, criteria, and evidence to further the study. This study used a moderate complexity alignment model developed by Webb (1997, 1999, 2005) and modified by DeLuca and Bellara (2013). It examined the data in three dimensions. A few adjustments were made to the existing methodology to adapt it to this study, the main change being the addition of an *assessed curriculum* component, which DeLuca and Bellara (2013) did not have.

## **Procedures**

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 researcher analyzed each document completely independent of each other. Official comparisons between documents were not made until all document analyses had concluded.

**Categorical concurrence.** Deductive coding was used to analyze the content data. The researcher considered the codes that were identified previously when classifying content. This was decided in an attempt to not strictly go by the categories identified in any one data source. Instead, anchors were generated from the larger body of research around best practice. The researcher did remain open to the discovery of new codes during the analysis process and these new codes were added when they were discovered. The final codes are as follows:

- 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

**Depth of knowledge.** The researcher coded each data source for DOK levels. Table 2 details the coding criteria as defined by DeLuca & Bellara (2013) and adapted from Webb (1997).

Table 2

*Depth of Knowledge Levels*

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).

There are only three DOK levels used in this study, as opposed to Webb’s four levels. Condensing the levels from four to three was a decision initially made by DeLuca & Bellara



(2013) and continued in this study. This was done for simplicity purposes and to keep a parallel structure with the range of knowledge levels. Table 3 shows the crosswalk from Webb’s (1997) levels to DeLuca & Bellara’s (2013) levels.

Table 3

*Crosswalk from Webb’s DOK Levels to DeLuca and Bellara’s DOK Levels*

Webb’s DOK Level		DeLuca & Bellara’s DOK Level
Level 1: Recall	→	Level 1: Low cognitive level
Level 2: Skill/Concept	→	Level 1: Low cognitive level
Level 3: Strategic Thinking	→	Level 2: Moderate cognitive level
Level 4: Extended Thinking	→	Level 3: High Cognitive Level

*Note.* (DeLuca & Bellara, 2013; Webb, 1997).

**Range of knowledge.** The researcher coded each data source for ROK levels. Table 4 details the coding criteria as defined by DeLuca & Bellara (2013) and adapted from Russell and Airasian’s (2011) typology of global standards, educational standards, and instructional standards, which in turn was adapted from the work done by Krathwohl and Payne (1971).

Table 4

*Range of Knowledge Levels*

Range of Knowledge Level	Coding Criteria
Level 1: Instructional objective	Specific objectives used to plan assessment use in daily teacher practice (e.g., test design, questioning approaches, etc.)
Level 2: Educational objective	Statements that describe teacher accomplishment 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).

**Coding process.** It would be impossible to assign DOK and ROK codes to every mention of assessment in the syllabi. Assessment mentions are often brief like “Introduction to Standard Assessments” or simply content topics like “Student Performance Assessments: Scoring Rubrics”. Because of this, the contents of the assessment were then categorized into two parts: objectives and supplementary pieces of content. The decision was made to code the objectives in all three categories and to code the supplementary pieces of content in only the CC category.

This plan was also applied to the edTPA assessment handbook. Ultimately, the only pieces of content resembling objectives within the handbook were the main topics of the rubrics themselves. These were always written as questions, such as rubric 14: “Analyzing students’ language use and science learning: How does the candidate analyze students’ use of language to

develop content understanding?” The researcher decided to examine this content as if it were written as an objective statement. So, in the case of rubric 14, it was coded as if it said, “The candidate analyzes students’ use of language to develop content understanding.”

A decision that had to be made during coding was how to handle certain pieces of content that appeared in the syllabi. First of all, many of the courses contain clinical experiences as a crucial element. These experiences might be observations or actual teaching, which would involve lesson planning and, sometimes, assessing the students. It was decided to not code these situations unless there were explicit mentions of assessment. If it was not said explicitly, the researcher chose not to assume anything. Likewise, several syllabi mention edTPA preparations for the students. The edTPA certainly contains many assessment elements (as can be seen in this study), but, again, unless the syllabus explicitly discusses preparing students for the assessment portions of the edTPA, it was not coded. Additionally, there were occasionally situations where the syllabi gave students options for what they might include in a portfolio and that they artifacts they use “could include” tests, quizzes, or other assessments. It was decided that, in these situations, the pieces of content *would be* coded. This was decided because there is explicit mention of assessment and, even though it might not be used, the researcher did not want to ignore this.

When content suggested multiple codes, the researcher decided that it could be double coded in the CC category, but not in other categories. This is because it is not possible for an objective to exist in two different DOK or ROK levels at the same time. Because of their nature, each level is mutually exclusive. Likewise, the researcher decided that every objective must be coded with a DOK and an ROK level, but that rule does not apply to the CC category. There were times that an objective was written very broadly (for example, “Design and implement

instruction on an assigned chemistry topic in consideration of NGSS, including creation of an assessment”) and, thus, did not fit in to one specific code.

**Organization of codes.** 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. For example, 18% of the InTASC standard 6 indicators were coded with the “statistical literacy” code. Displaying the data in this way made each data source comparable, since there are a different number of data pieces in each one. For the syllabi, the researcher first identified the frequencies by each individual syllabus and rubric and then later combined them into a full intended curriculum category. Additionally, data are presented in a matrix that allows for multiple dimensions to be compared at once. This allows for an even deeper identification of points of alignment and misalignment across the data sources.

The frequency tables showing results (Tables 5, 6, and 7) only include the objectives so additional tables were constructed to display the remaining content (Tables 8, 9, 10, 11, 12, 13, 14, and 15). This broke the analysis of the data into two different parts. Additionally, the frequency tables do not include the content coded with “no specific code” in the CC category. This content is also included in the additional tables.

**Member-checking.** To establish greater credibility and gather additional data, the researcher engaged in a technique known as member-checking. This technique consists of taking the researcher’s interpretations of the data back to the study participants and consulting with them on their views (Creswell & Miller, 2000). The researcher reached out to each department through email to identify the professors with the most familiarity with the syllabi who would be willing to participate. Each department responded and communication was established with these

individuals, each of which were professors. After the coding process, the researcher emailed each professor the results, the codebook, and a narrative of the process. A meeting was also set for the researcher to discuss the results with each individual participant. On the day of these in-person meetings, the conversations were open-ended and largely led by the professors' thoughts on the coding process. Participants were asked if the codes made sense, if they agreed with the coding, if they would make any changes, if they had anything to add, and whether the full account was an accurate analysis of their syllabi. The conversation was recorded and transcribed and this information was then incorporated into the findings and considered during the analysis.

**Quantitative data.** The quantitative edTPA student data provided an extra layer of information to the study. Tables were constructed with the descriptive statistics of the seven identified rubrics, plus the three edTPA tasks. The means were then compared using one-way repeated measures (within-subjects) ANOVAs. This statistic was appropriate because the same students are being assessed across the different rubrics. A post hoc test was conducted if the means were shown to have significant differences.

The first question being asked by the quantitative study is "Is there a statistically significant difference between any of the local individual rubric means concerning assessment?" The null hypothesis would be "There is no statistically significant difference between any of the local rubric means." The alternate hypothesis would be "There is a statistically significant difference between the local rubric means." The second question being asked by the quantitative study is "Is there a statistically significant difference between any of the local individual task means?" The null hypothesis would be "There is no statistically significant difference between any of the local task means." The alternate hypothesis would be "There is a statistically significant difference between the local task means."

## **Reliability and Validity**

Lincoln and Guba (1985) stressed the importance of trustworthiness in any qualitative study. This study followed many of their techniques for establishing credibility. The study followed Webb's established alignment methodology that has been tested and used by researchers for many years (Webb 1997, 1999, 2005). This methodology was originally designed for the study of standard and assessment alignment, but has been adapted by others to incorporate curriculum as well. This is also not the first study to apply this methodology to assessment education as this has been done previously by DeLuca and Bellara (2013) and explained in the literature review.

The most prominent method for establishing credibility was the aforementioned member-checking technique. This technique was also used for data collection purposes, but it served to increase the trustworthiness of the document analyses as well. Lincoln and Guba (1985) described member checks as the "most crucial technique" that qualitative researchers can use to establish credibility (p. 314). These follow-up interviews also serve as a method for triangulation. Information was gathered to triangulate the data gained from the syllabi.

Another element used in this study was an audit trail, which took the form of a personal journal used by the researcher. In this journal, the researcher kept important notes on how the study progressed, the issues that arose, the critical decisions that were made, and many other entries on important aspects of the process. Throughout the study, the researcher's mind changed on many elements along the way. The journal kept a record on these instances so they could be tracked and incorporated in the discussion. The researcher also practiced reflexivity by reacting to new information that arose and being transparent about it in the journal and the study's formal discussion (Lincoln & Guba, 1985).

Several other techniques for establishing credibility and transferability were also used in this study. There were regular peer debriefings with the researcher's dissertation committee and other researchers. Also, while creating the coding scheme, the researcher compared a committee member's independent coding of material with his own initial coding to check for reliability and to make any modifications. A thick description of the analysis was also provided (Lincoln & Guba, 1985).

### **Researcher Bias**

In an effort of full disclosure, it is important to note that the researcher has both been a student in this secondary education program (though he graduated nearly ten years prior to this study) and has served as an instructional assistant professor for CI 218 (though not at the time of the study). Because of his familiarity with the program, to deal with this bias, he kept a journal to further examine the deeper aspects of the study and his thoughts on the issue. This was done by frequently logging thoughts and procedures throughout the entire research process and spending time reflecting on the process every step of the way.

### **Ethical Considerations**

Data from the edTPA did not have identifiable information included when it was delivered to the researcher. For this reason, Institutional Review Board (IRB) approval was not required to obtain the initial data.

To conduct member-checking with participants, IRB approval was granted prior to reaching out. A recruitment letter was sent to possible participants, which can be found in Appendix K. Participants were given the option of not participating without fear of any kind of consequence. Participants signed a participation consent form and an audio-recording consent form (Appendix L).

## CHAPTER V: RESULTS

This chapter presents the information gathered in the study. It begins with a review of the coding process and the findings for each of the six content areas, the standards, and the edTPA handbook. These are presented first through code frequency tables for each alignment dimension and then by tables outlining the codes found in each data source. This is followed by the edTPA statistics and a narrative describing them.

### **Document Analysis Coding**

The researcher began the coding process with several codes in mind but the list of codes organically grew and transformed throughout the study. The final CC codes were *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*. All of these codes existed in some form from the beginning except for *language and literacy*. This category emerged when the researcher saw many instances of it in three different sources. The CC codebook can be found in Appendix M. The DOK and ROK levels remained consistent with DeLuca and Bellara's (2013) study throughout with minor changes for clarification purposes. The DOK codebook can be found in Appendix N and the ROK codebook can be found in Appendix O.

During the coding process, the researcher made the decision to only fully code the learning objectives written in the syllabi. This decision was made because it is not possible to code all content with a DOK and an ROK level. These codes were designed as an analysis of the learning action, which is not something that is present within every mention of assessment in a syllabus. So, each syllabus was scanned for every mention of assessment and each mention was



coded with a CC category. Then, additionally, a DOK and ROK category was coded for each learning objective. The non-learning objectives were logged as supplementary pieces of content.

In the syllabi, seven objectives were not coded with a CC category and instead labeled “no specific code”. This happened in cases where the objective wasn’t specific enough to be labeled with a code. For example, one syllabus had the objective “Teacher candidates understand and develop the knowledge, skills, and dispositions necessary to effectively assess student work samples related to school success.” This was labeled “no specific code” because it is too broad for any of the identified codes. It was still given a DOK and ROK code.

When constructing the frequency tables, the researcher made the decision to exclude the instances where no specific code was given in the CC category. This was because including it would greatly affect the percentages in the table, and the goal of the table was to identify the codes that appear most often, and then show the alignment between sources. The instances of “no specific code” would not allow for this. Discussing the instances of “no specific code” would help answer research question #2, but it would not help at all with research question #3, which is the question to which these frequency tables are specific.

Every assessment standard indicator in the InTASC was coded in all three categories because they are naturally written as objectives. In the edTPA handbook, the central question on each rubric was transitioned to a learning objective and coded in all three categories. For example, rubric 5 says “How are the informal and formal assessments selected or designed to monitor students’ progress toward using science concepts and scientific practices during inquiry to explain or predict a real-world phenomenon?” This was coded as if it said, “Teachers will select or design informal and formal assessments to monitor students’ progress toward using

science concepts and scientific practices during inquiry to explain or predict a real-world phenomenon.” Every other mention of assessment was only coded in the CC category.

In any of the data sources, if needed, the objective was coded with more than one CC code. This is not the case for the DOK or ROK codes, as a definitive code was always identified for them.

A “Relative Representation” percentage was calculated for each code in each data source. This is the percentage of the total coded items for each individual code. The supplemental, non-learning objective content will be used in the “Discussion” section to showcase any assessment-related lessons or activities that might be taking place. This content is not included in the comparative tables below. Table 5 shows the frequency of codes across the three major coding sources (syllabi, InTASC and edTPA) and Table 6 shows the frequency of codes in each individual syllabus. Items coded with “no specific code” in the CC category were left out of both tables.

Table 5

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

Category	Code	Syllabi	InTASC	edTPA
CC	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
DOK	Level 1: Low	13	32	0
	Level 2: Moderate	42	50	71
	Level 3: High	46	18	29
ROK	Level 1: Instructional	33	0	0
	Level 2: Educational	21	91	100
	Level 3: Global	46	9	0

Table 6

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

Category	Code	Content Area					
		BIO	CHE	COM	PHY	THT	CI
CC	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
DOK	Level 1: Low	17	0	17	0	-	20
	Level 2: Moderate	33	0	50	40	-	60
	Level 3: High	50	100	33	60	-	20
ROK	Level 1: Instructional	33	50	17	80	-	0
	Level 2: Educational	33	0	33	0	-	20
	Level 3: Global	33	50	50	20	-	80

Table 7 places the data into a frequency matrix that includes all three dimensions (CC, DOK, and ROK). Each cell of the table shows the frequency percentage across all three categories, giving a clearer picture of how each data source was specifically coded. Objectives coded with “no specific code” in the CC category were left out of this table.

Table 7

*Alignment Matrix Based on Code Frequency (Expressed as Percentages)*

		Syllabi n=28 InTASC n=32 edTPA n=8			ROK					
		Instructional			Educational			Global		
CC	DOK	Syllabi	InTASC	edTPA	Syllabi	InTASC	edTPA	Syllabi	InTASC	edTPA
Alignment to learning goals	Low	0.00	0.00	0.00	0.00	3.13	0.00	0.00	3.13	0.00
	Moderate	0.00	0.00	0.00	3.57	3.13	0.00	0.00	0.00	0.00
	High	3.57	0.00	0.00	0.00	3.13	0.00	0.00	0.00	0.00
Diverse learners	Low	0.00	0.00	0.00	0.00	3.13	0.00	0.00	3.13	0.00
	Moderate	0.00	0.00	0.00	7.14	6.25	0.00	0.00	0.00	0.00
	High	3.57	0.00	0.00	0.00	3.13	12.50	3.57	0.00	0.00
Feedback and motivation	Low	0.00	0.00	0.00	0.00	6.25	0.00	0.00	0.00	0.00
	Moderate	0.00	0.00	0.00	0.00	6.25	25.00	0.00	0.00	0.00
	High	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Formative and summative	Low	0.00	0.00	0.00	0.00	3.13	0.00	0.00	0.00	0.00
	Moderate	0.00	0.00	0.00	0.00	3.13	0.00	0.00	0.00	0.00
	High	0.00	0.00	0.00	0.00	0.00	0.00	3.57	0.00	0.00
Language and literacy	Low	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Moderate	0.00	0.00	0.00	3.57	0.00	12.50	3.57	0.00	0.00
	High	0.00	0.00	0.00	0.00	0.00	0.00	3.57	0.00	0.00

*(Table Continues)*

Syllabi n=28 InTASC n=32 edTPA n=8		ROK								
		Instructional			Educational			Global		
CC	DOK	Syllabi	InTASC	edTPA	Syllabi	InTASC	edTPA	Syllabi	InTASC	edTPA
Results to guide instruction	Low	0.00	0.00	0.00	0.00	3.13	0.00	3.57	0.00	0.00
	Moderate	0.00	0.00	0.00	7.14	0.00	12.50	3.57	0.00	0.00
	High	0.00	0.00	0.00	0.00	3.13	12.50	7.14	0.00	0.00
Statistical literacy	Low	0.00	0.00	0.00	0.00	6.25	0.00	0.00	0.00	0.00
	Moderate	3.57	0.00	0.00	3.57	3.13	12.50	3.57	3.13	0.00
	High	7.14	0.00	0.00	0.00	3.13	0.00	3.57	0.00	0.00
Student self- assessment	Low	0.00	0.00	0.00	0.00	3.13	0.00	0.00	0.00	0.00
	Moderate	0.00	0.00	0.00	0.00	3.13	0.00	0.00	3.13	0.00
	High	0.00	0.00	0.00	0.00	3.13	0.00	0.00	0.00	0.00
Tools and types	Low	0.00	0.00	0.00	3.57	0.00	0.00	0.00	3.13	0.00
	Moderate	0.00	0.00	0.00	3.57	12.50	0.00	7.14	0.00	0.00
	High	0.00	0.00	0.00	0.00	3.13	12.50	7.14	0.00	0.00

## **Biology**

Only one syllabus was analyzed for the subject: BIO 305. Table 8 displays the findings. The number of instances of each code is listed in parentheses and anything added during the member check is labeled with “MC” in parentheses. This syllabus had six objectives specific to assessment and 10 pieces of supplementary content. The most common CC codes found throughout were *statistical literacy* and *alignment to learning goals*, though others were present as well. For example, the objective “Analyze assessments for effectiveness in evaluating student learning of give objectives” was coded with both *statistical literacy* and *alignment to learning goals*. All three levels of both DOK and ROK were found within this syllabus. This course also includes 40 hours of clinical experiences.

During member checking, the participant did discuss the assessment discussions that are present throughout the course. She said, “One thing we talk about consistently in this methods course is that everything is assessment. Everything we do is assessment because you have formative and summative; you have formal and informal.” She also indicated that there is a mini edTPA that the students complete, which includes the assessment task. Also, she discussed a few other places within the department where students create lesson plans that include assessment. Because the researcher did not receive those syllabi, they were not included in the study.

Table 8

*Document Analysis Findings for BIO*

Source	Type	CC Codes	DOK Codes	ROK Codes
BIO 305	Objectives (6)	<ul style="list-style-type: none"> <li>• Alignment to learning goals (1)</li> <li>• Statistical literacy (2)</li> <li>• Tools and types (1)</li> <li>• No specific code (3)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (1)</li> <li>• Level 2 (2)</li> <li>• Level 3 (3)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (2)</li> <li>• Level 2 (2)</li> <li>• Level 3 (2)</li> </ul>
	Supplementary Content (10) and Member Check	<ul style="list-style-type: none"> <li>• Alignment to learning goals (2)</li> <li>• Feedback and motivation (1)</li> <li>• Formative and summative (MC)</li> <li>• Statistical literacy (3)</li> <li>• Tools and types (1)</li> <li>• No specific code (3)</li> </ul>	N/A	N/A

**Chemistry**

Three different chemistry syllabi were coded. CHE 160 is an introductory course for assessment. One objective and one supplementary piece of content was found that referenced assessment in this course. This objective, “Design and implement instruction on an assigned chemistry topic in consideration of NGSS, including creation of an assessment”, was not specific to any CC code, was a DOK level 3, and was an ROK level 1. This course also includes clinical experiences and observations. CHE 303 contained one objective that was specific to assessment, as well as four pieces of supplementary content, including mentions in the course schedule of “assessment in chemistry” and “assessment for learning”. The objective, “Create and use appropriate formative and summative assessment tools to improve and assess student learning,



making informed decisions based on the outcome of the assessments”, dealt with several CC codes (*formative and summative, tools and types, results to guide instruction*) and was coded with DOK level 3 and ROK level 3. Much of the supplementary content was not specific to any CC code except for one that was focused on feedback and motivation. This course also includes clinical experiences and observations. CHE 304 did not contain any mention of assessment at all, mostly because it is a course students take concurrently with student teaching. It is mainly focused on supporting them throughout that process.

During the member check, the participant confirmed these codes. She did stress the presence of assessment education within CHE 303. She stated, “In 303 they create the assessments, they give the assessments, and they assess the assessments.” Table 9 displays the findings.

Table 9

*Document Analysis Findings for CHE*

Source	Type	CC Codes	DOK Codes	ROK Codes
CHE 160	Objectives (1)	<ul style="list-style-type: none"> <li>• No specific code (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 3 (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (1)</li> </ul>
	Supplementary Content (1) and Member Check	<ul style="list-style-type: none"> <li>• No specific code (1)</li> </ul>	N/A	N/A
CHE 303	Objectives (1)	<ul style="list-style-type: none"> <li>• Formative and summative (1)</li> <li>• Results to guide instruction (1)</li> <li>• Tools and types (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 3 (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 3 (1)</li> </ul>
	Supplementary Content (4) and Member Check	<ul style="list-style-type: none"> <li>• Feedback and motivation (1)</li> <li>• No specific code (3)</li> </ul>	N/A	N/A
CHE 304	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (0) and Member Check	N/A	N/A	N/A

**Communications**

Three communications courses were coded. Assessment content was found in all three of them. COM 291 included one objective dealing with assessment and two additional pieces of content. COM 293 contained one objective and eight additional pieces of content. COM 391 contain four objectives and four additional pieces of content. Clinical experiences are present throughout the courses.

These courses contain a wide variety of CC, DOK, and ROK codes. There is one instance where the way the objective was written had an impact on its DOK level. In COM 391, there is an objective that states, “To enhance preservice teachers understanding of creating assessment and using assessment data to guide teaching.” This was coded as a DOK level 1 because the preservice teachers are “understanding” how to create and use assessments. If it was written differently, it could be classified as a level 3. Because of its form, however, it remains a level 1.

During the member check, the participant indicated that the department uses an edTPA-based template for lesson planning. This contains assessment as a major aspect. She also stressed that COM 391 is heavily focused on edTPA. The students look at sample edTPA feedback and they put together a mini-edTPA throughout the course. Table 10 displays the findings.

Table 10

*Document Analysis Findings for COM*

Source	Type	CC Codes	DOK Codes	ROK Codes
COM 291	Objectives (1)	<ul style="list-style-type: none"> <li>• Statistical literacy (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 2 (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 3 (1)</li> </ul>
	Supplementary Content (2) and Member Check	<ul style="list-style-type: none"> <li>• Feedback and motivation (1)</li> <li>• No specific code (1)</li> </ul>	N/A	N/A
COM 293	Objectives (1)	<ul style="list-style-type: none"> <li>• Diverse learners (1)</li> <li>• Language and literacy (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 3 (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 3 (1)</li> </ul>
	Supplementary Content (8) and Member Check	<ul style="list-style-type: none"> <li>• Alignment to learning goals (1)</li> <li>• Language and literacy (1)</li> <li>• Results to guide instruction (1)</li> <li>• Statistical literacy (1)</li> <li>• Tools and types (1)</li> <li>• No specific code (3)</li> </ul>	N/A	N/A
COM 391	Objectives (4)	<ul style="list-style-type: none"> <li>• Diverse learners (1)</li> <li>• Results to guide instruction (3)</li> <li>• Tools and types (1)</li> <li>• No specific code (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (1)</li> <li>• Level 2 (2)</li> <li>• Level 3 (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (1)</li> <li>• Level 2 (2)</li> <li>• Level 3 (1)</li> </ul>
	Supplementary Content (4) and Member Check	<ul style="list-style-type: none"> <li>• Feedback and motivation (1)</li> <li>• No specific code (3)</li> </ul>	N/A	N/A

## Physics

Four Physics syllabi were analyzed but only one of them was coded with any assessment codes. This is not surprising because the other three courses are mostly attached to clinical experiences (such as student teaching) and they mainly serve to supplement these experiences. PHY 323 is the main methods course in the department and assessment is featured in many places throughout this syllabus. Five different course objectives in PHY 323 were coded for assessment and twelve additional pieces of assessment-related content was found in the syllabus as well. There is a variety in the codes that were assigned to this course, but *statistical literacy* leads in the CC category. Examples of this include mentions of “rubric construction”, a “test item analysis”, and “scoring and evaluating”.

During member-checking, the participant confirmed that PHY 323 is where most of the assessment is being taught within the department. Though it is not present in the syllabus and, thus, was not coded, the participant identified *formative and summative* as being a topic within the course that was covered. The participant did indicate that PHY 324 does cover the edTPA assessment task, which naturally results in assessment instruction occurring. He said, “Task 3 on the edTPA which deals with assessment is our program’s weakest task and so, starting last year, we spent a good amount of time in 324 addressing assessment, pretty much piggy-backing on what we bring out in 323 but gearing it more towards the specifics of the edTPA.” Table 11 displays the findings.

Table 11

*Document Analysis Findings for PHY*

Source	Type	CC Codes	DOK Codes	ROK Codes
PHY 215	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (0) and Member Check	N/A	N/A	N/A
PHY 323	Objectives (5)	<ul style="list-style-type: none"> <li>• Alignment to learning goals (1)</li> <li>• Diverse learners (1)</li> <li>• Results to guide instruction (1)</li> <li>• Statistical literacy (2)</li> <li>• No specific code (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 2 (2)</li> <li>• Level 3 (3)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (4)</li> <li>• Level 3 (1)</li> </ul>
	Supplementary Content (11) and Member Check	<ul style="list-style-type: none"> <li>• Alignment to learning goals (3)</li> <li>• Diverse learners (1)</li> <li>• Formative and summative (MC)</li> <li>• Results to guide instruction (1)</li> <li>• Statistical literacy (4)</li> <li>• Tools and types (1)</li> <li>• No specific code (1)</li> </ul>	N/A	N/A
PHY 324	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (0) and Member Check	N/A	N/A	N/A

*(Table Continues)*

Source	Type	CC Codes	DOK Codes	ROK Codes
PHY 357	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (0) and Member Check	N/A	N/A	N/A

## Theater

Based on the three syllabi that were analyzed, no assessment content was coded for this study. There was no mention of assessment in any of the syllabi. The courses do include clinical experiences. During member-checking, the participant did indicate a few instances where the teaching of assessment did occur, however. He said THT 186 spends time on standards, lesson planning, and backwards design. “They are learning from the first semester how to begin with the end in mind with the assessment and what they are going to do to structure their lessons to meet that.” He mentioned *formative and summative* as being a topic that is certainly discussed in that class. He also indicated that THT 286 has a mini edTPA project built in, so assessment would be covered in Task 3 of that. Also, the participant indicated that the department lesson plan is edTPA-based, so any time lesson planning occurs, students must consider assessment. Table 12 displays the findings.

Table 12

*Document Analysis Findings for THT*

Source	Type	CC Codes	DOK Codes	ROK Codes
THT 186	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (0) and Member Check	• Formative and summative (MC)	N/A	N/A
THT 286	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (0) and Member Check	N/A	N/A	N/A
THT 386	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (0) and Member Check	N/A	N/A	N/A

**Curriculum and Instruction**

Three syllabi were coded in the School of Curriculum and Instruction. CI 214 did not have any objectives specific to assessment, but did have topics in the course outline and course assignments that focused on assessment. Because this course is meant to be introductory on the topic, some of it was not coded with a categorical content code because it was purposely intended to be broad. These broad items include a mention in the course outline of the “role of assessment” and a discussion of “standardized testing.” Other items were coded with the *statistical literacy* code because of the focus on student data. These include a mention in the course outline of the “use of student data” and an assignment called the “school data analysis.”

CI 218 had two different objectives that were coded in assessment and seven additional pieces of content were identified to cover assessment as well. The objectives were both relatively



broad, resulting in both being coded with an ROK level 3. One was not given a CC code because it was too broad and the other was coded with three different CC codes. The supplementary content deal with formative and summative assessment (on two different pieces of content), tools and types, results to guide instruction, and statistical literacy. A few other assessment-related pieces were identified but they were too broad to be coded with any specific code.

Three course objectives and one piece of supplementary content were coded for assessment in CI 221. The three course objectives were coded with tools and types, *language and literacy*, and *diverse learners*. All were a DOK level 2 and two were an ROK level 3 while one was an ROK level 2. The supplementary piece was a student task in the assignment section where students create an assignment. It was broad so it was not coded with any specific CC code.

During member checking, the participant agreed with the coding. He expressed the extensive amount of clinical experiences in the three course and how they contain assessment aspects. For example, in CI 214's clinical experiences, there are five themes students are supposed to be watching for and one is assessment. He also mentioned that CI 214 and 218 both dive into formative and summative assessment extensively. This code was not found in the CI 214 syllabus. Table 13 displays the findings.

Table 13

*Document Analysis Findings for CI*

Source	Type	CC Codes	DOK Codes	ROK Codes
CI 214	Objectives (0)	N/A	N/A	N/A
	Supplementary Content (4) and Member Check	<ul style="list-style-type: none"> <li>• Formative and summative (MC)</li> <li>• Statistical literacy (2)</li> <li>• No specific code (2)</li> </ul>	N/A	N/A
CI 218	Objectives (2)	<ul style="list-style-type: none"> <li>• Results to guide instruction (1)</li> <li>• Statistical literacy (1)</li> <li>• Tools and types (1)</li> <li>• No specific code (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (1)</li> <li>• Level 3 (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 3 (2)</li> </ul>
	Supplementary Content (7) and Member Check	<ul style="list-style-type: none"> <li>• Formative and summative (2)</li> <li>• Results to guide instruction (1)</li> <li>• Statistical literacy (1)</li> <li>• Tools and types (1)</li> <li>• No specific code (2)</li> </ul>	N/A	N/A
CI 221	Objectives (3)	<ul style="list-style-type: none"> <li>• Diverse learners (1)</li> <li>• Language and literacy (2)</li> <li>• Tools and types (2)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 2 (3)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 2 (1)</li> <li>• Level 3 (2)</li> </ul>
	Supplementary Content (1) and Member Check	<ul style="list-style-type: none"> <li>• No specific code (1)</li> </ul>	N/A	N/A

## InTASC Standard 6

Because of their obvious nature, the InTASC standard 6 indicators are all written as objectives, so nothing was identified as a supplementary piece of content. There are 22 indicators. *Language and literacy* was the only CC code not found in the standard. All of the rest of the CC codes were fairly evenly distributed, with *statistical literacy* and *tools and types* being the most prevalent but only each making up 18% of the total. All DOK levels were found and, as for ROK codes, only levels 2 and 3 were found (with the most being coded as ROK level 2). The full standard and indicators can be found in appendix B. Table 14 displays the findings.

Table 14

### *Document Analysis Findings for InTASC Standard 6*

Source	Type	CC Codes	DOK Codes	ROK Codes
InTASC Standard 6	Objectives (22)	<ul style="list-style-type: none"> <li>• Alignment to learning goals (4)</li> <li>• Diverse learners (5)</li> <li>• Feedback and motivation (4)</li> <li>• Formative and summative (2)</li> <li>• Results to guide instruction (2)</li> <li>• Statistical literacy (6)</li> <li>• Student self-assessment (4)</li> <li>• Tools and types (6)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 1 (7)</li> <li>• Level 2 (11)</li> <li>• Level 3 (4)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 2 (20)</li> <li>• Level 3 (2)</li> </ul>
	Supplementary Content (0)	N/A	N/A	N/A

## **edTPA Science Assessment Handbook**

Most of what was identified within the edTPA Science Assessment Handbook was considered supplementary except for the main questions within the seven rubrics that dealt with assessment. These main questions were transitioned into objective statements and coded that way. So, for example, Task 3, Rubric 14's main question was "How does the candidate analyze students' use of language to develop content understanding?" This was coded as "The candidate analyzes students' use of language to develop content understanding."

Six different CC codes were identified within the objectives and seven were identified within the supplementary pieces of content. All objectives were either coded with DOK level 2 or 3 and all were coded with ROK level 2. The assessment-related rubrics can be found in appendices D through J. Table 15 displays the findings.

Table 15

*Document Analysis Findings for edTPA*

Source	Type	CC Codes	DOK Codes	ROK Codes
edTPA Science Assessment Handbook	Objectives (7)	<ul style="list-style-type: none"> <li>• Diverse learners (1)</li> <li>• Feedback and motivation (2)</li> <li>• Language and literacy (1)</li> <li>• Results to guide instruction (2)</li> <li>• Statistical literacy (1)</li> <li>• Tools and types (1)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 2 (5)</li> <li>• Level 3 (2)</li> </ul>	<ul style="list-style-type: none"> <li>• Level 2 (7)</li> </ul>
	Supplementary Content (25)	<ul style="list-style-type: none"> <li>• Alignment to learning goals (3)</li> <li>• Diverse learners (3)</li> <li>• Feedback and motivation (4)</li> <li>• Language and literacy (4)</li> <li>• Results to guide instruction (5)</li> <li>• Statistical literacy (4)</li> <li>• Tools and types (1)</li> <li>• No specific code (4)</li> </ul>	N/A	N/A

**edTPA Data**

The local edTPA data that were examined only includes students within the secondary subjects highlighted in this study: biology, physics, chemistry, theater, and communications. The decision to present the numbers this way was made to better align the quantitative portion of the

study with the qualitative portion. Also, it is important to note that these scores are the initial (first chance) edTPA scores. The final scores would be slightly different. This decision was made to better capture how local students are actually performing prior to the opportunity to revise their work. There were only three cases (one in spring 2016 and two in spring 2017) where students did not pass on the first try and, thus, were given a chance to revise.

In only one instance, a student received an error code on rubric 11 of the fall 2016 assessment because there was a problem with the evidence and the rubric was not assigned a score. Basically, this can be viewed as the equivalent to an incomplete grade. A composite score was still given for the student by adding up the 14 remaining rubric scores and the student still received a passing score. For the purposes of this study, this student’s scores will be viewed in the same way that the university views them: by replacing the error code with a zero. This will affect the values for rubric 11 and the assessment task (which is the sum of rubrics 11 through 15).

Table 16 displays the content being assessed in each rubric.

Table 16

*edTPA Rubric and Content Being Assessed*

Rubric	Content
Rubric 5	Planning assessments to monitor and support student learning
Rubric 10	Analyzing teaching effectiveness
Rubric 11	Analysis of student learning
Rubric 12	Providing feedback to guide learning
Rubric 13	Student understanding and use of feedback
Rubric 14	Analyzing students’ language use and [subject] learning
Rubric 15	Using assessment to inform instruction

Table 17 displays the descriptive statistics for the scores. Rubrics 5, 10, 11, 12, 13, 14, and 15 are displayed because they aligned with the qualitative portion of the study. These are the rubrics that were identified as having assessment-related content.

Table 17

*Descriptive Statistics for the edTPA Assessment-Related Rubrics*

Rubric	Mean Score	Standard Deviation
Rubric 5	3.08	0.54
Rubric 10	2.79	0.62
Rubric 11	3.13	0.83
Rubric 12	3.48	0.74
Rubric 13	2.90	0.65
Rubric 14	2.99	0.72
Rubric 15	2.93	0.66

N=60

A one-way repeated-measures (within-subjects) ANOVA was conducted with the factor assessment rubrics (treatment) and the dependent variable being the rubric scores. For a one-way repeated-measures (within-subjects) ANOVA, the overall results based on the multivariate tests indicate a significant treatment effect, Wilks'  $\Lambda = 0.54$ ,  $F(6, 54) = 7.73$ ,  $p < .001$ ,  $\eta^2 = 0.46$ . A post hoc test (Bonferroni) was conducted to show the pairwise comparisons between rubrics. Twenty-one unique pairwise comparisons were conducted among the means for rubric 1, rubric 2, rubric 3, rubric 4, rubric 5, rubric 6, and rubric 7. Six of the twenty-one pairwise comparisons are significant, controlling for familywise error rate across twenty-one tests at the .05 level, using the Bonferroni procedure. Below is a summary that shows the significant differences between rubrics:

- Rubric 5 has a significantly higher mean than rubric 10 (mean difference = 0.29,  $p = .015$ )

- Rubric 12 has a significantly higher mean than rubric 5 (mean difference = 0.39,  $p=.008$ )
- Rubric 12 has a significantly higher mean than rubric 10 (mean difference = 0.68,  $p<.001$ )
- Rubric 12 has a significantly higher mean than rubric 13 (mean difference = 0.58,  $p<.001$ )
- Rubric 12 has a significantly higher mean than rubric 14 (mean difference = 0.48,  $p=.002$ )
- Rubric 12 has a significantly higher mean than rubric 15 (mean difference = 0.54,  $p<.001$ )

The null hypothesis was rejected and the alternate hypothesis was accepted. The alternate hypothesis was, “There is a statistically significant difference between the local rubric means.”

The full post hoc test results can be found in Appendix P.

The three edTPA tasks were also analyzed. The planning task is the sum of rubrics 1 through 5, the instruction task is the sum of rubrics 6 through 10, and the assessment task is the sum of rubrics 11 through 15. A reliability test was conducted on each of the three tasks. Cronbach’s alpha reliability coefficient for the planning task was .73, for the instruction task was .81, and for the assessment task was .73. The three tasks were found to be reliable.

The descriptive statistics for the three tasks are displayed in Table 18.



Table 18

*Descriptive Statistics for the edTPA Tasks*

Task	Mean Score	Standard Deviation
Planning Task	15.28	2.06
Instruction Task	15.06	2.45
Assessment Task	15.43	2.50

N=60

A one-way repeated-measures (within-subjects) ANOVA was conducted with the factor tasks (treatment) and the dependent variable being the task scores. For a one-way repeated measures (within-subjects) ANOVA, the overall results based on the multivariate tests do not indicate a significant treatment effect, Mauchly's Test of Sphericity = 0.57,  $F(2, 118) = 0.75$ ,  $p = .473$ ,  $\eta^2 = 0.013$ . The null hypothesis was accepted.

## CHAPTER VI: DISCUSSION

This chapter will discuss and analyze the results presented in chapter 5. This discussion and analysis will be guided by the three research questions for the study, each of which will be presented as a section heading. The discussion will stay focused on the research question at hand until the final cross-question discussion section which will analyze the conclusions drawn from the entire study.

As an aid to the reader, this chapter will begin with a restatement of the problem and a review of the methodology.

### **Restatement of the Problem**

Literature on the topic frequently shows teacher assessment literacy to be at a level lower than desired. Both qualitative and quantitative data often indicate that teachers do not possess the necessary skills to collect data on their students' learning, analyze these data, provide feedback to the students and other stakeholders, and use the data to inform their future instruction. Teaching these skills can be very difficult and this study set out to examine how it is being done in one university's secondary teacher education programs.

### **Review of the Methodology**

The study was centered on secondary teacher education programs at a large, Midwestern university. It attempted to answer three research questions by undertaking a mixed-methods design. A qualitative alignment study was conducted that examined syllabi from the methods courses for five different secondary subjects and the School of Curriculum and Instruction, along with standard 6 of the InTASC teaching standards and the science assessment handbook of the edTPA. All documents were coded based on a developed assessment codebook and then examined for alignment with the others. Additionally, the university's edTPA data for the

selected subjects was examined to identify the preservice teachers' performance in certain assessment-related areas.

### **Study Discussion**

All of the CC codes except one existed in some form from the beginning of this study. They were initially created based on important concepts identified within the literature on the topic. The code *language and literacy* is the only code that was completely discovered during the actual study. This code can be found in the COM syllabi, the CI syllabi, and in the edTPA science assessment handbook.

The *language and literacy* code was initially discovered when elements of it appeared in the edTPA handbook. As an example, rubric 14 contains this overarching question: "How does the candidate analyze students' use of language to develop content understanding?" After much deliberation, it was decided that this did not fit into any of the other established categories. Then, when similar objectives were observed in other sources (mainly the syllabi), it was clear a new category had been discovered.

Though it was not originally identified as a CC code, there is research that points to its importance. Language assessment literacy (LAL) is an important term in the linguistics field (Giraldo, 2018; Herrera & Macias, 2015; Inbar-Lourie, 2008; 2013; Scarino, 2013; Yan, Zhang, & Fan, 2018). LAL is essentially the ability to assess language usage and learning. This idea could expand beyond those tasked with helping others learn a language and could be applied to the language used in everyday learning. This is how this code is conceptualized here.

## **Research Question 1: Where is Assessment Being Taught to Undergraduate Secondary Education Preservice Teachers?**

This question was addressed primarily through a document analysis of the methods course syllabi. Prior to discussing these results, it is crucial to keep in mind that a syllabus does not necessarily capture everything that happens within a course. Examining a course in full would be much more time-consuming and involve observations, interviews, and other methods throughout the entire semester. Though the member-checking process helps fill in the gaps of the syllabus, the researcher still considers this examination to be somewhat incomplete.

All of the assessment instruction in this program happens in the course of regular classes. There is not an assessment-specific course in the observed college of education. In Mandinach et al. (2013), their survey of teacher preparation programs found that 62.4% of them indicated that they offer at least one course that is primarily focused on teaching data use to inform instructional decisions. This specific program would not fall into that group. Impara et al. (1993) and Marso and Pigge (1993) both indicated that an assessment course might not have a significant impact on assessment literacy. More recent studies on this topic would provide a more valuable viewpoint and are recommended for future study.

Within the examined syllabi, there were 24 assessment-related objectives found. Of those, six (25%) were in BIO, six (25%) were in COM, five (21%) were in PHY, five (21%) were in CI, two (8%) were in CHE, and none (0%) were in THT. Assessment education was found to appear somewhere in all examined courses except for THT. During member-checking, the participant did indicate some emphasis that is put on *formative and summative* assessment, but this was not found in the syllabus.

BIO is the subject that is tied for the largest number of assessment objectives, yet it is the subject with the fewest number of examined syllabi. They only require one methods course for education students (though during member-checking, the participant did stress that lesson plans were created in other, non-analyzed courses as well). It is tied for the most assessment objectives with COM, which is drawing from three syllabi instead of one. In the BIO syllabus, there are 20 total objectives listed and six of them (30%) were identified as relating to assessment. Within the current study, it was not possible to examine the impact this has on their students in comparison to students in other subjects, though this examination would be an interesting future research study.

Outside of the main course objectives, assessment was found somewhere in each of the subjects, either through supplementary pieces of content or indicated by the participant during member-checking. The codes *formative and summative*, *statistical literacy*, and *tools and types* were found in four different subjects, *alignment to learning goals*, *feedback and motivation*, and *results to guide instruction* were found in three different subjects, and *diverse learners* and *language and literacy* were each found in one subject.

It could be argued that the CI syllabi carry more weight in this study than the others because these courses are taken by every student in the secondary teacher education programs. If one were to subscribe to that argument, it would be important to note that the CI syllabi were not coded at all with the following CC codes: *alignment to learning goals*, *feedback and motivation*, and *student self-assessment*. That would mean that *alignment to learning goals* is never exposed to students in CHE or THT, *feedback and motivation* is never exposed to students in PHY or THT, and *student self-assessment* is never exposed to students in any of the studied subjects.

As mentioned previously, it is important to consider that this study did not examine clinical experiences. Any assessment skills that are learned during these opportunities for observations or teaching were not coded in this study. Though DeLuca and Klinger (2010) argue that direct instruction is necessary in assessment education, there are many that stress the importance of clinical experiences (Dorfman et al., 2006; Kelting-Gibson et al., 2013; Lian et al., 2014). This study was not able to capture the impact of clinical experiences on the assessment literacy of local preservice teachers.

### **Research Question 2: How Closely Does Assessment Instruction Align with Assessment Practices as Defined by InTASC and the edTPA?**

To answer this research question, a document analysis of all three sources (syllabi, standards, assessment handbook) was conducted and results were compared. Studying the alignment of intended, enacted, and assessed curriculum requires a consideration of the nature of the sources. It would be expected for the syllabi writers to make an effort to align to the standards. It would also be expected for the syllabi to align with the assessment because the education programs want to see their students succeed. So, when in this frame of mind, the instances of misalignment become especially noteworthy.

**Categorical concurrence.** The CC code with the closest alignment across sources is *diverse learners*, being 14% in the syllabi, 15% in the InTASC, and 13% in the edTPA. This shows the equal importance that all three sources put on being prepared to assess the wide range of learning styles within a classroom. Considering the importance of Individualized Education Programs (IEPs), Response to Intervention (RtI), and other aspects designed to help each and every student, this discovery fits in to these education priorities.

One of the CC codes with the least alignment across sources is *feedback and motivation*, with a 0% share of the syllabi, 12% of the InTASC, and 25% of the edTPA. This wide range is very interesting, especially noting the difference between the syllabi and the edTPA. Often the program designs their courses to prepare students for the edTPA but this is an occasion where the objectives do not match up.

There were two CC codes that did not show up in any of the syllabi's course objectives: *feedback and motivation* and *student self-assessment*. *Feedback and motivation* does show up quite a bit in the supplementary pieces of content within the syllabi, but *student-self assessment* was not coded anywhere within the analyzed courses. Hattie (2009)'s list of the biggest factors related to student achievement labels students self-reporting their grades as one of the top factors on the list. Its effect size cannot be ignored. Four different indicators within standard 6 of the InTASC standards were coded with this code, showing the importance they place on it. This code does not appear at all in the edTPA assessment handbook, however. No objectives or supplementary pieces of content were coded with this code. The intention to align with the edTPA might partially explain its absence in the syllabi.

The *formative and summative* CC code was only applied to the objectives in two data sources: the InTASC standard 6 indicators and the chemistry syllabi. It is only 4% in the syllabi, 6% in the InTASC, and 0% in the edTPA. It does appear more often in the various pieces of supplementary content from the other data sources and it was the most commonly inserted code from the member-checking process. This code might be considered a "building block" for assessment education to some because, to understand many of the other aspects of assessment, one must understand the basic types of assessment that exist. When thinking of it this way, it is easier to understand why it was the most commonly inserted code during member-checking.

Also, during member-checking, it was common for the participant to indicate that this aspect is taught during their earliest methods course where students learn the basics of assessment. This would explain its absence in many of the syllabi, but does not explain its absence in others.

The consensus among the literature is that formative assessment is crucial in education, yet it is underused and often misunderstood (Black & Wiliam, 1998; Buck et al., 2010; Buyukkarci, 2014; Frey & Smith, 2010; Gunn & Gilmore, 2014; Jones, 2014; Poth, 2013; Smith & Galvin, 2014; Yao, 2015). This study shows that these sources do not put a large amount of emphasis on the *formative and summative* code. This includes it not being coded at all in the edTPA. This might be explained by the difficulty to evaluate this concept through a performance assessment like the edTPA. Often, formative assessment is done quickly and without any documentation, so it could be hard to capture in this measure. It also might be explained by the *results to guide instruction* code being frequently found in the edTPA. This code focuses on using the data gathered during the formative process. It is a more tangible idea that can be examined using the edTPA. The scorers might use this concept to assess whether preservice teachers understand the purpose of formative assessment.

The *results to guide instruction* code was featured heavily in the syllabi (21%) and the edTPA (25%) but not as much in the InTASC standard 6 (6%). In fact, of all the CC codes, *results to guide instruction* is tied as the second-least appearing code within InTASC. Interestingly, the code it is tied with is *formative and summative*. Formative assessment also heavily deals with using assessment to inform instruction.

As mentioned previously, *language and literacy* arose during this study. It was identified enough in the documents that it was clear a new code had been discovered. Many of the reviewed literature reviews and frameworks did not place the same level of importance on this



topic as they did on other assessment literacy aspects. The InTASC standards follow this pattern by not including this topic in its recommended assessment skills.

Though this code was not found in the standards, it did represent 13% of the edTPA objectives and 11% of the syllabi objectives. Of the syllabi, it represents 13% of the COM objectives and 25% of the CI objectives. It is not surprising that this code was found in COM because of the subject's emphasis on language, but the large focus in CI is somewhat surprising. It could easily be theorized that this code would appear in the university's English and foreign language syllabi as well, though this was not examined in the present study.

The InTASC standard 6 indicators are relatively balanced in regard to the CC codes. Every code appears at least twice except for *language and literacy* which does not appear at all. This balance is not surprising due to the nature of the standards and their purpose. It is also not surprising that *language and literacy* does not appear because these standards were heavily considered when developing the original list of codes.

**Depth of knowledge.** The syllabi objectives were mostly coded with the *moderate* (42%) and *high* (46%) DOK levels with only 13% of objectives being coded at the *low* level. This suggests that the courses are asking preservice teachers to engage in higher order thinking while learning about assessment. This contrasts with the InTASC standard 6 indicators which were more heavily coded at the *low* (32%) and the *moderate* (50%) levels with the *high* level at only 18%. The edTPA does not put any emphasis on the *low* level (0%) but instead is heavy on the *moderate* level (71%) with the *high* level making up 29%.

**Range of knowledge.** Somewhat surprisingly, only one-third of syllabi objectives are written at the *instructional* ROK level (33%), which falls behind the *global* ROK level (46%). This indicates that the majority of course assessment objectives are written broadly, not getting

very specific for the exact skills students should be learning. It is clear from the coding process that both the InTASC standard 6 indicators and the edTPA Science Assessment Handbook put the emphasis on the *educational* ROK level. It was coded in 91% of the InTASC objectives and 100% of the edTPA objectives. This level, however, is coded the least within the syllabi (BIO: 33%, CHE: 0%, COM: 33%, PHY: 0%, CI: 20%).

**Alignment matrix.** The alignment matrix (Table 7) does suffer from a small sample size (syllabi n=28, InTASC n=32, edTPA n=8). This certainly must be considered when analyzing the data. Relative alignment can be seen often but there are a few places where apparent misalignments should be highlighted.

For the *tools and types* code with a moderate DOK and an educational ROK, there does seem to be a misalignment between the three data sources. The syllabi are at 3.57% (1 out of 28), the InTASC is at 12.50% (4 out of 32) and the edTPA is at 0.00% (0 out of 8). Likewise, in the *diverse learners* code with a moderate DOK and an educational ROK, the syllabi are at 7.14% (2 out of 28), the InTASC is at 6.25% (2 out of 32) and the edTPA is at 0.00% (0 out of 8). These are both areas where the other two sources outweigh the edTPA in their instructional priorities. Because of its nature, every edTPA objective was coded as an educational ROK so it is worth noting that it is missing in these two instances. In the case of *tools and types* with moderate DOK and educational ROK, this is InTASC's biggest representation for the whole study. This misalignment with edTPA is interesting and might be a possible explanation for the lower emphasis in this realm for the syllabi. This logic does not hold true for *diverse learners*, moderate DOK, educational ROK, however. Here the InTASC and syllabi are in close alignment even though they do not match up with the edTPA.

*Feedback and motivation*, moderate DOK, educational ROK is one area that is emphasized in all data sources except for the syllabi (syllabi = 0.00%, InTASC = 6.25%, edTPA = 25.00%). As mentioned previously, this gap in the syllabi is concerning, especially when considering its emphasis in other data sources.

Several areas are emphasized in one data source but none of the others. This is true for *statistical literacy*, high DOK, instructional ROK (syllabi = 7.14%, InTASC = 0.00%, edTPA = 0.00%); *statistical literacy*, low DOK, educational ROK (syllabi = 0.00%, InTASC = 6.25%, edTPA = 0.00%); *results to guide instruction*, high DOK, global ROK (syllabi = 7.14%, InTASC = 0.00%, edTPA = 0.00%); *tools and types*, moderate DOK, global ROK (syllabi = 7.14%, InTASC = 0.00%, edTPA = 0.00%); and *tools and types*, high DOK, global ROK (syllabi = 7.14%, InTASC = 0.00%, edTPA = 0.00%).

There are no categorical concurrence themes that align perfectly between the syllabi and the standards. This is an interesting find because of the university's desire to align their syllabi to the standards. For complete alignment, the two sources would need to match in all three categories. This finding might warrant a further examination of the university's alignment processes in general.

### **Research Question 3: How Do Secondary Education Preservice Teachers Perform in the Realm of Student Assessment as Measured by the edTPA?**

The researcher identified seven edTPA rubrics that contained assessment elements. Five of them (rubrics 11, 12, 13, 14, and 15) constitute what edTPA calls the assessment task. The other two (rubrics 5 and 10) were identified as containing assessment elements within them. To answer the research question, descriptive statistics were analyzed for each rubric and each task and then a one-way repeated measures (within-subjects) ANOVA was conducted for the rubrics

and again for the tasks. It is important to remember that this is a teacher performance assessment, not an assessment literacy measurement tool like those discussed by Brown (2004; 2006; 2007; 2008a; 2008b; 2009), Mertler and Campbell (2005), Plake and Impara (1992), or Wayman et al. (2016). The edTPA was created for a separate purpose and this should always be considered when analyzing results.

The mean scores for all rubrics hover around 3.00, which is predictable due to that score being the center of the continuum. Rubrics 5, 11, and 12 are slightly above 3.00 and rubrics 10, 13, 14, and 15 are slightly below. These means cannot be assumed to be significant, however. This is why the one-way repeated measures (within-subjects) ANOVA was conducted. This test identified if there was a significant difference within this group and where specifically the significant differences existed. It was found that there was a significant difference within the means of this group.

It is interesting to note that the two rubrics that address feedback, rubric 12 (providing feedback to guide learning) and rubric 13 (student understanding and use of feedback) have a significant mean difference. Rubric 12 has a significantly higher mean than rubric 13 (mean difference = 0.58,  $p < .001$ ). To receive a high score on rubric 12 (found in appendix G) the preservice teacher must demonstrate specific feedback to students on both their strengths and their weaknesses. To receive a high score on rubric 13 (found in appendix H) the preservice teacher must demonstrate how he/she will support students in understanding and using this feedback. This shows a disconnect between those two skills. Preservice teachers from this university are better at providing feedback than they are at working with students to use the feedback. Interestingly, rubric 12's mean is significantly higher than five of the six other

analyzed rubrics. The only one it is not significantly higher than is rubric 11 (analysis of student learning), which has the second highest mean score.

The lowest mean score belongs to rubric 10 (analyzing teaching effectiveness), which is significantly lower than both rubric 5 and rubric 12. This rubric focuses on examining student learning patterns to adjust teaching practices. To do this well, candidates must bring in research or theory to justify their changes.

In addition to examining the assessment-related rubrics, the researcher also examined the three tasks on the edTPA. The assessment task had the highest mean score (15.43) followed by the planning task (15.28) and then the instruction task (15.06). A one-way repeated measures (within-subjects) ANOVA showed no significant difference in these means, however. Though it is the highest mean, it cannot be said that the assessment task is significantly higher than either of the other two.

So, according to these results, local students are not necessarily doing better or worse in the realm of assessment in comparison to the other two realms (planning or instruction). This is an important finding. Though the literature frequently indicates that assessment literacy is at a low level, this study offers no evidence to support that theory.

The results were not compared to external data (state or national means) for several reasons. Firstly, the researcher was not able to obtain the complete data sets for the state or national means, which prevents any high-level statistical analysis. Secondly, the researcher was not able to separate only the data from the five identified subjects from the rest of the state or national data, making the comparison disjointed. Thirdly, the state and national data are only available for final attempts, not initial attempts, which was the focus of this study. Because of these reasons, the local data could not be examined in comparison to the results of students

outside of this specific university. If these barriers could be overcome, this would add another very interesting layer of information to future studies.

### **Cross-Question Discussion**

As discussed in the literature review, there is often a concern that preservice teachers are leaving their education programs with a good theoretical understanding of assessment but this does not necessarily translate to good habits in the classroom (DeLuca, 2012; DeLuca & Lam, 2014; Siegel, 2013; Siegel & Wissehr, 2011). Wallace and White (2015) discussed how teachers in their study generally focused on *how to assess* before they learned *why to assess*. The literature often claims that preservice teachers are not able to tie it all together. Many researchers call for more practical experience to help preservice teachers put their learning into action (Dorfman et al, 2006; Kelting-Gibson et al., 2013; Lian et al., 2014). It may not be highlighted much within this study, but it is incredibly important to note that practical experiences do occur throughout these courses and throughout the program as a whole. This study does not allow for a discussion of whether the practical experiences are beneficial for assessment education or whether they are enough, but they do play a big role within the program.

Though this study does nothing to support a discussion of the impact of practical experience, the edTPA does attempt to capture practical understanding of assessment by taking the form of a performance assessment. The edTPA asks preservice teachers to submit artifacts and write-ups from their experiences in the classroom. It could be debated how accurate a picture the scorers get during this process, however. Without utilizing observations, interviews, or actually getting to know the individuals, it is always difficult to get a true understanding of what they know. This must be kept in mind throughout the discussion of the edTPA results in this study.

One way this study can contribute to this discussion is to look at the DOK levels. Just how deep is the knowledge that the courses are trying to impart? A high DOK level could support the idea that preservice teachers are getting that practical experience necessary for deeper learning. The document analysis indicated that 46% of the syllabi assessment objectives were written at a high DOK and another 42% were written at a moderate DOK. Only twice was an objective coded with a low DOK level, once with the *results to guide instruction* code and once with the *tools and types* code.

Several CC codes (and the actual skills they represent) lend themselves to being difficult to assess through measures like the edTPA. For example, the code *formative and summative* is not one that can easily be measured in regard to preservice teachers' theoretical vs. practical understanding. This code is centered around the distinction between the two types of assessment. It was only found in one syllabus (CHE).

To be clear, the *formative and summative* code was barely found in the other sources either, which might indicate the need to combine this code with others. There are other codes that center around aspects of formative assessment, such as *results to guide instruction* or *feedback and motivation*. The decision was made to keep them separate because the *formative and summative* code is focused on teaching preservice teachers the difference between formative and summative assessment. This includes explicit teaching of what formative assessment is and how it is different from the more commonly known summative assessment. What this study showed is that this explicit lesson is not appearing in any of the sources, but many aspects of formative assessment are.

The CC code *results to guide instruction* is tied for the highest frequency in both the syllabi and in the edTPA. This study indicates that it is taught frequently and it is assessed

frequently. Yet, according to the quantitative study, the rubrics that this code most frequently appeared in (rubric 10 and rubric 15) both had lower mean scores. They were not both significantly lower than all of the rubrics, but rubric 10 (analyzing teaching effectiveness) was significantly lower than rubric 5 and rubric 12 while rubric 15 (using assessment to inform instruction) was significantly lower than rubric 12. Part of this can be credited to rubrics 5 and 12 having especially high means (particularly rubric 12) but it also cannot be ignored that there is a pattern between the two *results to guide instruction* rubrics. It should also be pointed out that, as seen in the alignment matrix (Table 7), *results to guide instruction* is one of only two CC codes that was coded with a low DOK level (in one instance) and that the number of instances of it being coded with a high DOK level (two) lags behind the number of instances of it being coded with a moderate DOK level (three). So, when considering the importance of teaching practical knowledge, only 33.3% of the objectives coded with *results to guide instruction* were also coded with a high DOK level.

The lowest mean score belongs to rubric 10 (analyzing teaching effectiveness), which is significantly lower than both rubric 5 and rubric 12. Rubric 10 was coded with the CC codes *results to guide instruction* and *diverse learners*, the DOK level 3, and the ROK level 2. When looking at the alignment matrix (Table 7), these combinations (*results to guide instruction*, DOK level 3, and ROK level 2; *diverse learners*, DOK level 3, ROK level 2) can both be found in the InTASC standards, but neither can be found in the syllabi.

To return to the discussion of the *formative and summative* code, another possible merging partner would be the *feedback and motivation* code. A full discussion of the *feedback and motivation* code is a fascinating one. This code was not found in a single syllabus objective. There are two edTPA rubrics that address feedback (rubrics 12 and 13) and they were shown to



be significantly different from one another. Rubric 12 (providing feedback to guide learning) has the highest mean score of the analyzed rubrics, being significantly higher than every other rubric except rubric 11. Local students are performing significantly better on this rubric than all but one other rubric (rubric 11), yet the document analysis indicated that this topic is not being covered very heavily.

Because rubric 13 (student understanding and use of feedback) is significantly lower than rubric 12, it is possible that this is where the lack of instruction on the topic becomes problematic. Hattie (2009) does not separate the ideas of providing feedback to guide learning and student understanding and use of feedback. One might look at the first and consider it incomplete until the second has also been mastered. Providing feedback is important, but it does not accomplish much unless students are supported in understanding and using it.

Another interesting discussion is the coding of the ROK levels and the alignment of these codes among the various sources. ROK looks at the scale of the objective and what the learner will be able to do once it is completed. As the ROK level gets higher (*global* being the highest), the time it takes to master the objective will most likely get longer. Surprisingly, the majority of the syllabi assessment objectives (46%) are written at a *global* level. One would expect the objectives of a specific course to be more *instructional* or *educational*. This does not match up with the standards or the edTPA which are heavily written at the *educational* level (91% and 100% respectively). This is a clear area of misalignment. One could wonder what impact a better alignment in this area might have on local edTPA scores.

One thing that was not specifically found in this study is the sociocultural context needed for proper assessment as described by Willis et al. (2013). They stated, “Assessment literacy is a dynamic context dependent social practice that involves teachers articulating and negotiating

classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students” (p. 2). This concept was not identified through the coding process. This is not to say it does not exist at all in these sources, but it was not prevalent enough to emerge as a new code. As Willis et al. (2013) stress, teachers must learn how to connect the dots between their own cultural knowledge of all stakeholders and the actual assessment practice. Not making this connection can result in a limited understanding of the knowledge that students have gained.

## CHAPTER VII: CONCLUSIONS

Though all secondary education teachers take the courses that are required within the CI and EAF departments, they all take different courses depending on their content area. As seen in this study, the assessment methods that are taught within these content-area courses can vary widely. All students are at least exposed to the five assessment objectives that are found within the CI methods courses, but after that it will vary from zero to six more, depending on their subject (at least according to the subjects in this study). The current study did not allow for an examination of the impact of this variation in assessment instruction.

This study examined the alignment of assessment content, DOK, and ROK among the syllabi, the InTASC assessment standard and indicators, and the edTPA. There were multiple areas of misalignment. One notable example is the absence of the code *feedback and motivation* in the syllabi and its strong emphasis in the edTPA. Another notable area of misalignment is in ROK, where the syllabi do not match up with the levels of the standards or the edTPA.

The study also examined the local edTPA results for the preservice teachers who have completed the observed education program. The results show that preservice teachers are generally performing no better or worse on the assessment task than they are on the other two tasks (planning and instruction). It was also noted that the edTPA rubric 12 (Providing Feedback to Guide Learning) had a statistically higher mean score than many of the other rubrics. Considering the absence of the *feedback and motivation* code in the syllabi, this is a fascinating finding. The other rubric that deals with feedback, rubric 13 (Student Understanding and Use of Feedback) had a statistically lower mean than rubric 12, which might indicate that preservice teachers are better at providing students with feedback than they are at supporting them as they use the feedback to impact their learning.

## Implications for Practice

This study was situated in the theoretical framework, FAST (Figure 1 and Table 1). It examines the prevalence of the “what teachers should know” components and used them as the CC codes during the coding process. This particular study only looks at the university setting in the “where it should be learned” realm, leaving the K-12 school setting for a future study. The discussion of theoretical vs. practical knowledge and the examination of the DOK and ROK levels falls into the “how it should be learned” realm, though this study only addresses the tip of the iceberg in that area. Future studies could dive into the “how it should be learned” realm at a much deeper level.

It should be noted that further discussion must be had around the *formative and summative* code. This code was not found very often and this could very likely be because of the need to combine codes. The distinction between formative and summative assessment is not often explicitly seen, yet many aspects of formative assessment are, which could indicate that it is taught in conjunction with these other ideas. *Results to guide instruction* and *feedback and motivation* would be two candidates for combination with this code.

The study indicated that the CC code *feedback and motivation* needs to be taught at a deeper level than it currently is. The edTPA results show that preservice teachers can successfully give feedback, but they struggle in supporting students in using it. This is evidence that action should be taken.

From this study, conversations can begin in the “how it should be learned” tier. It is important to consider the DOK and ROK levels of the standards and edTPA. This study indicated that there is not strong alignment, especially at the ROK level. The syllabi writers should consider ROK when writing their objectives and activities. Bringing these objectives

down to the *instructional* or *educational* levels might give them more direction for what should be accomplished.

To use the FAST framework, educators can begin by asking the three important questions: “What should teachers know?”, “Where should it be taught?”, and “How should it be taught?” Educators could use this study to begin answering those questions and then move forward to examine their current situations and discuss any gaps or alignment issues.

### **Limitations**

The limitations to this study must be considered when discussing the results. One limitation is the small sample size being used in this study. This study only looks at one college of education and syllabi from only six different areas. Broadening the scope of this study could alter the findings. Also, the examined edTPA data only come from 60 preservice teachers. Increasing this number would help the study’s validity as well.

As mentioned many times, it is crucial to consider that syllabi are not able to capture everything that happens within a course. To fully capture all that occurs, the researcher would need to engage in hours of observations, professor and student interviews, etc. Often, the syllabi only give a broad overview of what occurs.

Along these same lines, it is important to remember that this study does not touch on practical experience very heavily. Much can (and should) happen in these clinical experiences that shape a preservice teacher’s learning. Here, they see assessment in action and are often asked to create assessments themselves. This study does discuss these situations, but they are not included in the coding or frequency tables.

Also, this study discusses syllabi objectives extensively, but it is possible that these objectives do not accurately represent what happens in the course. This is an issue in itself that

needs to be addressed, but it needs to be considered when discussing the results of this study. As an example, it is possible that the DOK and ROK levels are not accurately simply because the professor did not put as much thought into the objective as the researcher might like. It might not be an accurate representation of what the course's students are being ask to learn.

### **Recommendations for Further Research**

This study only addresses a small corner of this massive topic. Using similar methodologies or undertaking new methods altogether, there are multiple next steps researchers can take to better understand the way to improve the assessment literacy within the teaching force.

One avenue would be taking this study's alignment model even further and include multiple new sources of data. The Danielson Model would provide an interesting dimension to this study as it is the most popular method used in teacher evaluation today. Comparing university assessment instruction to the way teachers are evaluated in their jobs would provide for a closer look at the priorities of each. Additionally, it would be interesting to examine employer perceptions in this processes, possibly through the use of interviews or surveys. In what ways does university assessment instruction align with the views of the people who will be hiring the graduates? Likewise, there would be value in interviewing or surveying the teacher candidates themselves to examine how their views align with the other data sources.

There would also be value in expanding this alignment study to include multiple universities. This study only examined one university's efforts, but replicating this study with another university or expanding the study to include multiple universities would provide a great deal of information to add to the conversation.

The researcher recommends increasing the sample size of the data, possibly through the inclusion of more syllabi. An increase in the sample size would improve the study's reliability. The alignment matrix (Table 7) would be one specific aspect that would be greatly improved with a larger sample size.

Another logical next step would be to examine the edTPA data for individual subjects and to run a statistical procedure to view how different subjects respond to their assessment instruction. The current study was unable to do this due to the limitations of gathering the correct data, but a future study might be able to shed more light on how specific strategies and methods affect student performance on their end-of-program assessment.

Keeping the idea of an alignment study but using a different method, the researcher recommends a study that incorporates the Surveys of Enacted Curriculum (SEC). The SEC is a method that utilizes a tested survey tool to gather data from professors on how much time and emphasis they place on certain aspects of instruction within a specific course (Blank, 2002; Porter, 2002). This method would add more to the discussion because it can more closely examine what is actually happening in the course. As previously discussed, there are limitations to using the syllabus to provide these data. The SEC method would assist in remedying those problems.

One question that this study could not answer is whether better alignment would produce better results. For example, in this study, it was observed that there was not alignment in the ROK levels between the syllabi and the edTPA. If there was better alignment, would scores improve? This study might require a control group and an experimental group. The researcher could also examine any correlations between the frequency numbers in the qualitative portion of this study with the scores in the quantitative portion of this study.

Another interesting idea would be to examine the way assessment strategies are being modeled by the professor within a specific course. When the professor assesses his or her own students, how is he or she doing it? What methods are being used? Is this having an impact on how their students learn to assess?

Impara et al. (1993) and Marso and Pigge (1993) both indicated that an assessment-specific course does not have a great impact on preservice teachers. Being that these are older studies, it is recommended that this issue be examined again. This could be done by examining the differences between programs that include an assessment-specific course and programs that, like the one studied here, incorporate assessment throughout.

As noted previously, the study at hand only addressed parts of the “where it should be learned” and “how it should be learned” realms of the FAST framework. These are areas where future research is needed. This process could begin by examining educator beliefs in regards to these realms. A study could also examine specific colleges of education or district professional development initiatives to see where and how these issues are being tackled.

### **Final Thoughts**

Assessment literacy is crucial to the education field. Collecting, analyzing, and using assessment data are part of the foundation of the education process and can greatly improve student learning if used correctly. Unfortunately, many argue that teachers do not possess the necessary skills to do this. Today, this is more important than ever considering that student performance data are often required in teacher evaluations, yet teachers often have a hard time understanding the very factors that are used to assess their job performance.

This study found that secondary education preservice teachers at a large Midwestern university statistically do not perform any better or worse on the edTPA’s assessment task than



on either of the other two tasks. There are specific assessment areas in which they perform better and specific areas in which they perform worse. Different content areas teach assessment at different levels and in different quantities. Areas of alignment and misalignment with the standards and edTPA were found and discussed.

There is work that can and should be done to improve assessment literacy overall. The conversation of whether teachers can transfer their theoretical assessment knowledge into practical assessment knowledge is still ongoing and must be studied more. Ultimately, an improved assessment literacy is needed to place teachers in a position to better impact student learning.

## REFERENCES

- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The Professional knowledge base of science teaching* (Vol. 12, pp. 205-221). Dordrecht: Springer.
- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91-106.
- Alkharusi, H. (2011). Psychometric properties of the teacher assessment literacy questionnaire for preservice teachers in Oman. *Procedia - Social and Behavioral Sciences*, 29, 1614-1624.
- Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal Of Teacher Education*, 39(2), 113-123.
- Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20-34.
- Armstrong, C. J. (2011). *Belizean primary school teachers' understanding of assessment; assessment practices, and use of student assessment data* (Unpublished doctoral dissertation). Oklahoma State University, Stillwater, OK.
- Atkin, J. M., Black, P., and Coffey, J. (2001). *Classroom assessment and the National Science Education Standards*. Washington, DC: National Academies Press.
- Bailey, J., Little, C., Rigney, R., Thaler, A., Weideman, K., & Yorkovich, B. (2010) *Assessment 101: Assessment made easy for first-year teachers*. ERIC.
- Bangert, A., & Kelting-Gibson, L. (2006). Teaching principles of Assessment Literacy through teacher work sample methodology. *Teacher Education and Practice*, 19(3), 351-364.

- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teacher beliefs* (pp. 284-300). New York: Routledge.
- Barnes, N., Fives, H., & Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education, 65*, 107-116.
- Bennett, K., & Cunningham, A. C. (2009). Teaching formative assessment strategies to pre-service teachers: Exploring the use of handheld computing to facilitate the action research process. *Journal of Computing in Teacher Education, 25*(3), 99-105.
- Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21-29.
- 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.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27-40.
- Brookhart, S. M. (2001). *The "standards" and classroom assessment research*. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, Dallas, TX.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3-12.

- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles and Practice*, 11(3). 305-322.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99, 166-170.
- Brown, G. T. L. (2007, Dec. 4-7). *Teachers' conceptions of assessment: Comparing measurement models for primary and secondary teachers in New Zealand*. Paper presented at the New Zealand Association for Research in Education (NZARE) annual conference, Christchurch, NZ.
- Brown, G. T. L. (2008a). Assessment literacy training and teachers' conceptions of assessment. In C. Rubie-Davies & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 285-302). New York: Nova Science.
- Brown, G. T. L. (2008b). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science.
- Brown, G. T. L. (2009). Teachers' self-reported assessment practices and conceptions: Using structural equation modelling to examine measurement and structural models. In T. Teao & M. S. Khine (eds.) *Structural Equation Modelling in Educational Research: Concepts and Applications* (pp. 243-266). Rotterdam, NL: SensePublishers.
- Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi experimental study of Indian teachers in private schools. *International Journal of Educational Research*, 71, 50-64.

- Brown, G. T. L., and Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of Multidisciplinary Evaluation*, 6(12), 68-91.
- Brown, G. L., Kennedy, K. J., Fok, P. K., Chan, J. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, 16(3), 347-363.
- Brown, G. T. L. & Lake, R. (2006, Nov. 26-30). *Queensland Teachers' conceptions of teaching, learning, curriculum and assessment: Comparisons with New Zealand teachers*. Paper presented at the Annual Conference of the Australian Association for Research in Education (AARE), Adelaide, Australia.
- Brown, G. T. L. & Lingbiao, G. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education*, 2(1).
- Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education*, 26, 319–337.
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, 15, 75–89.
- Buck, G. A., Trauth-Nare, A., & Kaftan, J. (2010). Making formative assessment discernable to pre-service teachers of science. *Journal of Research in Science Teaching*, 47(4), 402-421.
- Buyukkarci, K (2014). Assessment beliefs and practices of language teachers in primary education. *International Journal of Instruction*, 7(1), 107-120.

- 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: SAGE.
- Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice*, 26(1), 9-18.
- Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom Assessment for Student Learning: Doing it Right, Using it Well*, 2nd ed. Upper Saddle River, NJ: Pearson.
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Cornish, L., & Jenkins, K. A. (2012). Encouraging teacher development through embedding reflective practice in assessment. *Asia-Pacific Journal of Teacher Education*, 40(2), 159-170.
- Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W., and Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, 39(3), 124-130.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- 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 [http://www.ccsso.org/Resources/Programs/Interstate\\_Teacher\\_Assessment\\_Consortium\\_\(InTASC\).html](http://www.ccsso.org/Resources/Programs/Interstate_Teacher_Assessment_Consortium_(InTASC).html).
- Danielson, C. (2007). *Enhancing Professional Practice: A Framework for Teaching* (2<sup>nd</sup> ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson Group. (2017). *Charlotte Danielson*. Retrieved from <http://www.danielsongroup.org/charlotte-danielson/>.
- Darling-Hammond, L., & Bransford, J. (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco, CA: Jossey-Bass.
- Data Quality Campaign. (2014). *Teacher data literacy: It's about time*. Washington, DC: Data Quality Campaign. Retrieved from <https://dataqualitycampaign.org/wp-content/uploads/2016/03/DQC-Data-Literacy-Brief.pdf>.
- Dayal, H. C., & Lingam, G. I. (2015). Fijian teachers' conceptions of assessment. *Australian Journal of Teacher Education*, 40(8), 42-58.
- 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., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator*, 48(2), 128-142.

- 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.
- 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.
- DeLuca, C., & Lam, C. Y. (2014). Preparing teachers for assessment within diverse classrooms: An analysis of teacher candidates' conceptualizations. *Teacher Education Quarterly*, 41(3), 3-24.
- Dillon, P., Erkens, C., Sanna, D., & Savastano, L. F. (2015). Crowdlearning: 8 districts pool resources to focus on assessment literacy. *Journal of Staff Development*, 36(3), 28-32.
- DinanThompson, M., & Penney, D. (2015). Assessment literacy in primary physical education. *European Physical Education Review*, 21(4), 485-503.
- Dorfman, A. B., Galluzzo, G. R., & Meisels, S. J. (2006). Learning to teach: Developing assessment skills when program and placement are aligned. *Journal of Early Childhood Teacher Education*, 27(3), 231-247.
- edTPA (November 2017). *Educative Assessment & Meaningful Support: 2016 edTPA Administrative Report*. Retrieved from [https://secure.aacte.org/apps/rl/res\\_get.php?fid=3621&ref=rl](https://secure.aacte.org/apps/rl/res_get.php?fid=3621&ref=rl).



- Elliott, E. J. (2010, April 29). *Assessment As a Critical Element in Clinical Experiences for Teacher Preparation*. NCATE. Retrieved from <http://www.ncate.org/LinkClick.aspx?fileticket=oo50CSYDEFM%3D&tabid=715>.
- Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In C. Wyatt-Smith, V., Klenowski, & P. Colbert (Eds.), *The enabling power of assessment: Designing assessment for quality learning* (pp. 140-162). New York: Springer.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4.
- Falter Thomas, A., & Sondergeld, T. (2015). Investigating the Impact of Feedback Instruction: Partnering Preservice Teachers with Middle School Students to Provide Digital, Scaffolded Feedback. *Journal Of The Scholarship Of Teaching And Learning*, 15(4), 83-109.
- Fan, Y. C., Wang, T. H., & Wang, K. H. (2011). A Web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education*, 57(2), 1727-1740.
- Fraenkel, J. R. & Wallen, N. E. (2006). *How to Design and Evaluate Research in Education* (6<sup>th</sup> ed.). New York: McGraw-Hill.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Frey, B. B., & Schmitt, V. L. (2010). Teachers' classroom assessment practices. *Middle Grade Research Journal*, 5(3), 107-117.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.

- Gareis, C. R., & Grant, L. W. (2015). Assessment Literacy for Teacher Candidates: A Focused Approach. *Teacher Educators' Journal*, 2015, 4-21.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *PROFILE: Issues in Teachers' Professional Development*, 20(1), 179-195.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2), 14-18.
- 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.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607-621.
- 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*. Washington, DC: National Council on Teacher Quality. Retrieved from [http://www.nctq.org/dmsView/What\\_Teacher\\_Prep\\_Programs\\_Teach\\_K-12\\_Assessment\\_NCTQ\\_Report](http://www.nctq.org/dmsView/What_Teacher_Prep_Programs_Teach_K-12_Assessment_NCTQ_Report).
- Gunn, A. C., & Gilmore, A. (2014). Early childhood initial teacher education students' learning about assessment. *Assessment Matters*, 7, 24-38.
- Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. *Educational Measurement: Issues and Practice*, 4(1), 15-18.

- Hailaya, W., Alagumalai, S., & Ben, F. (2014). Examining the utility of Assessment Literacy Inventory and its portability to education systems in the Asia Pacific region. *Australian Journal of Education, 58*(3), 297-317.
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. New York: Routledge, 2009.
- Herrera, L. & Macias, D. (2015). A call for language assessment literacy in the education and development of teachers of English as a foreign language. *Colombian Applied Linguistics Journal, 17*(2), 302-312.
- Herrington, A., Herrington, J., & Glazer, E. (2002, Oct. 26-29). *Authentic approaches to learning assessment strategies: Beginning teachers' practice in classrooms*. Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Athens, GA.
- Highland, F. B. (2015). *Evaluation of a systematic approach to develop teachers' assessment literacy* (Unpublished doctoral dissertation). College of William and Mary, Williamsburg, VA.
- Hill, M., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing Assessment-capable Teachers: What Should Preservice Teachers Know and Be Able to Do?. *Assessment Matters, 2*, 44-64.
- Hill, M. F., Ell, F., Grudnoff, L., & Limbrick, L. (2014). Practise what you preach: Initial teacher education students learning about assessment. *Assessment Matters, 7*, 90-112.
- 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. *Assessment Matters, 7*, 4-23.

- Howley, M. D., Howley, A., Henning, J. E., Gillam, M. B., & Weade, G. (2013). Intersecting domains of assessment knowledge: School typologies based on interview with secondary teachers. *Educational Assessment, 18*(1), 26-48.
- Huai, N., Braden, J. P., White, J. L., & Elliott, S. N. (2006). Effect of an internet-based professional development program on teachers' Assessment Literacy for all students. *Teacher Education and Special Education, 29*(4), 36-52.
- Illinois State Board of Education (2013). *Illinois Professional Teaching Standards*. Retrieved from [http://www.isbe.net/peac/pdf/il\\_prof\\_teaching\\_stds.pdf](http://www.isbe.net/peac/pdf/il_prof_teaching_stds.pdf).
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Teachers' assessment background and attitude toward testing. *Theory into Practice, 32*, 113-117.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402.
- Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923-2931). Oxford: Blackwell.
- Jie-Qi, C., & McNamee, G. (2006). Strengthening early childhood teacher preparation: Integrating assessment, curriculum development, and instructional practice in student teaching. *Journal of Early Childhood Teacher Education, 27*(2), 109-128.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing, 27*(4), 555-584.
- Jones, J. (2014). Student teachers developing a critical understanding of formative assessment in the modern foreign languages classroom on an initial teacher education course. *Language Learning Journal, 42*(3), 275-288.

- Kelting-Gibson, L., Karsted, K., & Weikert, A. (2013). Assessment-ready preservice teachers. *Teacher Education and Practice*, 26(3), 529-541.
- Kiomrs, R., Abdolmehdi, R., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156-161.
- 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]. Retrieved from <http://www.amazon.com>.
- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22(3), 255-276.
- Krathwohl, D. R., & Payne, D. A. (1971). Defining and assessing educational objectives. In R. L. Thorndike, *Educational Measurement*. Washington, DC: American Council on Education. 17-41.
- Lai, M. K. & McNaughton, S. (2016). The impact of data use professional development on student achievement. *Teaching and Teacher Education*, 60, 434-443.
- Lee, J., & Son, J. (2015). Two teacher educators' approaches to developing preservice elementary teachers' mathematics assessment literacy: Intentions, outcomes, and new learning. *Teaching and Learning Inquiry*, 3(1), 47-62.
- Lee, J., & Yoon, J. Y. (2008). Teaching early childhood teacher candidates how to assess children's inquiry skills in science learning. *Contemporary Issues in Early Childhood*, 9(3), 265-269.

- Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice*, 17(1), 7-21.
- Leirhaug, P. E., MacPhail, A., & Annerstedt, C. (2016). "The grade alone provides no learning": Investigating assessment literacy among Norwegian physical education teachers. *Asia-Pacific Journal of Health, Sport and Physical Education*, 7(1), 21-36.
- Levy-Vered, A., & Alhija, F., N. A. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation*, 21(5-6), 378-406.
- Lian, L. H., Yew, W. T., & Meng, C. C. (2014). Enhancing Malaysian teachers' assessment literacy. *International Education Studies*, 7(10), 74-81.
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.
- Livingston, K. & Hutchinson, C. (2016). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice*, 24(2), 290-307.
- Lomax, R. G. (1996). On becoming assessment literate: An initial look at pre-service teachers' beliefs and practices. *The Teacher Educator*, 31(4), 292-303.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Loreneta-Catalán, E., & Kirk, D. (2016). Student teachers' understanding and application of assessment for learning during a physical education teacher education course. *European Physical Education Review*, 22(1), 65-81.

- Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23(2), 26-32.
- Lyon, E. G. (2013). Learning to assess science in linguistically diverse classrooms: Tracking growth in secondary science preservice teachers' assessment expertise. *Science Education*, 97(3), 442-467.
- MacLellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education*, 20(5), 523-535.
- Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71-85.
- Mandinach, E. B., & Gummer, E. S. (2011, May 12). *The complexities of integrating data-driven decision making into professional preparation in schools of education: It's harder than you think*. Retrieved from <http://educationnorthwest.org/sites/default/files/resources/Gummer-Mandinach%20full%20report.pdf>.
- 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. G. (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.

- Mandinach, E. B., Gummer, E. G., & Friedman, J. M. (2013). *An analysis of the survey of schools of education on use of data in their teacher preparation programs: An interim report*. San Francisco, CA: WestED. Retrieved from <https://datafordecisions.wested.org/wp-content/uploads/2014/08/Dell-Survey-Interim-Report.pdf>.
- Marso, R. N., & Pigge, F. L. (1993). Teachers' testing knowledge, skills, and practices. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 129-185).
- Martone, A. & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332-1361.
- Marzano, R. J. (2010). *Formative assessment & standards-based grading*. Bloomington, IN: Solution Tree Press.
- Matthews, J., Trimble, S., & Gay, A. (2007). But what do you do with the data? *Principal Leadership*, 7(9), 31-33
- Mayor, S. (2005). Pre-service teachers' developing perspectives on assessment and remediation of struggling readers. *Reading Improvement*, 42(3), 164.
- McMillan, J. H. (2001b). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.
- McMillan, J. H. (2008). *Assessment essentials for standards-based education* (2nd Ed.) Thousand Oaks, CA: Corwin Press.



- McMillan, J. H. Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research, 95*(4), 203.
- Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports*. U.S. Department of Education. Retrieved from <https://www2.ed.gov/rschstat/eval/data-to-inform-instruction/report.doc>
- Mertler, C. A. (2004) Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education, 33*(1), 49-64.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101-113.
- Mertler, C. A. & Campbell, C. (2005, April 11-15). *Measuring teachers' knowledge & 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.
- Mitton-Kukner, J. J., Munroe, E., & Graham, D. (2015). The Challenge of Differing Perspectives Surrounding Grades in the Assessment Education of Pre-Service Teachers. *Canadian Journal Of Higher Education, 45*(4), 322-342.
- Mooi, L. M., Periasamy, S., Ming, C. C., & Osman, S. (2014). Exploring the outcomes and issues of the implementation of an assessment course in a teacher education programme. *Procedia – Social and Behavioral Sciences, 114*, 883-888.
- Morrison, J. A. (2005). Using science notebooks to promote preservice teachers' understanding of formative assessment. *Issues in Teacher Education, 14*(1), 5-21.
- Munroe, E., Foran, A., MacLeod, K., Graham, D., Lunney-Borden, L., & Curry, A. (2012). Walking our talk about assessment with preservice teachers. *In Education, 18*(2), 53–66.

National Council for Accreditation of Teacher Education. (2008). *Unit accreditation standards*.

Retrieved from

<http://www.ncate.org/Standards/UnitStandards/UnitStandardsinEffect2008/tabid/476/Default.aspx>.

O'Connor, K. (2002). *How to Grade for Learning: Linking Grades to Standards*, 2nd Ed.

Thousand Oaks, CA: Corwin Press.

Odom, A. L., & Bell, C. V. (2017). Developing PK-12 preservice teachers' skills for understanding data-driven instruction through inquiry learning. *Journal of Statistics Education*, 25(1), 19.

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

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Phanachalaem, K., Sujiva, S., & Tangdhanakanond, K. (2016). The state of teachers' educational data use in Thailand. *Procedia - Social and Behavioral Sciences*, 217(1), 638-642.

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, 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
- Pierce, R., Chick, H., & Gordon, I. (2013). Teachers' perceptions of the factors influencing their engagement with statistical reports on student achievement data. *Australian Journal of Education*, 57(3), 237-255.
- Plake, B. S., & Impara, J. C. (1992). *Teacher competencies questionnaire description*. Lincoln, NE: University of Nebraska.
- Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of Classroom Assessment*. 55-68. New York, NY: Academic Press.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or Fundamental? *Theory into Practice*, 48(4), 4-11.
- 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*. Consortium for Policy Research in Education. ERIC.
- Poth, C. (2013). What assessment knowledge and skills do initial teacher education programs address? A Western Canadian perspective. *Alberta Journal of Educational Research*, 58(4), 634-656.

- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*, 115-131.
- REL West & WestEd. (2009). *Multi-State Review of Professional Teaching Standards*. Retrieved from <https://www.ctc.ca.gov/docs/default-source/educator-prep/cstp/review-of-other-state-standards.pdf>.
- Rorsyth, R., Cullen, R., Ringan, N., & Stubbs, M. (2015). Supporting the development of assessment literacy of staff through institutional process change. *London Review of Education, 13*(3), 34-41.
- Russell, M. K., & Airasian, P. W. (2011). *Classroom assessment: Concepts and applications* (7<sup>th</sup> ed.). New York, NY: McGraw-Hill.
- Sato, M., Chung, R. R., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of national board certification. *American Educational Research Journal, 45*(3), 669-700.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327.
- Schafer, W. D., & Lizzitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education, 38*(3), 57-63.
- Schaffer, C. L. (2014). When assessment and accountability intersect, good things can happen. *Journal of Curriculum & Instruction, 8*(1), 31-47. doi:10.3776/joci.2014.v8n1p31-47
- Schneider, M. C., & Randel, B. (2010). Research on characteristics of effective professional development programs for enhancing educators' skills in formative assessment. xin H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 251-276). New York: Routledge.

- Shepard, L. A. (2003). Reconsidering large-scale assessment to heighten its relevance to learning. In J. M. Atkin, J. Coffey, and National Science Teachers Association (Eds.), *Everyday Assessment in the Science Classroom* (pp. 121-146). Arlington: NSTA Press.
- 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: Josey-Bass.
- Shulha, L. M. (1999). Understanding novice teachers' thinking about student assessment. *Alberta Journal of Educational Research*, 45(3), 288-303.
- Siegel, M. A. (2013). Developing preservice teachers' expertise in equitable assessment for English learners. *Journal of Science Teacher Education*, 25(3), 289-308.
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, 22(4), 371-391.
- Smith, L. F., & Galvin, R. (2014). Toward assessment readiness: An embedded approach in primary initial teacher education. *Assessment Matters*, 7, 39-63.
- Smith, L. F., Hill, M. F., Cowie, B., & Gilmore, A. (2014). Preparing teachers to use the enabling power of assessment. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 418-445). Dordrecht: Springer
- Sondergeld, T. A., Bell, C. A., & Leusner, D. (2010). Understanding how teachers engage in formative assessment. *Teaching & Learning*, 24(2), 72-86.
- Stanford Center for Assessment, Learning, & Equity. (2014). *edTPA (2014) Crosswalk: Interstate Teacher Support and Assessment Consortium (InTASC) Model Core Teaching Standards (2013)*. Retrieved from <https://secure.aacte.org/apps/rl/resource.php?resid=298&ref=edtpa>.

- Stanford Center for Assessment, Learning, & Equity. (Sept. 2016). Secondary English Language Arts: Assessment Handbook. *edTPA*.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan*, 83, 758-765.
- Stiggins, R. J., & Knight, T. (1997). *But are they really learning?* Portland, OR: Assessment Training Institute.
- Stobaugh, R., Tassell, J., & Norman, A. D. (2010). Improving pre-service teacher preparation through the teacher work sample: Exploring assessment and analysis of student learning. *Action in Teacher Education*, 32(1), 39-53.
- 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/>.
- Thomas, S. & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practices: "That's my grade. That's me." *Educational Psychologist*, 32(2), 107-123.
- Tomlinson, C. A. (2005). Grading and differentiation: Paradox or good practice. *Theory into Practice*, 44(3), 262-269.
- Volante, L, & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30, 749-770.

- 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.
- Wang, J. R., Kao, H. L., & Lin, S. W. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching and Teacher Education, 26*, 522-529.
- Wayman, J. C., Wilkerson, S. B., Cho, V., Mandinach, E. B., & Supovitz, J. A. (2016). *Guide to using the Teacher Data Use Survey*. Washington, DC: US Department of Education, Insitute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia. Retrieved from <https://files.eric.ed.gov/fulltext/ED569169.pdf>.
- 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.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *Australian Educational Researcher, 40*(2), 241-256.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice, 17*(1), 59-75.

- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching And Teacher Education, 58*, 149-162.
- Yan, X., Zhang, C., & Fan, J. J. (2018). “Assessment knowledge is important, but ...”: How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System, 74*, 158-168.
- Yao, Y. (2015). Teacher perceptions of classroom assessment: A focus group interview. *SRATE Journal, 24*(2), 51-58.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323-342.



APPENDIX A: ILLINOIS PROFESSIONAL TEACHING STANDARDS ASSESSMENT

STANDARD AND INDICATORS

**Standard 7 – Assessment** – The competent teacher understands and uses appropriate formative and summative assessments for determining student needs, monitoring student progress, measuring student growth, and evaluating student outcomes. The teacher makes decisions driven by data about curricular and instructional effectiveness and adjusts practices to meet the needs of each student.

Indicator	The Competent Teacher...
7A	understands the purposes, characteristics, and limitations of different types of assessments, including standardized assessments, universal screening, curriculum-based assessment, and progress monitoring tools;
7B	understands that assessment is a means of evaluating how students learn and what they know and are able to do in order to meet the Illinois Learning Standards;
7C	understands measurement theory and assessment-related issues, such as validity, reliability, bias, and appropriate and accurate scoring;
7D	understands current terminology and procedures necessary for the appropriate analysis and interpretation of assessment data;
7E	understands how to select, construct, and use assessment strategies and instruments for diagnosis and evaluation of learning and instruction;
7F	knows research-based assessment strategies appropriate for each student;
7G	understands how to make data-driven decisions using assessment results to adjust practices to meet the needs of each student;
7H	knows legal provisions, rules, and guidelines regarding assessment and assessment accommodations for all student populations; and
7I	knows assessment and progress monitoring techniques to assess the effectiveness of instruction for each student.
7J	uses assessment results to determine student performance levels, identify learning targets, select appropriate research-based instructional strategies, and implement instruction to enhance learning outcomes;

*(Table Continues)*

- 7K appropriately uses a variety of formal and informal assessments to evaluate the understanding, progress, and performance of an individual student and the class as a whole;
- 7L involves students in self-assessment activities to help them become aware of their strengths and needs and encourages them to establish goals for learning;
- 7M maintains useful and accurate records of student work and performance;
- 7N accurately interprets and clearly communicates aggregate student performance data to students, parents or guardians, colleagues, and the community in a manner that complies with the requirements of the Illinois School Student Records Act [105 ILCS 10], 23 Ill. Adm. Code 375 (Student Records), the Family Educational Rights and Privacy Act (FERPA) (20 USC 1232g) and its implementing regulations (34 CFR 99; December 9, 2008);
- 7O effectively uses appropriate technologies to conduct assessments, monitor performance, and assess student progress;
- 7P collaborates with families and other professionals involved in the assessment of each student;
- 7Q uses various types of assessment procedures appropriately, including making accommodations for individual students in specific contexts; and
- 7R uses assessment strategies and devices that are nondiscriminatory, and take into consideration the impact of disabilities, methods of communication, cultural background, and primary language on measuring knowledge and performance of students.

---

(Illinois State Board of Education, 2013).

## APPENDIX B: INTASC ASSESSMENT STANDARD AND INDICATORS

**Standard 6 – 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.

---

### Performances

---

6(a) The teacher balances the use of formative and summative assessment as appropriate to support, verify, and document learning.

6(b) The teacher designs assessments that match learning objectives with assessment methods and minimizes sources of bias that can distort assessment results.

6(c) The teacher works independently and collaboratively to examine test and other performance data to understand each learner’s progress and to guide planning.

6(d) The teacher engages learners in understanding and identifying quality work and provides them with effective descriptive feedback to guide their progress toward that work.

6(e) The teacher engages learners in multiple ways of demonstrating knowledge and skill as part of the assessment process.

6(f) The teacher models and structures processes that guide learners in examining their own thinking and learning as well as the performance of others.

6(g) The teacher effectively uses multiple and appropriate types of assessment data to identify each student’s learning needs and to develop differentiated learning experiences.

6(h) The teacher prepares all learners for the demands of particular assessment formats and makes appropriate accommodations in assessments or testing conditions, especially for learners with disabilities and language learning needs.

6(i) The teacher continually seeks appropriate ways to employ technology to support assessment practice both to engage learners more fully and to assess and address learner needs.

---

### Essential Knowledge

---

6(j) The teacher understands the differences between formative and summative applications of assessment and knows how and when to use each.

6(k) The teacher understands the range of types and multiple purposes of assessment and how to design, adapt, or select appropriate assessments to address specific learning goals and individual differences, and to minimize sources of bias.

*(Table Continues)*

6(l) The teacher knows how to analyze assessment data to understand patterns and gaps in learning, to guide planning and instruction, and to provide meaningful feedback to all learners.

6(m) The teacher knows when and how to engage learners in analyzing their own assessment results and in helping to set goals for their own learning.

6(n) The teacher understands the positive impact of effective descriptive feedback for learners and knows a variety of strategies for communicating this feedback.

6(o) The teacher knows when and how to evaluate and report learner progress against standards.

6(p) The teacher understands how to prepare learners for assessments and how to make accommodations in assessments and testing conditions, especially for learners with disabilities and language learning needs.

---

#### Critical Dispositions

---

6(q) The teacher is committed to engaging learners actively in assessment processes and to developing each learner's capacity to review and communicate about their own progress and learning.

6(r) The teacher takes responsibility for aligning instruction and assessment with learning goals.

6(s) The teacher is committed to providing timely and effective descriptive feedback to learners on their progress.

6(t) The teacher is committed to using multiple types of assessment processes to support, verify, and document learning.

6(u) The teacher is committed to making accommodations in assessments and testing conditions, especially for learners with disabilities and language learning needs.

6(v) The teacher is committed to the ethical use of various assessments and assessment data to identify learner strengths and needs to promote learner growth.

---

(Council of Chief State School Officers, 2011).

## APPENDIX C: THE FRAMEWORK FOR TEACHING (DANIELSON FRAMEWORK)

---

### Domain 1: Planning and Preparation

---

#### 1a Demonstrating Knowledge of Content and Pedagogy

- Content knowledge  
Prerequisite relationships
- Content pedagogy

#### 1b Demonstrating Knowledge of Students

- Child development
- Learning process
- Special needs
- Student skills, knowledge, and proficiency
- Interests and cultural heritage

#### 1c Setting Instructional Outcomes

- Value, sequence, and alignment
- Clarity
- Balance
- Suitability for diverse learners

#### 1d Demonstrating Knowledge of Resources

- For classroom
- To extend content knowledge
- For students

#### 1e Designing Coherent Instruction

- Learning activities
- Instructional materials and resources
- Instructional groups
- Lesson and unit structure

#### 1f Designing Student Assessments

- Congruence with outcomes
- Criteria and standards
- Formative assessments
- Use for planning

---

### Domain 2: Classroom Environment

---

#### 2a Creating an Environment of Respect and Rapport

- Teacher interaction with students
- Student interaction with students

#### 2b Establishing a Culture for Learning

- Importance of content
- Expectations for learning and achievement
- Student pride in work

*(Table Continues)*

---

2c Managing Classroom Procedures

- Instructional groups
- Transitions
- Materials and supplies
- Non-instructional duties
- Supervision of volunteers and paraprofessionals

2d Managing Student Behavior

- Expectations
- Monitoring behavior
- Response to misbehavior

2e Organizing Physical Space

- Safety and accessibility
- Arrangement of furniture and resources

---

Domain 3: Instruction

---

3a Communicating with Students

- Expectations for learning
- Directions and procedures
- Explanations of content
- Use of oral and written language

3b Using Questioning and Discussion Techniques

- Quality of questions
- Discussion techniques
- Student participation

3c Engaging Students in Learning

- Activities and assignments
- Student groups
- Instructional materials and resources
- Structure and pacing

3d Using Assessment in Instruction

- Assessment criteria
- Monitoring of student learning
- Feedback to students
- Student self-assessment and monitoring

3e Demonstrating Flexibility and Responsiveness

- Lesson adjustment
- Response to students
- Persistence

*(Table Continues)*

---

## Domain 4: Professional Responsibilities

---

### 4a Reflecting on Teaching

- Accuracy
- Use in future teaching

### 4b Maintaining Accurate Records

- Student completion of assignments
- Student progress in learning
- Non-instructional records

### 4c Communicating with Families

- About instructional program
- About individual students
- Engagement of families in instructional program

### 4d Participating in the Professional Community

- Relationships with colleagues
- Participation in school projects
- Involvement in culture of professional inquiry
- Service to school

### 4e Growing and Developing Professionally

- Enhancement of content knowledge and pedagogical skill
- Receptivity to feedback from colleagues
- Service to the profession

### 4f Showing Professionalism

- Integrity/ethical conduct
- Service to students
- Advocacy
- Decision-making
- Compliance with school/district regulations

(Danielson, 2007)

APPENDIX D: EDTPA SCIENCE RUBRIC 5: PLANNING ASSESSMENTS TO MONITOR  
AND SUPPORT STUDENT LEARNING

How are the informal and formal assessments selected or designed to monitor students' progress toward using science concepts and scientific practices during inquiry to explain or predict a real-world phenomenon?

Level 1	Level 2	Level 3	Level 4	Level 5
<p>The assessments only provide evidence of students' ability to memorize and follow prescribed procedures.</p> <p>OR</p> <p>Candidate does not attend to ANY ASSESSMENT requirements in IEPs and 504 plans.</p>	<p>The assessments provide limited evidence to monitor students' understanding of</p> <ul style="list-style-type: none"> <li>• science concepts,</li> <li>• the phenomenon,</li> <li>AND</li> <li>• the application of scientific practices during scientific inquiry during the learning segment.</li> </ul>	<p>The assessments provide evidence to monitor students' understanding of</p> <ul style="list-style-type: none"> <li>• science concepts,</li> <li>• the phenomenon,</li> <li>AND</li> <li>• the application of scientific practices during scientific inquiry during the learning segment.</li> </ul>	<p>The assessments provide multiple forms of evidence to monitor students' progress toward developing understanding of</p> <ul style="list-style-type: none"> <li>• science concepts,</li> <li>• the phenomenon,</li> <li>AND</li> <li>• the application of scientific practices during scientific inquiry throughout the learning segment.</li> </ul>	<p>Level 4 plus:</p> <p>The assessments are strategically designed to allow individuals or groups with specific needs to demonstrate their learning.</p>



APPENDIX E: EDTPA SCIENCE RUBRIC 10: ANALYZING TEACHING EFFECTIVENESS

How does the candidate use evidence to evaluate and change teaching practice to meet students' varied learning needs?

<b>Level 1</b>	<b>Level 2</b>	<b>Level 3</b>	<b>Level 4</b>	<b>Level 5</b>
Candidate suggests changes unrelated to evidence of student learning.	Candidate proposes changes to teacher practice that are superficially related to student learning needs (e.g., task management, pacing, improving directions).	Candidate proposes changes that address students' collective learning needs related to central focus.  Candidate makes superficial connections to research and/or theory.	Candidate proposes changes that address individual and collective learning needs related to the central focus.  Candidate makes connections to research and/or theory.	Level 4 plus: Candidate justifies changes using principles from research and/or theory.

**APPENDIX F: EDTPA SCIENCE RUBRIC 11: ANALYSIS OF LEARNING**

How does the candidate analyze evidence of student learning related to conceptual understanding, the use of scientific practices during inquiry, and evidence-based explanations or reasonable predictions about a real-world phenomenon?

<b>Level 1</b>	<b>Level 2</b>	<b>Level 3</b>	<b>Level 4</b>	<b>Level 5</b>
<p>The analysis is superficial or not supported by either student work samples or the summary of student learning.</p> <p>OR</p> <p>The evaluation criteria, learning objectives, and/or analysis are not aligned with each other.</p>	<p>The analysis focuses on what students did right OR wrong.</p>	<p>The analysis focuses on what student did right AND wrong.</p> <p>AND</p> <p>Analysis includes some differences in whole class learning.</p>	<p>Analysis uses specific examples from work samples to demonstrate patterns of learning consistent with the summary.</p> <p>AND</p> <p>Patterns of learning are described for whole class.</p>	<p>Analysis uses specific evidence from work samples to demonstrate the connections between quantitative and qualitative patterns of learning for individuals or groups.</p>

APPENDIX G: EDTPA SCIENCE RUBRIC 12: PROVIDING FEEDBACK TO GUIDE

LEARNING

What type of feedback does the candidate provide to focus students?

<b>Level 1</b>	<b>Level 2</b>	<b>Level 3</b>	<b>Level 4</b>	<b>Level 5</b>
<p>Feedback is unrelated to the learning objectives OR is developmentally inappropriate.</p> <p>OR</p> <p>Feedback contains significant content inaccuracies.</p> <p>OR</p> <p>No feedback is provided to one or more focus students.</p>	<p>Feedback is general and addresses needs AND/OR strengths related to the learning objectives.</p>	<p>Feedback is specific and addresses either needs OR strengths related to the learning objectives.</p>	<p>Feedback is specific and addresses both strengths AND needs related to the learning objectives.</p>	<p>Level 4 plus: Feedback for one or more focus students</p> <ul style="list-style-type: none"> <li>• provides a strategy to address an individual learning need</li> </ul> <p>OR</p> <ul style="list-style-type: none"> <li>• makes connections to prior learning or experience to improve learning.</li> </ul>

APPENDIX H: EDTPA SCIENCE RUBRIC 13: STUDENT UNDERSTANDING AND USE  
OF FEEDBACK

How does the candidate support focus students to understand and use the feedback to guide their further learning?

Level 1	Level 2	Level 3	Level 4	Level 5
<p>Opportunities for understanding or using feedback are not described.</p> <p>OR</p> <p>Candidate provides limited or no feedback to inform student learning.</p>	<p>Candidate provides vague description of how focus students will understand or use feedback.</p>	<p>Candidate describes how focus students will understand or use feedback related to the learning objectives.</p>	<p>Candidate describes how s/he will support focus students to understand and use feedback on their strengths OR weaknesses related to the learning objectives.</p>	<p>Candidate describes how s/he will support focus students to understand and use feedback on their strengths AND weaknesses related to the learning objectives.</p>

APPENDIX I: EDTPA SCIENCE RUBRIC 14: ANALYZING STUDENTS' LANGUAGE USE  
AND SCIENCE LEARNING

How does the candidate analyze students' use of language to develop content understanding?

Level 1	Level 2	Level 3	Level 4	Level 5
<p>Candidate identifies student language use that is superficially related or unrelated to the language demands (function, vocabulary and/or symbols, and additional demands).</p> <p>OR</p> <p>Candidate's description or explanation of language use is not consistent with the evidence submitted.</p>	<p>Candidate describes how students use only one language demand (vocabulary, and/or symbols, function, syntax, discourse).</p>	<p>Candidate explains and provides evidence of students' use of</p> <ul style="list-style-type: none"> <li>• the language function</li> <li>AND</li> <li>• one or more additional language demands (vocabulary and/or symbols, syntax, discourse).</li> </ul>	<p>Candidate explains and provides evidence of students' use of</p> <ul style="list-style-type: none"> <li>• the language function,</li> <li>• vocabulary and/or symbols,</li> <li>AND</li> <li>• additional language demand(s) (syntax, discourse)</li> </ul> <p>in ways that develop content understandings.</p>	<p>Level 4 plus: Candidate explains and provides evidence of language use and content learning for students with varied needs.</p>

APPENDIX J: EDTPA SCIENCE RUBRIC 15: USING ASSESSMENT TO INFORM

INSTRUCTION

How does the candidate use the analysis of what students know and are able to do to plan next steps in instruction?

Level 1	Level 2	Level 3	Level 4	Level 5
<p>Next steps do not follow from the analysis.</p> <p>OR</p> <p>Next steps are not relevant to the learning objectives assessed.</p> <p>OR</p> <p>Next steps are not described in sufficient detail to understand them.</p>	<p>Next steps primarily focus on changes to teaching practice that are superficially related to student learning needs, for example, repeating instruction, pacing, or classroom management issues.</p>	<p>Next steps propose general support that improves student learning related to assessed learning objectives.</p> <p>Next steps are loosely connected with research and/or theory.</p>	<p>Next steps provided targeted support to individuals or groups to improve their learning relative to</p> <ul style="list-style-type: none"> <li>• conceptual understanding,</li> <li>• use of scientific practices during inquiry, AND/OR</li> <li>• construction of evidence-based explanations of or reasonable predictions about a real-world phenomenon.</li> </ul> <p>Next steps are connected with research and/or theory.</p>	<p>Next steps provide targeted support to individuals or groups to improve their learning relative to</p> <ul style="list-style-type: none"> <li>• conceptual understanding,</li> <li>• use of scientific practices during inquiry, AND/OR</li> <li>• construction of evidence-based explanations of or reasonable predictions about a real-world phenomenon.</li> </ul> <p>Next steps are justified with principles from research and/or theory.</p>

## APPENDIX K: MEMBER-CHECKING RECRUITMENT LETTER

Dear \_\_\_\_

You may remember from an earlier correspondence that I am conducting a dissertation research study on where and to what extent assessment is being taught to undergraduate secondary education preservice teachers within [name of institution]'s school of education. I want to thank you for granting me permission to use the syllabi from \_\_\_\_ for my study. I have since had my dissertation proposal approved by my committee and I am now in the process of completing my study.

I have conducted a document analysis on the syllabi. I have classified and coded them in three different dimensions: categorical concurrence, depth of knowledge, and range of knowledge. In an effort to increase the validity of my data, I am interested in sharing my findings with you or someone in your department and engaging in a member check. I believe that my study will be greatly benefitted by the input from those who work most closely with the course syllabi that are being analyzed. I want my study to accurately represent your department and I feel this member check is a necessary step in doing so.

If you agree to it, I would ask that you or someone you designate from your department review my analyses. Attached to this email is my full analyses of the syllabi as well as a narrative of what I have done. I would then be interested in setting up an appointment to meet with you to discuss your review of my analyses. I am somewhat flexible on time and I could meet with you on campus sometime within the months of March or April. I do live in [nearby city], but I am happy to meet you in person. If scheduling doesn't permit an in-person meeting, we could also discuss a virtual meeting.

There is, of course, no obligation to participate, nor will there be any consequences to declining. I've attached an informed consent form to assist your department in the decision on participating.

Thank you for your time and I look forward to hearing back from you.

Sincerely,

Daniel Frederking

## APPENDIX L: INFORMED CONSENT FORM

### **Introduction**

This research study is being conducted by Daniel Frederking at [name of institution] to explore where assessment is being taught to undergraduate secondary education preservice teachers. You must be over 18 years of age to participate.

### **Procedures**

If you choose to take part in this research study you will be asked to review the researcher's analysis of selected syllabi from your department to determine where and to what extent assessment is covered in the course. You will then be asked to meet with the researcher to discuss your review. The meeting will take approximately 15 minutes.

### **Risks/Discomforts**

The risks associated with this research are no greater than those encountered in everyday life.

### **Benefits**

There are no direct benefits to participants. However, your participation will help ensure the researcher is representing your department correctly within the study.

### **Confidentiality**

Names of participants will not be recorded. Participants' role in the department will be identified, but the name of the university will not be included.

### **Participation**

Participating in this study is voluntary. Refusal to participate involves no penalty. You may discontinue participation at any time without penalty. You can also skip any part of the process (review or meeting) if you'd prefer.

### **Conflicts of Interest**

This study is conducted by a doctoral student in Curriculum and Instruction at [name of institution]. The researcher is also a former undergraduate secondary education major at [name of institution] and has taught one semester of CI 218 to undergraduate secondary education majors.

### **Questions about the Research**

For questions about this research Daniel Frederking can be contacted at [phone] or [email].

---

You will be given a copy of this consent form for your records.

I consent to participating in the above study.

Signature \_\_\_\_\_

Date \_\_\_\_\_

**Please mark one of the below choices:**



\_\_\_\_ I grant permission for this conversation in which I am participating to be audio-recorded.

\_\_\_\_ I DO NOT grant permission for the conversation in which I am participating to be audio-recorded.

Signature \_\_\_\_\_

Date \_\_\_\_\_

If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Research Ethics & Compliance Office at [name of institution] at [phone] or via email at [email].

## APPENDIX M: WORKING CODEBOOK: CATEGORICAL CONCURRENCE

### **Label:** Alignment to learning goals

**Definition:** A learning goal is the objective that the teacher intends for the student to meet at the end of a given period of time. Assessments should align to these goals. The document mentions the learning goal in conjunction with assessment.

**Criteria/Indicators:** Standards-based grading, alignment studies

**Exclusions:** Mentions of standards, objectives, or learning goals alone is not enough. It must be discussed in connection with assessment.

### **Examples:**

InTASC 6(r): “The teacher takes responsibility for aligning instruction and assessment with learning goals.”

BIO 305 Syllabus: “Objectives: Analyze assessments for effectiveness in evaluating student learning of given objectives.”

### **Label:** Formative and Summative

**Definition:** Formative assessment is data collection *during* the learning process when there is still time to adjust the learning. Summative assessment is data collection *after* the learning process when it is time to evaluate how much they learned. The document refers to the two types and teaches students to distinguish between them.

**Criteria/Indicators:** Most often, this code will directly mention the words “formative” and/or “summative”.

**Exclusions:** Not to be confused with the “feedback and motivation” code or the “results to guide instruction” code. This category is specifically about the distinction between the two forms of assessment.

### **Examples:**

InTASC 6(a): “The teacher balances the use of formative and summative assessment as appropriate to support, verify, and document learning.”

CI 218 Syllabus: “Topical Outline: Effective Formative/Summative Assessment Methods”

**Label:** *Statistical literacy*

**Definition:** Statistical literacy is the knowledge and know-how needed to design assessments that are valid and reliable and to accurately understand the data that is collected through the assessment process. The document refers to grading practices, assigning grades, validity, reliability, or data interpretation.

**Criteria/Indicators:** Include validity and reliability issues. Include data interpretation and analysis.

**Examples:**

InTASC 6(l): “The teacher knows how to analyze assessment data to understand patterns and gaps in learning, to guide planning and instruction, and to provide meaningful feedback to all learners.”

CI 218 Syllabus: “Topical Outline: Grade Management”

PHY 323 Syllabus: “Performance Objectives: 1a. critique in writing a poorly written physics examination”

**Label:** *Feedback and motivation*

**Definition:** Feedback is an important factor in the assessment process because it gives students specific details on what they did correctly and incorrectly and what they can do to fix their problems. Many studies point to student motivation as a direct result of effective feedback.

**Criteria/Indicators:** The document refers directly to feedback and/or student motivation.

**Exclusions:** Simply mentioning students being “engaged” in an assessment is not enough. The engagement must come from the teacher’s efforts to involve the students in the conversations about their work.

**Examples:**

InTASC 6(d): “The teacher engages learners in understanding and identifying quality work and provides them with effective descriptive feedback to guide their progress toward that work.”

edTPA Task 3, Rubric 12: “Feedback is specific and addresses both strengths AND needs related to the learning objectives.”

**Label:** Student self-assessment

**Definition:** When students effectively assess their own learning, studies have shown a large increase in student achievement. Teachers must understand the value of student self-assessment and create opportunities to incorporate it into classroom lessons and activities.

**Criteria/Indicators:** The document refers to the theory and/or purposes of student self-assessment and its inclusion in the classroom.

**Examples:**

InTASC 6(f): “The teacher models and structures processes that guide learners in examining their own thinking and learning as well as the performance of others.”

**Label:** Tools and types

**Definition:** Teachers must have a full arsenal of assessment tools and types of assessments that they can use for the correct classroom purposes. These include tests, quizzes, presentations, projects, observations, interviews, portfolios, etc. The document mentions an understanding of specific assessment tools and types.

**Criteria/Indicators:** Mention of understanding or creating assessment instruments; mention of variety of assessment types. Also could include mentions of using multiple assessment types to triangulate student data or incorporating technology into the assessment process.

**Examples:**

InTASC 6(k): “The teacher understands the range of types and multiple purposes of assessment and how to design, adapt, or select appropriate assessments to address specific learning goals and individual differences, and to minimize sources of bias.”

CI 218 Syllabus: “Topical Outline: Traditional and Authentic Assessment Methods”

**Label:** Diverse learners

**Definition:** Quality assessors must be aware of the cultural and educational differences present within their classrooms. When assessing, these factors must be considered so no student is put at a disadvantage due to a factor that has nothing to do with what is being assessed. The document mentions an awareness of diversity in relation to assessment.

**Criteria/Indicators:** Specific mention of diversity, differentiation

**Examples:**

Rubric 5: “Candidate does not attend to ANY ASSESSMENT requirements in IEPs and 504 plans.”

InTASC 6(h): “The teacher prepares all learners for the demands of particular assessment formats and makes appropriate accommodations in assessments or testing conditions, especially for learners with disabilities and language learning needs.”

**Label:** Results to guide instruction

**Definition:** The assessment data must be incorporated into the planning of instruction. Teachers must be open to adjusting their plan based on the student learning (or lack thereof) that is evidenced by the assessments. The document mentions the need for instruction to be adjusted after data collection. This could apply to formative or summative assessment.

**Criteria/Indicators:** Specific mention of informing instruction

**Examples:**

Rubric 15: “How does the candidate use the analysis of what students know and are able to do to plan next steps in instruction?”

CI 218 Syllabus: “Topical Outline: Reflection of Assessment to Inform Teaching

**Label:** Language and literacy

**Definition:** Students must be given the opportunity to express themselves through language and quality assessors must consider literacy when designing assessments. Since literacy is one of the fundamental skills that students should possess, it cannot be ignored in assessment.

**Criteria/Indicators:** There is mention of language use or literacy skills in relation to assessment.

**Examples:**

edTPA: Task 3: Rubric 14: “Analyzing students’ language use and science learning: How does the candidate analyze students use of language to develop content understanding?”

COM 293: “Course Goals: to sharpen their ability to analyze and evaluate the communication of others in a diverse environment.”

## APPENDIX N: WORKING CODEBOOK: DEPTH OF KNOWLEDGE

### **Label:** Level 1: Low cognitive level

**Definition:** Ability to identify, define, recognize and recall assessment knowledge

**Criteria/Indicators:** Verbs “identify, define, recognize, recall, understand”

**Exclusions:** Statements that do not contain a verb unless meaning is explicitly implied.

#### **Examples:**

CI 218 Syllabus: “Course Overview: Teacher candidates understand and develop the knowledge, skills, and dispositions necessary to effectively assess student work samples related to school success.”

BIO 305 Syllabus: “Objective: Determine appropriate assessment strategies for different lessons.”

### **Label:** Level 2: Moderate cognitive level

**Definition:** Ability to apply and analyze assessment knowledge. Establish connections between assessment knowledge, teaching practice, and personal experiences

**Criteria/Indicators:** Using assessment knowledge to select appropriate instruments, gather data, provide feedback. Verbs “apply, analyze”

**Exclusions:** Statements that do not contain a verb unless meaning is explicitly implied.

#### **Examples:**

InTASC 6(t): “The teacher is committed to providing timely and effective descriptive feedback to learners on their progress.”

COM 291 Syllabus: “Course Objective: Validating teaching: Implementing reliable and relevant student/faculty assessment mechanisms to determine if competencies have been achieved.”

### **Label:** Level 3: High cognitive level

**Definition:** 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.

**Criteria/Indicators:** Creating effective instruments, making decisions based on data. Verbs “evaluate, synthesize, create, develop”

**Exclusions:** Statements that do not contain a verb unless meaning is explicitly implied.

**Examples:**

CI 218 Syllabus: “Course Objectives: Develop and implement effective assessment instruments and interpret student performance data to inform instruction.”

CHE 303 Syllabus: “Course Objectives: Create and use appropriate formative and summative assessment tools to improve and assess student learning, making informed decisions based on the outcome of the assessments.”

## APPENDIX O: WORKING CODEBOOK: RANGE OF KNOWLEDGE

### **Label:** Level 1: Instructional objective

**Definition:** Specific objectives used to plan assessment use in daily teacher practice (e.g., test design, questioning approaches, etc.).

**Criteria/Indicators:** Focus on narrow topics of learning. Could take days or hours.

**Exclusions:** Only objectives and goals will be considered.

#### **Examples:**

PHY 323 Syllabus: Performance Objectives: 1c. create a detailed, multidimensional scoring rubric for a complex task.”

CHE 160 Syllabus: “Course Objectives: Design and implement instruction on an assigned chemistry topic in consideration of NGSS, including creation of an assessment.”

### **Label:** Level 2: Educational objective

**Definition:** Statements that describe teacher accomplishment 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)

**Criteria/Indicators:** Narrow enough to help teachers plan teaching, but broad enough to suggest a range of student outcomes. Could take weeks or months.

**Exclusions:** Only objectives and goals will be considered.

#### **Examples:**

InTASC 6(o): “The teacher knows when and how to evaluate and report learner progress against standards.”

edTPA Task 1, Rubric 3: “Using knowledge of students to inform teaching and learning. How does the candidate use knowledge of his/her students to justify instructional plans?”

### **Label:** Level 3: Global objective

**Definition:** Very broad statements of intended learning that require years to accomplish (e.g., teachers will be assessment literate)



**Criteria/Indicators:** Very general and encompass a large number of smaller objectives. Could take years.

**Exclusions:** Only objectives and goals will be considered.

**Examples:**

InTASC 6(k): “The teacher understands the range of types and multiple purposes of assessment and how to design, adapt, or select appropriate assessments to address specific learning goals and individual differences, and to minimize sources of bias.”

CHE 303 Syllabus” “Course Objectives: Create and use appropriate formative and summative assessment tools to improve and assess student learning, making informed decisions based on the outcome of the assessments.”

APPENDIX P: BONFERRONI POST HOC TEST FOR RUBRICS

Rubric	Rubric	Mean Diff.	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
5	10	.292*	0.082	0.015	0.032	0.551
	11	-0.042	0.122	1.000	-0.429	0.346
	12	-.392*	0.104	0.008	-0.723	-0.060
	13	0.183	0.087	0.841	-0.094	0.461
	14	0.092	0.096	1.000	-0.214	0.397
	15	0.150	0.089	1.000	-0.133	0.433
10	5	-.292*	0.082	0.015	-0.551	-0.032
	11	-0.333	0.113	0.092	-0.691	0.024
	12	-.683*	0.110	0.000	-1.031	-0.335
	13	-0.108	0.088	1.000	-0.388	0.172
	14	-0.200	0.105	1.000	-0.534	0.134
	15	-0.142	0.087	1.000	-0.419	0.136
11	5	0.042	0.122	1.000	-0.346	0.429
	10	0.333	0.113	0.092	-0.024	0.691
	12	-0.350	0.134	0.241	-0.776	0.076
	13	0.225	0.110	0.945	-0.124	0.574
	14	0.133	0.112	1.000	-0.222	0.489
	15	0.192	0.116	1.000	-0.176	0.560
12	5	.392*	0.104	0.008	0.060	0.723
	10	.683*	0.110	0.000	0.335	1.031
	11	0.350	0.134	0.241	-0.076	0.776
	13	.575*	0.104	0.000	0.244	0.906
	14	.483*	0.117	0.002	0.112	0.855
	15	.542*	0.114	0.000	0.179	0.904
13	5	-0.183	0.087	0.841	-0.461	0.094
	10	0.108	0.088	1.000	-0.172	0.388
	11	-0.225	0.110	0.945	-0.574	0.124
	12	-.575*	0.104	0.000	-0.906	-0.244
	14	-0.092	0.091	1.000	-0.380	0.197
	15	-0.033	0.071	1.000	-0.259	0.193

(Table Continues)

Rubric	Rubric	Mean Diff.	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
14	5	-0.092	0.096	1.000	-0.397	0.214
	10	0.200	0.105	1.000	-0.134	0.534
	11	-0.133	0.112	1.000	-0.489	0.222
	12	-.483*	0.117	0.002	-0.855	-0.112
	13	0.092	0.091	1.000	-0.197	0.380
	15	0.058	0.085	1.000	-0.211	0.328
15	5	-0.150	0.089	1.000	-0.433	0.133
	10	0.142	0.087	1.000	-0.136	0.419
	11	-0.192	0.116	1.000	-0.560	0.176
	12	-.542*	0.114	0.000	-0.904	-0.179
	13	0.033	0.071	1.000	-0.193	0.259
	14	-0.058	0.085	1.000	-0.328	0.211

\*The means are significant at the .05 level