

---

Theses and Dissertations

---

Summer 2012

# Learning, knowing, and doing classroom assessment: exposure and understanding rates of assessment knowledge among elementary pre-service teachers

Martin Swen Moe  
*University of Iowa*

Copyright 2012 Martin Swen Moe

This dissertation is available at Iowa Research Online: <http://ir.uiowa.edu/etd/3352>

---

## Recommended Citation

Moe, Martin Swen. "Learning, knowing, and doing classroom assessment: exposure and understanding rates of assessment knowledge among elementary pre-service teachers." PhD (Doctor of Philosophy) thesis, University of Iowa, 2012.  
<http://ir.uiowa.edu/etd/3352>.

---

Follow this and additional works at: <http://ir.uiowa.edu/etd>



Part of the [Teacher Education and Professional Development Commons](#)

LEARNING, KNOWING, AND DOING CLASSROOM ASSESSMENT: EXPOSURE  
AND UNDERSTANDING RATES OF ASSESSMENT KNOWLEDGE AMONG  
ELEMENTARY PRE-SERVICE TEACHERS

by

Martin Swen Moe

An Abstract

Of a thesis submitted in partial fulfillment of the requirements for the Doctor of  
Philosophy degree in Teaching and Learning (Curriculum and Supervision) in the  
Graduate College of The University of Iowa

July 2012

Thesis Supervisor: Professor Peter Hlebowitsh

## ABSTRACT

The use of assessments and assessment results is critical to classroom instruction. Teachers use the information gathered from assessments to guide instruction within their classrooms and to otherwise inform their judgment in areas of the curriculum requiring individual and group-based responsiveness and differentiation.

Much time and care go into preparing teachers to be able to instruct from the standpoint of a specific content area and age level. Unfortunately, this preparation does not always extend itself to understanding various assessment practices. The teaching of assessment practices should be a common experience in teacher education programs. This study aimed to determine what one large scale teacher education program situated in a large public university did to advance the skills of assessment knowledge among its elementary pre-service teachers. The focus was on the perceived exposure and attained knowledge levels of pre-service teachers on assessment topics. The perceived levels of exposure and knowledge were compared to the beliefs that elementary principals held on the importance of the same assessment topics. Surveys were given to student teachers and principals in a way that allowed the information to be compared.

The analysis of the survey results found that the exposure levels reported by the pre-service teachers was lower than the importance levels placed on the respective topics by the principals. However, the pre-service teachers reported higher levels of knowledge on key assessment topics than the levels of knowledge that principals believed beginning teachers should possess on respective assessment topics.

Principals felt strongly that it is important for beginning teachers to enter their teaching careers with knowledge of various assessment topics and skills. In many cases,

there were misalignments between what principals identified as important and what pre-service teachers were exposed to. The schisms have implications for teacher education, as they speak to a potentially different way to design what assessment knowledge gets taught and how it gets taught at the pre-service level.

Abstract Approved: \_\_\_\_\_  
Thesis Supervisor  
\_\_\_\_\_  
Title and Department  
\_\_\_\_\_  
Date

LEARNING, KNOWING, AND DOING CLASSROOM ASSESSMENT: EXPOSURE AND  
UNDERSTANDING RATES OF ASSESSMENT KNOWLEDGE AMONG ELEMENTARY  
PRE-SERVICE TEACHERS

by

Martin Swen Moe

A thesis submitted in partial fulfillment of the  
requirements for the Doctor of  
Philosophy degree in Teaching and Learning (Curriculum and Supervision) in  
the Graduate College of  
The University of Iowa

July 2012

Thesis Supervisor: Professor Peter Hlebowitsh

Graduate College  
The University of Iowa  
Iowa City, Iowa

CERTIFICATE OF APPROVAL

---

PH.D. THESIS

---

This is to certify that the Ph.D. thesis of

Martin Swen Moe

has been approved by the Examining Committee  
for the thesis requirement for the Doctor of  
Philosophy degree in Teaching and Learning  
(Curriculum and Supervision)  
at the July 2012 graduation.

Thesis Committee: \_\_\_\_\_

Peter Hlebowitsh, Thesis Supervisor

\_\_\_\_\_  
Bruce Fehn

\_\_\_\_\_  
Gregory Hamot

\_\_\_\_\_  
David Bills

\_\_\_\_\_  
Marcus Haack

To Donna, Jakob, Elias, and Jude

## TABLE OF CONTENTS

LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
CHAPTER I: INTRODUCTION.....	1
Foundation of the Study .....	1
Introduction to the Problem.....	5
Assessments and Teacher .....	10
Teacher preparation programs and assessments.....	12
A Disconnect Between Curriculum, Instruction and Assessments .....	15
Purpose of the Study.....	17
Rationale for the Study.....	17
Significance of the Study .....	19
Research Questions .....	20
CHAPTER II: LITERATURE REVIEW .....	22
History of Assessments .....	22
Accountability and Assessment – Standardized Tests Remain a Focus .....	28
Role of Assessment in Education.....	40
Assessment Standards .....	46
Assessment: Are teachers prepared to assess students? .....	49
CHAPTER III: METHODOLOGY .....	58
Introduction .....	58
Research Questions .....	58
Instrument Development .....	59
Selection of Participants.....	63
Survey Development and Collection.....	66
Data Analysis .....	68
Limitations of Study.....	70
CHAPTER IV: RESULTS.....	74
Question I – To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skills? .....	75
Question II – Among elementary school principals, what assessment practices are most important to know among beginning teachers working in the classroom.....	91
Question III – How do elementary school principals’ expectations for teacher assessment knowledge compare with self-reported pre-service teachers’ levels of exposure.....	104
Question IV – What are the self-reported levels of understanding among	



pre-service teachers of various assessment concepts and skills? .....	117
Question V – How do elementary school principals conceive of assessment understanding for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers? .....	121
Question VI – What are the attitudes of pre-service teachers toward the uses of standardized achievement results in the classroom? .....	129
Question VII – What are the attitudes of elementary school principals toward standardized achievement results in the classroom? How congruent are these with the attitudes of pre-service teachers? .....	131
 CHAPTER V: RESULTS DISCUSSION .....	 136
Introduction and Description of the Study .....	136
Analysis .....	137
Assessment Exposure .....	138
Assessment Knowledge .....	149
Perceptions of Standardized Assessments .....	152
Implications .....	158
Further Research .....	167
 APPENDIX A – INVITATION FOR PRE-SERVICE TEACHERS TO COMPLETE SURVEY .....	 171
 APPENDIX B – INVITATION FOR PRINCIPALS TO COMPLETE SURVEY .....	 172
 APPENDIX C – ADMINISTRATOR SURVEY – SECTION I – IMPORTANCE OF ASSESSMENT KNOWLEDGE .....	 173
 APPENDIX D – ADMINISTRATOR SURVEY – SECTION II – PERCEPTIONS OF STANDARDIZED ACHIEVEMENT ASSESSMENTS IN ELEMENTARY SCHOOLS .....	 176
 APPENDIX E – ADMINISTRATOR SURVEY – SECTION III – PERCEPTIONS OF NEW TEACHERS’ EXPERTISE IN ASSESSMENT WHEN THEY ENTER THE TEACHING PROFESSION .....	 177
 APPENDIX F – PRE-SERVICE TEACHER SURVEY – SECTION I – INSTRUCTIONAL EXPOSURE TO ASSESSMENT KNOWLEDGE .....	 178
 APPENDIX G – PRE-SERVICE TEACHER SURVEY – SECTION II – PERCEPTIONS OF STANDARDIZED ACHIEVEMENT ASSESSMENTS IN ELEMENTARY SCHOOLS .....	 180
 APPENDIX H – PRE-SERVICE TEACHER SURVEY – SECTION III – PERCEPTIONS OF ASSESSMENT KNOWLEDGE .....	 181

APPENDIX I – QUESTION I – CATEGORY #1 .....	182
APPENDIX J – QUESTION I – CATEGORY #2 .....	183
APPENDIX K – QUESTION I – CATEGORY #3 .....	184
APPENDIX L – QUESTION I – CATEGORY #4 .....	185
APPENDIX M – QUESTION I – CATEGORY #5 .....	186
APPENDIX N – QUESTION II – CATEGORY #1 .....	187
APPENDIX O – QUESTION II – CATEGORY #2 .....	188
APPENDIX P – QUESTION II – CATEGORY #3 .....	189
APPENDIX Q – QUESTION II – CATEGORY #4 .....	190
APPENDIX R – QUESTION II – CATEGORY #5 .....	191
APPENDIX S – QUESTION V – ITEM 3.7 .....	192
APPENDIX T – QUESTION V – ITEM 3.6 .....	193
APPENDIX U – QUESTION V – ITEM 3.3 .....	194
APPENDIX V – QUESTION V – ITEM 3.4 .....	195
APPENDIX W – QUESTION V – ITEM 3.1 .....	196
APPENDIX X – QUESTION V – ITEM 3.5 .....	197
APPENDIX Y – QUESTION V – ITEM 3.2 .....	198
BIBLIOGRAPHY .....	199

## LIST OF TABLES

Table 3.1.	Research Questions – Component Alignment .....	63
Table 3.2.	Distribution of the Pre-service Teacher Sample Population.....	64
Table 3.3.	Principal Sample Population.....	65
Table 3.4.	Component Analysis.....	70
Table 4.1.	Question I – Assessment topics exposure ratings – 1 – 25 .....	91
Table 4.2.	Question II – Assessment topics level of importance #1 -25.....	104
Table 4.3.	Question III – Pre-service Teacher and Principal Comparison, Values of Corresponding Scales .....	105
Table 4.4.	Research Questions Comparison of Means and Range .....	135
Table II	Item Response Comparison .....	182
Table J1	Understanding the strengths and limitations of different assessment types, Item Response Comparison.....	183
Table K1	Constructing or evaluating test items, Item Response Comparison .....	184
Table L1	Interpretation of different score results, Items Response Comparison ....	185
Table M1	Exposure to key assessment terms, Item Response Comparison.....	186
Table N1	Create and use different assessment methods.....	187
Table O1	Understanding the strengths and limitations of different assessment types .....	188
Table P1	Principals – Learning how to construct or evaluate test items.....	189
Table Q1	The interpretation of different types of standardized score results, item response comparison.....	190
Table R1	Knowledge of key assessment terms, item response comparison.....	191

## LIST OF FIGURES

Figure 4.1.	Question I – To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skills .....	77
Figure 4.2.	Question I - Category #1 – Exposure to the creation and use of different assessments .....	80
Figure 4.3.	Question I - Category #2 – Understanding the strengths and limitations of different assessment types .....	83
Figure 4.4.	Question I - Category #3 – Learning how to construct or evaluate test items .....	85
Figure 4.5.	Question I - Category #4 – Exposure to the interpretation of different types of standardized score results .....	87
Figure 4.6.	Question I – Category #5 – Exposure to key assessment terms .....	89
Figure 4.7.	Question II – Among elementary school principals, what assessment principles and practices are most important to know among beginning teachers working in classrooms?.....	93
Figure 4.8.	Question II - Category #1, Principals - Create and use different assessment methods .....	95
Figure 4.9.	Question II - Category #2: Understanding the strengths and limitations of different assessment types .....	97
Figure 4.10.	Question II - Category #3, Principals: Learning how to construct or evaluate test items .....	99
Figure 4.11.	Question II - Category #4, Principals - The interpretation of different types of standardized score results .....	100
Figure 4.12.	Question II – Category #5 – Knowledge of key assessment terms.....	102
Figure 4.13.	Question - Category #1 - Pre-service teacher and principal comparison - Create and use different assessment methods .....	107
Figure 4.14.	Question III - Category #2 - Principal and Pre-service teacher comparison - Understanding the strengths and limitations of different assessment types.....	109
Figure 4.15.	Question III - Category #3 – Principal and Pre-service teacher comparison - Learning how to construct or evaluate test items.....	111

Figure 4.16.	Question III - Category #4 - Principal and Pre-service teacher comparison - The importance of being able to interpret different types of standardized score results.....	113
Figure 4.17.	Question III - Category #5 - Principal and Pre-service teacher comparison - Exposure to key assessment terms .....	115
Figure 4.18.	Question III - How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure? .....	117
Figure 4.19.	Question IV - What are the self-reported levels of understanding among pre-service teachers of various assessment concepts and skills?.....	119
Figure 4.20.	Question IV - Pre-service teacher perceptions of assessment knowledge compared to levels of exposure .....	121
Figure 4.21.	Question IV - How do elementary school principals conceive of assessment practices for beginning classroom teachers?.....	127
Figure 4.22.	Question V - How do elementary school principals conceive of assessment practices for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers? .....	128
Figure 4.23.	Question VI – What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?.....	130
Figure 4.24.	Question VII – What are the attitudes of elementary school principals toward standardized achievement results in the classroom?.....	132
Figure 4.25.	Question VII - Principal and Pre-service teacher comparison – Attitudes towards the use of standardized achievement results in the classroom .....	133
Figure S1.	Sufficient knowledge on how to use assessment results to help plan instruction, Item Response Comparison .....	192
Figure T1.	Sufficient knowledge to use data to interpret test results for students and parents, Item Response Comparison .....	193
Figure U1.	Sufficient knowledge to use data to interpret test results, Item	

	Response Comparison .....	194
Figure V1.	Sufficient knowledge to use performance assessment methods, Item Response Comparison.....	195
Figure W1.	Sufficient knowledge of classroom assessments and procedures, Item Response Comparison.....	196
Figure X1.	Sufficient knowledge to grade student work, Item Response Comparison.....	197
Figure Y1.	Sufficient knowledge to write sound test items, Item Response Comparison.....	198

## CHAPTER I

### INTRODUCTION

#### **Learning, Knowing, and Doing Classroom Assessment: Exposure and Understanding Rates of Assessment Knowledge Among Elementary Pre-service Teachers**

##### **Foundation of the Study**

After graduating from the University of Iowa in 1997, I felt prepared for the professional experience I was about to undertake. I was embarking on an education career by taking a job as a fifth grade classroom teacher with the Chicago Public Schools (CPS). Teaching within CPS brought more challenges and learning experiences for me than I could have imagined or felt prepared to handle. While I felt like I became a good teacher, there were aspects to my teaching I would have liked to have gotten better at.

Through my interactions with students, I learned how to deal with classroom management issues and how to teach students that possessed a wide range of reading and writing skill levels, a dearth of parental involvement, and a lack of peer support school. At the same time, while these are all critical issues, perhaps the reality, with which I was least prepared, was the high level of importance placed upon standardized assessments and the difficulty inherent to assessing what my students were learning in my classroom. Philosophically speaking, I believed that students need to learn a core curriculum in an engaging and relevant manner, and to have an experience that continuously built upon itself. I especially appreciate the work of John Dewey and his emphasis on the ‘continuity of experience’ that so deeply affects the quality of the teaching and learning (Dewey 1938, 27). The experience that we, as teachers, present to our students sets the stage for how they will embrace education. Educative experiences provide students with

a desire to continue their education and to build and maintain their ‘love of learning.’ Ralph Tyler made this same point by stressing that learning experiences should, “offer satisfying activities that build an appreciation for learning and for the content and skills embodied in those activities,” (Hlebowitsh 2005, 161). I never envisioned myself in a situation where I would spend so many hours reviewing basic skills and having students do hours of rote test preparation. This type of teaching did not provide meaningful experiences to the students; nor did it do anything to instill a desire to learn. Yet it happened to me. And it seemed to me that the idea of assessment was largely responsible for the problem, as it was pushing students away from learning.

A key task given to us each year at my former school was to ensure that students showed growth on the annual standardized assessments. We spent a minimal amount of time discussing what the scores actually meant; the push was always to get the scores higher. One part of the process embedded in my memory includes receiving previous editions of the assessments and having students take and retake the measurements. We also provided practice booklets for the students, who would complete these two, three, or more times. The reasoning behind this was to have students familiar with the assessment and format of the questions, so that when they were taking the ‘real’ test, they would not feel that it was something they had never experienced before. The end of January typically marked the time to set aside class time allotted for science, social studies, or other ‘extra’ curricula, so that we could focus on reading, writing, mathematics, and test preparation. Students were bored, and teachers were bored. Score inflation tactics prevailed; but actual learning was hardly evident.



My students' scores always did trend in the right direction. But, as a teacher, the scores did not mean anything to me. First, I did not know how to use the results to pinpoint the areas where my students' needed additional help or where they were already comfortable. I was not analyzing the data to see how my teaching should change, what areas were not being learned, or how I could better prepare the students. I did not analyze the scores of the students coming into my classroom the next year to better prepare to teach the critical areas in which they needed the most support. I simply saw the scores and celebrated the fact that they increased. I was not alone in thinking of the assessment results in this manner. I do not recall any teacher at my school at that time that critically analyzed the results or used them to improve their teaching within their classrooms.

The lack of knowledge that I had in relation to the purpose and use of standardized assessments and their results extended to the more frequently used quizzes, tests, and other assessments I gave my students. In most instances, the assessment given to my students was something identified or created at the last moment. Assessments were not thought out or intricately aligned to the curriculum taught. There was no effort to ensure that the grades or components that were used to grade were proportionate or aligned with what was taught in the class. Little was done to differentiate the assessments based on student needs or in turn to differentiate the curriculum based on assessment results. Too often the assessments were scored after a new subject or topic was already being covered, leaving little to no opportunity to use the results to guide instruction.

A related issue was that I didn't fully comprehend what all could serve as an assessment. My perception was that quizzes, tests, and formal projects were the assessments and everything else was either class work or homework. This was a major oversight on my part. However, this experience is what has led to the topic of this research. I first realized all of the invalid uses of assessment results that I used in my teaching experience when I began taking coursework in my graduate studies. This was the first time that the "light came on" for me in this area. The course, Introduction to Educational Measurement, taught by David Frisbie, was an enlightening experience for me. It brought me right back to my classroom and constantly had me thinking of all of the missed opportunities that I had had. Moving back to the Chicago Public School system, this time as a central office administrator, has allowed me to talk with teachers and principals about the topic. I continue to hear a similar story-line, that teachers don't fully grasp the power and attributes of assessments within classroom instruction. While there is a heavy emphasis on analyzing data today, it is mostly in relation to the battery of standardized assessments given to our students, and mainly done in order to ensure that test scores trend up. A major focus remains on standardized assessment results and these are lost opportunities for teachers to improve instruction.

All of this helps to highlight the general lack of preparation teachers have in using assessments and assessment results to improve classroom instruction. Assessment was not a topic that I considered to be critical to teaching. It was simply something that was done in order to have a grade for the students and to see which students were ready to advance to the next grade level. Many of the education courses did provide discussions on the various assessments available, but at least for me, the connections to the critical

role assessment plays was not there. This is one of the major regrets that I have as I look back on my time as a pre-service teacher. I was able to improve my ability to assess students as I gained experience, but this was due to my maturation as an educator and not related much to what I learned in my teacher education program. Assessment is a critical component of education and an area that pre-service teachers must be properly prepared to use within the classroom. This is an area that teacher education programs can help students to better understand.

### **Introduction to the Problem**

Assessment is a term that catches the attention of the public, in large part because of its association with the concept of high stakes standardized tests and the *No Child Left Behind (NCLB)* legislation. The term can have a frightening dimension, as many educators do not have a true understanding of how assessments work or how they should be used. To use assessments properly calls for a basic understanding of various testing and instructional principles. Rick Stiggins (1999) identifies six classroom assessment competencies educators should possess, including:

- 1) To understand the full range of users and uses of the assessment;
- 2) To be clear about the achievement targets students are expected to hit;
- 3) To be prepared to use a full range of assessments;
- 4) To understand how to assemble the exercises of the chosen method to sample student achievement;
- 5) To understand the sources of bias that exist within each assessment;
- 6) To understand the relation between assessment and student motivation

Assessments are an integral piece of teaching and when utilized properly the benefits to teachers are significant.

Assessments are vital to helping educators learn about various strengths and weaknesses of their students. In reality, teachers assess students every day through a

variety of means. Studies show that teachers spend anywhere from 30 – 50% of their instructional time assessing students (Schafer 1993). The key to making the assessment valuable depends upon the fidelity with which they implement the various means of assessments and how they interpret and use the results. The definition provided in the *Standards for Teacher Competence in Educational Assessment of Students* (1990), states that assessment is "the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy," (1). Simply put, teachers must constantly gather evidence to help determine what students are learning, what they are missing, and what might need to be done to make instructional adjustments- both for the entire class and individual students.

Many teachers equate assessment with the infamous once-a-year standardized tests given in schools, the basic chapter tests and quizzes found within many textbook series, or even with those created by individual teachers. Whether these associations are brought about by the immense media coverage of how well students perform, by state mandates to attain certain scores, or by principals and parents who are preoccupied with score results, the effect for teachers is always the same - producing score improvements. However, to limit assessment to such a narrow view is a gross misrepresentation of what assessments can do and how they should interact with the educational process. Madaus and Kelleghan stress that assessment is what allows us to see what a person knows or can do (Hlebowitsh 2005). This does not mean what students can do once a year, but what students can do over time, along comprehensive lines. This is not an argument that standardized tests are not worthwhile, but rather that they are one piece of the assessment

puzzle teachers must use to guide classroom instruction. In this way, it is imperative that teachers are knowledgeable about assessments and their various uses.

Mandated assessments do serve important purposes, including but not limited to monitoring student achievement, providing accountability, reporting to parents, and determining grade level promotions (Linn and Gronlund 1995). Teachers and other educators must appreciate the purpose of assessments and their benefits in order to ensure that there is an accurate view of what students are learning. The assessment literacy that teachers possess will affect their ability to properly use assessments. Decisions made by the classroom teacher are never ending and assessments provide the tool to ensure these decisions are informed (Nitko 2001).

The teaching and learning process is complex and a part of “*a continuous and interrelated series of instructional decisions concerning ways to enhance student learning,*” (Linn and Gronlund 2000, 31). Linn and Gronlund, as well as the other measurement specialists, point out the importance of assessments in the classroom decision making process, (Popham 1999; Schafer 1991; Stiggins 1991). Popham (1999) describes four traditional purposes used to describe the need for assessment that dates back to the 1950s and 1960s. These include diagnosing students’ strengths and weaknesses, monitoring students’ progress, assigning grades, and determining one’s own instructional effectiveness. By assessing the strengths and weaknesses of students, the teacher can make decisions about which instructional objectives to pursue. When assessing the progress of the students, a teacher may decide which aspects of the instructional program need to be changed. The assigning of grades based upon the assessment used is obvious, as are using assessment results to determine the success of

instruction (Popham 1999). This process is one that is not always easy and contains many decisions and outcomes. As teachers have an ever greater array of assessments at their disposal, the likelihood that they will have one or more that will provide them with information that will benefit the teaching and learning process increases. For this reason, Linn and Gronlund (2000) point out that it is important to have the knowledge of many different types of assessment data available to use for different situations.

*NCLB* put an emphasis on information that can be derived from assessment results. While this can provide a positive effect for schools, there are concerns as well. On this positive note, there is a push to develop assessment systems that align with standards that are used to guide student learning. There seems to be some recognition that these assessments must capture higher-order thinking skills, provide more accurate measures of student growth, and better inform classroom instruction (U.S. Department of Education 2010). Additionally there is a call to “elevate the teaching profession” by highlighting and rewarding ‘excellence’. States and districts are being asked to develop and implement practices that will evaluate teachers and principals and identify the most effective ones based on student growth and other factors (4).

The decisions made using assessment results are not limited to teachers but include the public and educational leaders as well. The assessment results attendant to the ESEA Reauthorization Act or *No Child Left Behind (NCLB)* exemplify this. Today, the results stemming from the state assessments associated with *NCLB* are scrutinized by federal and state officials, as well as by the general public. Title I schools are receiving sanctions for low student achievement scores and non-Title I schools with low student scores are labeled as low-achieving by the media. Obviously this use of assessments has

had an effect on the dynamics of assessment and their role in the educational process. President Obama and Secretary of Education Arne Duncan were working to enact a new version of the Reauthorization of the Elementary and Secondary Education Act. This act was going to place an even greater emphasis on standardized assessment results and seek to enhance reforms made over the past years that were spurred by the administration's American Recovery and Reinvestment Act of 2009. There were four key areas:

- 1) Improving teacher and principal effectiveness;
- 2) Providing information to families to help them evaluate and improve their children's schools, and to educators to help them improve their students' learning;
- 3) Implementing college- and career-ready standards and developing improved assessments aligned with those standards; and
- 4) Improving student learning and achievement in America's lowest-performing schools by providing intensive support and effective interventions (U.S. Department of Education 2010)

While the Reauthorization has stalled, Obama's administration has moved its educational reforms across the country through the use of the Race to the Top (RTTT) program. RTTT moves the increased federal involvement in public education that came about through the implementation of *NCLB*. Cash-strapped states were offered the possibility to tap into the \$4 billion set aside for those that qualify for the grants in exchange for writing proposals that meet the four federal priorities. RTTT had similarities to *NCLB* but added components including what became the eventual adoption of the Common Core State Standards and high-quality teachers no longer being defined by only have proper credentials, but also be evaluated based on student performance (Manna and Ryan 2011).

The Obama administration's effort to install national content standards in literacy and math also has assessment implications. Most states are now moving forward with the

development of national group of *Common Core State Standards (CCSS)*. The CCSS were developed with a collaboration of the Gates Foundation, the Council of Chief State School Officers and the National Governors Association (Phillips and Wong 2010). Though the CCSS were not developed by the federal government, they do support their implementation and have invested a lot of money through Race to the Top and other federal investments in state assessment systems. Central to the CCSS is the alignment of assessments. With common standards across a vast majority of the states, national assessments are now being developed. Phillips and Wong articulate that the assessments tied to the CCSS will allow for more good practice, because unlike *NCLB*, there will (hopefully) not be the prescriptive accountability measures tied directly to the results (2010). Additionally, the standards are being designed to be more conceptual and less procedural, which will allow for a different type of assessment. Sample instructional packages are being developed which include curriculum (modifiable) that are tied to performance tasks (to be used in a formative manner) along with summative end-of-year assessments that are not intended to be used for accountability purposes (2010).

### **Assessments and Teachers**

The term assessment encompasses an array of definitions, concepts, and activities. The question arises, what do teachers need to know about assessments in order to be effective in the classroom? Further, whose responsibility is it to ensure that teachers are knowledgeable of assessment concepts? Ronald N. Marso and Fred L. Pigge (Wise 1993) addressed the issue of teachers' testing knowledge. They noted that before standards for teacher competence in the assessment were published in 1990, the testing community had not provided clear expectations for classroom teachers. However,



Diamond and Fremer (1989, as cited by Marso and Pigge), note that these standards can be found in the 1985 Standards for Educational and Psychological Testing, jointly developed by the AERA, the APA, and the NCME. These standards have since been enhanced by the 1988 Code of Fair Testing practices in Education, again jointly sponsored by the previously mentioned three professional organizations. The Code is focused upon standardized educational testing but addresses the practices of both test developers and test users. The main function is to address test and test score misuses. Neither the Code nor the standards address teacher-devised testing (1993).

Rick Stiggins (1985, as cited by Marso and Pigge) observed that the measurement community has provided less professional guidance for teacher-made testing than it has for standardized testing. The lack of focus on teacher-devised testing has occurred despite of the fact that the measurement profession perceives teacher-made tests, not standardized tests, to be the foremost assessment influence in K-12 classrooms (Wise 1993). The measurement community, as it turns out, has concerns about the assessment knowledge of professional educators. An example of this can be seen by Diamond and Fremer (1989, as cited by Marso and Pigge) who noted that the Institute for Research on Teaching, which coordinated the development of the previously described fair testing code, was critical of the inadequate training of educational personnel in the interpretation and use of tests (Wise 1993).

Concern over the lack of assessment knowledge held by teachers is not a recent phenomenon and has existed for decades. Rita O'Sullivan and Marla K. Chalnick (1991, as cited by Marso and Pigge) note that in 1955, Noll, Thorndike, and Hagan suggested all teachers should know how to integrate measurement and evaluation knowledge with

instructional knowledge. Gullickson (1986 as cited by Marso and Pigge) identified this concern over the inadequacy of teacher skill in the area of testing and evaluation dates back to James Conant's book, *The Education of American Teachers* (1963); to Mayo's survey (1964) of principals, superintendents, and professors of what teachers ought to know about testing; and to Mayo's (1967) survey of pre-service teachers on what they know about classroom testing (Wise 1993).

### **Teacher Preparation Programs and Assessments**

The perceived lack of teacher knowledge in regards to assessment can be attributed to any number of influences. One area to examine is teacher preparation. Gullickson and Hopkins (1987, as cited by Marso and Pigge) argue that pre-service instruction in educational assessment is simply not up-to-par. One problem is that courses that specifically aim at assessment tend to put a narrow focus on statistical manipulations. While statistical concepts are an important feature of assessment, these can be taught with a minimal emphasis on the computational components, in the interests of putting more emphasis on application concerns (Wise 1993). They also note that assessment coursework in the pre-service curriculum occurs too early and fails to take advantage of the practicum and student teaching contexts, where application opportunities are more relevant (1993).

Richard J. Stiggins (1991) expands on this point. Using informal surveys of in-service teachers, Stiggins found that approximately 50% to 75% of in-service teachers have taken at least one course in measurement and evaluation, as a part of their undergraduate or graduate professional preparation program. Less than 1% of that population said that their course work in this area has provided them with a "relevant and helpful experience

in terms of preparing them to meet the day to day demands of classroom assessment," (7). The magnitude of this failure is appreciated when one realizes that teachers spend more than one third to one half of their professional time dealing with assessment related activities (Stiggins 1998).

The lack of articulation between the classroom needs of teachers and the content of pre-service assessment courses is commonplace. Studies by Airasian and Madaus (1983) and Stiggins and Bridgeford (1985) show that typical measurement training does not address the practical classroom demands that teachers face. This in part explains the lack of interest by teachers to have better training in assessment. Gullickson (1986), in fact, identified discrepancies between college measurement course topics and practicing teachers' perceptions of what testing topics and skills are actually needed in the classroom. Classroom teachers perceive informal observations and direct pupil communications as forms of assessment that are central to instructional decisions and that require little to no statistical know-how. Gullickson (1986) also found that pre-service measurement instruction typically focuses more on paper-and-pencil assessments and the statistical analyses of data than on the informal assessments teachers prevalently use (Wise 1993).

Other studies support Gullickson's (1986) findings. Gullickson and Ellewein (1985) and Marso and Pigge (1988) found few practicing teachers use statistical analysis procedures in interpreting pupil test performance (Wise 1993). Kellaghan, Madaus, and Airasian (1982) reported that measurement training has not made much of an impact upon teachers' testing practices, and concluded that it is unlikely to do so unless the

training focuses on the actual demands of pupil assessment in the classroom, as seen by teachers, (1993).

Teachers tend to show a concern for assessment issues that deal directly with classroom instructional decisions. The focus is on questions such as: "How do I best prepare the test for a given course? How do I use test information to make specific kinds of decisions? Or how do I evaluate ongoing classroom actions?" These are the types of issues that teachers typically address in the classroom. Gullickson (1986) found that pre-service measurement classes, on the other hand, tend to show more concern for the ways that test results can be analyzed, summarized, and used to improve test quality. Standardized testing issues were also important in these classes and included topics such as differentiating between norm- and criterion-referenced tests, and dealing with concepts such as norms, norm interpretation, validity, and reliability (Wise 1993). Obviously there is overlap between what is desired by teachers and what is offered by many professors, but the differences are striking.

Interestingly, all pre-service teachers do not necessarily take a measurement course. Gullickson (1986) points out that many teachers have only a minimal exposure to educational measurement in their pre-service courses. A survey by Gullickson and Hopkins (1987) showed that about 70% of teacher education programs offered a measurement course. Of these, only about three fourths required such a course for the pre-service teachers. And when the course is optional, typically less than one quarter of the students will use that option (1987). Schafer and Lissitz (1987) state that in many teacher education programs the teaching of assessment resides within other courses and not in a separate course. Thus pre-service teachers may receive their assessment

instruction from professors who do not have expertise in that area (Quilter 1999). As many pre-service teachers learn about assessment from professors without assessment expertise, it brings up the question of how much they are learning about assessment in these courses. Additionally, course content within teacher education programs are affected by licensure requirements set by states. States where assessment competence is required will see more assessment content within teacher education courses (Trevisan 2002).

### **A Disconnect Between Curriculum, Instruction, and Assessment?**

Three important components of teaching consist of curriculum, instruction, and assessment. They all are connected in a number of ways and depend upon each other to provide students with a sound education. However, a problem arises when the educators do not possess equal strengths and knowledge in each of the three component areas. As W. James Popham (2004) put it, "educators who work in one of those three specializations are not only unknowledgeable about what goes on in the other two specializations, but they are also blissfully complacent about their ignorance," (418).

The problem fundamentally is a lack of coordination between curriculum, instruction, and assessment. Popham indicates that this problem is largely a function of separate, specialized training. The evidence of the separation can be seen in the different programs, journals, and associations that exist for each of the specialties (2004). In order for teachers and other educators to be able to utilize assessments to their maximum potential, some coordination has to be forged between curriculum, instruction, and assessment. The experts in each of the fields need to work together to understand what

exactly teachers and educators need to understand in order to use assessments effectively.

Such imbalanced specialization is not likely to go away anytime soon. The recent call for accountability as featured in the current *NCLB* environment as well as the policies underway under President Obama's administration have put an undue focus on measurements of proficiency. We can currently see the focus placed on the standardized state tests used to make decisions about teaching, without much consideration given to broader instructional and curricular perspectives (Popham 2004). The *NCLB* legislation has focused on standardized assessment scores to determine the effectiveness of schools and districts. Not only are the scores on these assessments scrutinized by the state and federal governments, but by the general public as well. The literal existence of a school, Title I schools in particular, could potentially rest with the ability of teachers to 'get' their students to receive proficient scores on the state assessments. This leads teachers and schools to have a narrow focus on the specific skills being measured, typically related to math and reading achievement. This may then lead to a lack of focus on the normative dimensions of the schools that are not easily measured (Hlebowitsh 2005). The existence of the high stakes atmosphere created by the implementation of the *NCLB* legislation may cause assessment tunnel vision. It is important that teachers and their schools continue to utilize a variety of assessments in order to properly and fully assess students.

The education environment continues to move towards a greater focus on accountability as the Obama administration installs its own federal policy. The focus now is going to be more directly upon individual teachers. The increased use of performance-based pay and the value-added analysis of teacher effectiveness are at the

forefront of the new federal effort. It is obvious that good teachers are critical to a good education. Research has shown that a high quality teacher is the most significant impact upon student achievement. As teacher effectiveness increases, it is lower-achieving students who benefit the most (Goe 2007).

### **Purpose of the Study**

Assessment is critical to the teaching process. It is important that our schools of education prepare future generations of teachers to be able to make valid judgments from assessments. And it's not just about the knowledge of assessments; it's also about making thoughtful applications of assessments in the classroom.

Assessments serve a variety of purposes and are indeed an integral part of the educational system. Teachers should not be afraid of them and should learn to apply their results widely. At a time when there are many concerns about the state of education, providing our new teachers with a greater understanding of assessment and assessment uses seems critical.

### **Rationale for the Study**

Assessments are recognized as an important component of teaching and learning, as documented in the previous section. At the same time, research identifies that this recognition is not necessarily articulated in the preparation of new teachers. Despite the studies that show that teachers spend up to 50% of instructional time assessing students, few states require pre-service teachers to take an assessment course and only a handful require new teachers to exhibit competence in assessment (Stiggins 1998).

Approximately 50% of the states are without any requirement for teachers to show assessment competence. Studies show that students within different teacher programs

receive varying levels of instruction tied to assessment. The schools, however, are putting a heavy emphasis on the use of standardized assessment results to make instructional decisions.

The purpose of this dissertation is not to look at what is wrong with the use of standardized assessments, or even the issues involved with *No Child Left Behind*. The purpose is to take a close look at how well aspiring teachers are prepared to deal with the concept and application of assessment and assessment results within the teaching and learning process. Are pre-service teachers receiving an introduction into this topic in a way that will allow them to begin to successfully use assessments in their classrooms? Do they fully understand the purposes behind assessments and the ways in which they can and should be used? Are they aware of the critical decisions that should not be made about students without properly assessing them? By taking a close look at the studies and scholarship on the topic and by taking a close look at one teacher education program, I will be able to say something about exposure and understanding rates of assessment knowledge among pre-service teachers.

The instructional leader in a school is the building principal. The school principal should have the pulse of each school – understanding the issues that exist in relation to the learning that is occurring within the classrooms. Her voice is an important one to consider when looking at whether pre-service teachers are ready to enter the teaching force. As they interact with the new teachers within their buildings and observe as they teach their classes, building principals have a first-hand opportunity to know whether their new teachers are using assessments successfully. It is important to hear the views of principals on whether the new teachers entering their schools are actually



prepared to use assessments wisely. This study provides a sample of perspectives of elementary school principals on the topic of teachers' assessment literacy levels.

The combination of the perspectives of the pre-service teachers and principals will allow for an examination of how closely the exposure and understanding rates of assessment knowledge among pre-service teachers report matches principals' expectations of performance in the area of assessment knowledge. Once a teacher has begun her teaching career she will have several opportunities to enhance her assessment literacy and ability. However, learning and mastering new assessment skills while also managing the demands placed on teachers is very difficult.

Ideally, a principal will serve as the instructional leader of the school and will identify the areas of support most needed by teachers. Assessment, which is not typically something someone can master by reading an article or attending a few short professional development workshops, is one area that will likely challenge a principal. This makes it critical that new teachers enter the profession with the knowledge base they will need.

### **Significance of the Study**

This study compares and contrasts reported rates of exposure to and understanding of assessment knowledge among pre-service teachers to principals' expectations of assessment knowledge.

Recent school reform efforts have put assessment measures at the center of attention. After *NCLB* was enacted, teachers were put under increased pressure to use standardized assessments for high-stakes accountability purposes. This led many teachers to focus on 'teaching to the test.' The adoption of CCSS has also put assessment concerns into play. The CCSS linked assessments will be based more on performance

assessments and on an increased reliance on formative assessments. Regardless of the national or state strategy that becomes the focus of the educational landscape, teachers will be put under increased expectations to use multiple forms of assessment and to understand the types of assessment needed in the classroom. There is a lot to learn and to be able to do. This makes it an especially critical feature of teacher education programs.

This study is a case study analysis of one teacher education program. It will provide an opportunity to identify potential areas of assessment that are not addressed and to begin a discussion of what pre-service teachers should know about assessment from the standpoint of the classroom.

### **Research Questions**

- I. To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skills, including:
  - a. exposure to creating and using different assessment methods;
  - b. exposure to understanding the strengths and limitations of different assessment types;
  - c. exposure to learning how to construct or evaluate tests and test items;
  - d. exposure to the interpreting different types of standardized test scores; and,
  - e. exposure to key assessment terms.
- II. Among elementary school principals, what assessment principles and practices are most important to know among beginning teachers working in classrooms?
- III. How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure?
- IV. What are the self-reported levels of understanding among pre-service teachers of various assessment concepts and skills?
- V. How do elementary school principals conceive of assessment practices for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers?

- VI. What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?
- VII. What are the attitudes of elementary school principals toward standardized achievement results in the classroom? How congruent are these attitudes with the attitudes of pre-service teachers?

## **CHAPTER II: LITERATURE REVIEW**

We have entered into an educational period some commentators refer to as the ‘Age of Accountability’. And principals and teachers are the two parties at the ‘front lines,’ as they are charged with the task of ensuring that students meet public accountability measures. As a result, educators need to have a thorough knowledge of assessment practices and principles.

Assessing students is not a clear cut or uncomplicated process. At the same time it is, and always has been, a process critical to a good education. This being the case, the question of how well educators understand and implement assessments and use the results to inform their teaching and learning is an important one. Assessments ‘for learning’ and ‘of learning’ must be used in coordination with one another to ensure that teachers understand what students know. Teachers must understand the purposes of the assessments they give. There is an increasing call from scholars in the assessment field for teacher and school administrator preparation programs to make certain that all their graduates are deeply literate and capable of using a wide range of assessment practices to promote and document student learning (Noonan and Renihan 2006).

### **History of Assessments**

Assessment has played a variety of roles in the history of schooling. Going back to the 19<sup>th</sup> century, teachers relied primarily on recitation to learn how well their charges learned the target material. Teachers were not all of one mind in judging recitations and would often allow factors related to other behaviors to influence their judgments (Giordano 2005). Arguments were made that the way to assess students’ learning lacked standardization. Individual teachers placed a greater emphasis on

different foci, and used different standards to judge quality. Opinions, not facts, guided the grading of many teachers, and the curriculum taught to students was idiosyncratic and depended upon the whim or interest of the teacher (Giordano 2005). Frances Lowell (in 1919, as quoted by Giordano 2005) noted that “teachers frequently persuade themselves that since a child is sweet, docile, and attractive, he must necessarily be bright...” (27).

The grading practices used in the schools during the early part of the 20<sup>th</sup> century were frequently questioned. Different experts found that many practices had no good relation to student achievement. In 1912, S.S. Calvin, found grades were assigned in an “inconsistent and unpredictable fashion” and too often the grades were given arbitrarily. A New York superintendent named Bliss noted frequent issuing of grades determined in part by personal bias (Giordano 2005, 118). Concerns over how students were graded and the actual meaning or usefulness of those grades helped set the conditions for the growth of standardized tests. The need for a reliable and consistent method of assessment did exist and continues to exist today. Research has shown that grading decisions are still based on informal observations of students that take into consideration criteria such as student effort, good behavior, extra credit, and so on (Remesal 2010).

The development of intelligence tests in the early parts of the 20<sup>th</sup> century represented some of the earliest efforts to bring standardized testing routines to the school. Intelligence tests started to gain popularity in the early 1900s. Their use was strongly influenced by the eugenics movement. Alfred Binet typically receives credit for developing the modern form of intelligence testing. He used higher-order cognitive tasks and identified two criteria for test item writing: 1) task performance had to increase with age, and 2) task performance had to be related to school achievement (Embretson 2003).

American psychologists such as Terman, Goddard, and Yerkes promoted intelligence test results as a scientifically exact measure of a fixed trait that conformed to the laws of Mendelian genetics and emphasized the biological nature of intelligence. Goddard (1920) and Terman (1916) believed in the ordering of individuals based on the scale of intelligence and believed it accounted for behavior and cognitive performance. The differences in capacity were to be dealt with by the use of a highly differentiated curriculum (Shepard 2000). Terman, responsible for adapting the Binet-Simon scale for use in the United States, made some methodological changes as well. He standardized the directions and instructions, and added internal consistency as a criterion to item selection. These efforts led to the development of the Stanford-Binet intelligence test, which remains in use today as the Stanford-Binet IV (Embretson 2003).

Tracking students for instruction by ability also began in the early twentieth century. The differentiation occurring within schools at the time was made easier by intelligence testing, and reflected the economic purpose attached to education at the time. By tracking, or differentiating the curriculum, the system would allow the construct of ability to be used to sort students into differentiated economic roles that they would presumably move into later in life. Here was a way, its proponents stated, to provide all students with an equal opportunity to develop to their fullest capacity (Urban and Wagoner, Jr. 2009). Unfortunately, it was not until later in the century that the potential harm of labeling children was brought up, (Hobbs 1975, cited by Sheppard 2000). The inaccuracy of classifications based on single tests (Education for all Handicapped Children Act of 1975, P.L. 94-142) was not even questioned. Today, scientists and the public are more likely to assign a much more limited role to heredity, though there are

still teachers, policymakers, and the public who see family background and cultural difference as fixed characteristics accounting for school failure. The use of readiness measures and achievement tests to categorize students' learning capacity has unfortunately resulted in simplified school experiences for students from lower socio-economic strata. "More subtly perhaps, the sorting and classification model of ability testing for purposes of curriculum differentiation has left a legacy that also affects the conception of assessment within classrooms. Even when aptitude measures are replaced by achievement tests, there is still the tendency to use test results to assign students..." (Shepard 2000, 8-9).

The value of testing is also tied to the problem of grading, which many see as too subjective and scattered to be meaningful (Giordano 2005). Many standardized tests claim to provide diagnostic information to teachers and educators. When used properly, such tests could help teachers identify student strengths and weaknesses, and give them an opportunity to develop remedial solutions to weaknesses (2005).

The use of standardized tests received additional support when the military began to use them as a tool to help determine whether recruits could serve effectively. The test was used to slot those who were deemed to have high intelligence so that they could be assigned to special tasks based upon their test results (Giordano 2005). The Army Alpha and Army Beta tests are commonly cited as the initial efforts at group intelligence testing. These tests, developed under the guidance of Yerkes (1920), differed from Binet's efforts by aligning more closely with Kilpatrick's vision. After World War I, a number of school districts followed the basic model of the Army tests and used a point scale method (Embretson 2003).

Testing was also used to deal with the practice of allowing students to gain admission into post-secondary institutions. Many believed that the enrollments in higher education had to be cut down and that a testing method to select students capable of succeeding in college was needed to make the selection process reliable. Standardized tests became common tools for college admissions and still are today (Giordano 2005).

The use of standardized tests in our school system continued to grow unimpeded. In 1928, Symonds noted fifteen applications for standardized tests. They included: informing pupils of their achievement, incentivizing study, promoting competition, determining promotion, diagnosing weak areas of pupil achievement, determining quality of instruction, determining admission to high school, placing pupils in appropriate school settings, determining admission to college, providing reports to parents, determining credits, honors, educational and vocational guidance, rating teachers, predicting a pupil's success, and studying the efficiency of the school (Giordano 2005). What is interesting is that over eighty years later, all of these applications are still prevalent in our educational system.

The growth of testing has also brought about improvements in the psychometric theory underpinning it. Spearman published a series of papers that provided fundamental new components for testing. He introduced the concept of reliability and the expanded relationship to validity, true and error variance, and test length. The Spearman-Brown formula, which corrects validity correlations due to unreliability, was first published in 1904, yet remains important today (Embretson 2003). Another critical concept was the use of internal consistency, which was used to improve reliability and was introduced by Terman in 1916. Scoring was yet another component that saw improvements. In 1914,



Kelly made a proposal to use *z*-scores to represent abilities, thus providing a normative standard scoring method. The use of a standard score system is still widely used today (2003).

The early intelligence tests were criticized. Giordano (2005) identified a variety of complaints - the tests were too easy, too hard, not independent of school training, too varied in the number of tests given to different age groups, not properly standardized, and given to incomparable groups. It was argued that intelligence tests were designed without a clear definition of what intelligence actually was and that the designers of the tests were making guesses at abstract mental principles. Others, including school board members, parents, and teachers expressed concern that such tests would lead to many students to be labeled as inferior. The debate on intelligence testing was, in fact, a contributing factor in the wider 'nature versus nurture' debate (2005). The issue of labeling continues today with the implementation of *NCLB* and its use of descriptors such as 'in need of assistance.'

Other criticisms noted the cost to secure the results and the understanding that not all components of learning can be tested. Issues of reliability and validity prevailed too. How could anyone decide to retain or to promote a child based off of one twenty-minute exam? (Giordano 2005) Criticisms were also directed at the format of the exam – how the questions were put together, how the administrative instructions were stated, and that how the exam forms were developed (2005).

Today, a major push has been put in place to use the results of standardized tests to determine the effectiveness of not only individual schools but individual teachers as well. Benchmarks and progress on the tests are used to identify schools that are failing

and point out individual teachers who are either succeeding or failing in their efforts to educate students. This is not a new phenomenon. During early to mid-parts of the twentieth century, various school districts would use student test scores to determine teacher quality. Brooks Buckingham (1922, as cited in Giordano 2005, 124) stated that, “at the same time, we were measuring the ability of the teachers to get results.” He compared this use of student test results to the product-based procedures for assessing the efficiency of workers in other occupations, such as doctors, lawyers, wood choppers, and farmers (Giordano 2005). Brooks, a New Hampshire superintendent, actually paid \$2 per week on bonuses to teachers who had students who earned high scores on their standardized tests. Of course, various critics agreed that this was an unfair practice and that teachers can’t determine the ability levels of the students they have in their classrooms, and that this ‘incentive’ would lead to teaching to the test (Giordano 2005).

#### **Accountability and Assessment – Standardized Tests Remain a Focus**

*No Child Left Behind* is often associated with the intense focus placed upon accountability and the examination of standardized test results. As noted previously, many educators have been promoting and using standardized tests in an ever-increasing fashion. Many of the concerns and issues raised in the early part of the twentieth century are still factors today. World Wars I and II both helped push the idea of using standardized tests for the purpose of predicting performance and increased the call to have tests to predict the best candidates for post-secondary education and high-skilled employment. The Scholastic Aptitude Test (SAT), beginning in the 1950s, was used by a number of post-secondary institutions for this purpose. Over time, the SAT scores were also seen as a method of measuring the quality of public schooling (Peterson and West

2003). Following World War II the general sense was, though not unanimous, that our public schools were prospering. As the SAT scores began a downward trajectory throughout the 1970s rumbling about the decline translated into a narrative that was critical of all public schooling, even though the decline had more to do with demographic factors (Peterson and West 2003).

In order to have a better overall gauge of student performance, the federal government funded the National Assessment of Educational Progress (NAEP), which was to be given to a random sample of students at ages nine, thirteen, and seventeen. The early results of the NAEP substantiated the SAT decline. The NAEP results showed, especially for seventeen year old students, that scores were either declining or stagnant. And the results for the younger students, while not showing a decline, also were not showing much growth. Further, international surveys of educational achievement were showing that the scores of the U.S. students began to fall behind other nations as the students got older (Peterson and West 2003).

Concern for the quality of our public education system was increasing as these achievement results seemed to all tell a negative story. This led to a greater call for accountability, and which reached its high point in 1982 when President Ronald Reagan made reforming the public schools a high priority. Secretary of Education Terrel H. Bell appointed a national commission that issued a report in 1983 called "*A Nation at Risk*" (Peterson and West 2003). Key issues in the report referred to poor performance on international tests, the illiteracy rate of 17 year-olds', low student achievement results, and the increasing call for remedial education (Linn 1986). While *A Nation at Risk* itself did not lead to any immediate federal mandates or legislation, the impact was quite

significant at the state level. Politicians representing different party affiliations focused on the demise of education in our schools and used it as a catalyst to call for a greater sense of accountability. In effect, the groundwork was being put into place for the NCLB legislation that would come almost twenty years later.

Under the direction of Margaret Spellings, the U.S. Department of Education released *A Nation Accountable: Twenty-five Years After a Nation at Risk (A Nation Accountable)* in April 2008. *A Nation at Risk* identified five key areas to change in education: curriculum content, standards and expectations of students, time devoted to education, teacher quality, and educational leadership and financial support of education. *A Nation Accountable* looked at the progress made in these areas. The report noted that by 2005 almost 65% of high school graduates would take the additional classes as recommended, but that a little over a third of these graduates still did not. Another issue within the curriculum content that was scrutinized was the issue of high school students having too many easy courses to choose from at the time of the initial report. The newer report found that this did not improve over the twenty-five years (U.S. Department of Education 2008). Standards and expectations were identified as the Commission felt that the educational community needed to have higher expectations and measurable standards. The new report found that this was an area that did improve over the time and many states began to implement standards-based education. In addition to the states developing standards, the federal government also worked to advance this area. Evidence of that can be seen from a 1989 meeting of the nation's governors organized by George H.W. Bush that discussed the adoption of national K-12 performance standards for the year 2000, and two initiatives during President Clinton's administration – the *Improving America's*

*Schools Act of 1994* and the *Goals 2000: Educate America Act* (1994) (Department of Education 2008). Teacher quality was another issue that was brought to our attention. Too many teachers presumably lacked the knowledge, skills, and training they needed to be successful in the classroom. The report called for linking teacher pay to teacher performance in the classroom. Some large districts began to experiment with this, including Denver and New York. Additionally, the Teacher Incentive Fund was a federal grant program developed and initiated under President George W. Bush and his administration (Department of Education 2008).

In 2001, the federal government reauthorized the Elementary and Secondary Education Act (ESEA) of 1965, now known as the *No Child Left Behind (NCLB) Act of 2001* (P.L. 107-110). *NCLB* is an important reform effort in that it was legislation that was supported by bipartisan majorities in Congress, as well as by the business and civil rights communities. This was legislation that was built on the efforts of the 1980's and 1990's (U.S. Department of Education 2008). The legislation set requirements for states to administer annual standards-based tests, measure students' yearly progress, and to disseminate report cards that described progress at the state, school district, and school levels. The act required that all states must administer both math and language arts test to all students in grades three through eight. Each state was given the latitude to either select or design its own test, as the test was aligned with each state's academic and performance standards (Popham 2002).

*No Child Left Behind (NCLB)* was developed in part to address a concern that our public schools and educational system were not providing all students with a quality education. *Goals 2000* provided a framework for standards-based reform. States would

move to develop content standards for the subjects, performance standards, and assessment standards to see what students can actually do. Goals 2000 was important for firmly establishing the standards-based reform movement on a national level; it was *NCLB* that moved it to an outcome-based accountability system (Valli 2008). While the reasons for the academic struggles of our students are vast and often intertwined, the thought that additional accountability would help improve the situation was clear. *NCLB* was developed in such a manner that it would place a heavy accountability role upon the teachers, schools, districts, and states to ensure that all students become proficient in math and reading. From the time of the initial year of the act, the 2001-2002 school year, states had 12 years (2012-2013) to ensure that all students meet or exceed a “proficient” level of academic achievement as defined by each state (Popham 2002). States were required by the law to establish a baseline of adequate yearly progress (AYP), using data from the 2001-2002 school year as the base. A minimum baseline was to be set using the performance of the lowest-achieving demographic subgroup or the school that scored at the 20<sup>th</sup> percentile in the state. Once the baseline was set, schools were to have all students reach proficiency within twelve years. States then had to establish goals for yearly progress (the first set by the end of year two and subsequently at least once every three years). The failure of any subgroup (economically disadvantaged, from major racial and ethnic groups, with disabilities, and with limited English proficiency) within a school to meet the state-determined state-set progress objectives would lead to that school not meeting AYP. However, if the subgroup that performed below expectations was able to reduce the proportion of the subgroup performing below the expectations by at least ten percent, the school will be regarded as having made AYP (Popham 2002).

The educational law built in what is essentially a five-year sequence of actions for schools unable to meet the requirements set forth. If a school was unable to meet the required progress for two consecutive years, the school received technical assistance, became eligible for federal school improvement funds, and provided students with the opportunity to transfer to another public school (not in need of improvement), within that school district. A third year of failing to meet the set requirements led to the provision of supplemental instructional services, such as tutoring to low-achieving disadvantaged students. A fourth year of failing to meet AYP required the school to design a plan of corrective action. This included options such as replacing certain staff members or implementing a new curriculum. The fifth year of failure to meet expectations required the school to restructure, which included the option of reopening as a charter school (Popham 2002).

A key instrument of *NCLB* in holding schools accountable in the public is that by 2002-2003, schools had to issue report cards to parents and the public. The information contained in those report cards included information about the students' performance on the annual tests. This information had to be disaggregated by subgroup; a comparison of students' performance relative to the state's annual objectives for student progress; two-year trend data on students' achievement; the percent of students not tested; graduation rates; information on teachers' qualifications, and information on the performance of school districts in the state that are making adequate yearly progress. The school-level report card was required to indicate whether the school was identified for improvement (Popham 2002).

*NCLB*, at its core, provides our school system with a push to ensure that our students are properly educated. Our schools have historically been faced with major gaps in the quality of education that exists between the races and between different socioeconomic groups. The achievement gaps that existed were not being systematically addressed, and *NCLB* provided the impetus for states, districts, and schools to begin to think about how to close those gaps. Additionally concerns existed that there was little alignment across student populations as to what students were exposed to. One of the benefits that came about with the implementation of *NCLB* was the increased alignment between teaching and the state standards for curriculum content. And districts, schools, and teachers began to actually use the student data that they had to help drive the educational decisions that were being made (Center on Education Policy 2008).

*NCLB* was not welcomed by all factions in the educational system. Critics claimed that the law simply created a high stakes testing climate in schools that would come at the expense of a comprehensive education. Concerns were also raised about the variability across the states in the difficulty of the state assessments and over the construct validity of the AYP proficiency cut scores (Peterson and West 2003).

Another major issue brought up had to do with the latitude teachers had to develop and implement their instructional plans. With the new accountability features, teachers would now be pressured to focus more acutely on the standards and skills mandated by the state (Peterson and West 2003). A concern that was brought up, and has proven to be a reality, is that a narrowing of the curriculum would occur. While there is no debate that the teaching of mathematics and reading are crucial to one's education, there are many non-tested features of the curriculum that are important to a child's



education too. The consequences attached to the *NCLB* legislation effectively caused many school districts, schools, and teachers to decrease the amount of time spent on these ‘other’ areas or to altogether delete them from the curriculum. Prior to the implementation of this legislation, evidence already suggested that this would happen. A 2000 RAND study found that an accountability reform movement in the state of Washington led to increases of 50 percent or more in classroom time devoted to reading, writing, and mathematics. Simultaneously declines of 40 percent or more were reported in the time spent teaching subjects including social studies, science, the arts, and health and fitness (Peterson and West 2003). A study completed by the Center for Education Policy in 2008 identified the number of minutes of instruction that were dedicated to the non-tested subject areas pre- and post-implementation-*NCLB*. Social studies went from 239 instructional minutes per week to 164, from 226 minutes to 152 minutes in science, from 154 minutes to 100 minutes in art and music, from 115 minutes to 75 minutes in physical education, and from 184 minutes to 144 minutes for recess (Berliner 2009). A study completed in Chicago by Hong and Youngs (2008) found that the district significantly narrowed the curriculum and moved away from an emphasis on working with students to help them improve higher-order thinking, writing, and problem-solving skills (Berliner 2009). Lipman (2004) reported that the narrowing of the curriculum was more often done in the lower-performing schools, which were heavily populated with poor, minority students. Lipman (2004) argued that the education of these students focused almost primarily on fragmented facts and learning basic test taking skills, while more affluent students were provided with a richer and more rigorous curriculum, (Berliner 2009). A study by Woodworth, Gallagher, and Guha (2007) focused on the

teaching of the arts in California and found that the amount of art education a student received was heavily dependent upon the student's income level. They also found that almost twice the number of students in low poverty schools received instruction in the arts as compared to students attending high poverty schools (Berliner 2009).

The amount of time spent or not spent on different subject matter was not the only influence that was attached to *NCLB* and other instances of high-stakes accountability reforms associated with standardized tests. The Center on Education Policy conducted a study in 2008 on the impact of *NCLB* by completing case studies at six schools in Illinois. The findings showed that the legislation impacted how teachers were teaching. Teachers within these schools reported that they were spending more time doing different forms of test preparation. They noted that they were spending time familiarizing students with types of test questions. They also found that teachers were not able to teach 'more creative, broader-themed, or project-oriented lessons' until after the state test was administered (Center on Education Policy 2008). There was also a finding of increased use of closed questions and more seat work. Adding to this, the findings showed that while both schools identified for improvement and higher performing schools used a substantial amount of time in this manner, those in the higher performing schools spent more time implementing a wider range of instructional practices (Center on Educational Policy 2008). One teacher noted that the accountability tied to the high-stakes tests gave credence to 'bad' teachers who now have support for their preferred teaching method – skill and drill - essentially providing 'a license for bad teaching' (Jones, Jones, and Hargrove 2003). As one teacher opined, "I'm not the teacher I used to be. I used to be great, and I couldn't wait to get to school every day because I loved being great at what I

do. All of the most powerful teaching tools I used to use every day are no good to me now because they don't help children get ready for the test, and it makes me like a robot instead of a teacher," (Jones, Jones, and Hargrove 2003, 43). Some evidence also indicated that teaching was focusing more narrowly on the students closest to the proficiency markers on the state tests. By focusing on these students, a school or teacher is more likely to have a greater number of students be proficient, thus increasing the likelihood of the school meeting the adequate yearly progress (AYP) (Center on Educational Policy 2008), and ironically, leaving behind the most and least advanced students.

A narrowing of the curriculum, an alteration in the teaching methods used, and a narrowed focus on a smaller set of students are a few of the negative factors that the *NCLB* legislation and other state-level accountability reforms have unleashed (Center on Educational Policy 2009).

President Obama and Secretary of Education, Arne Duncan, continued the push to reform education based on a standards and accountability-based focus. While noting that *NCLB* has its faults, they also feel that only a revision of it is necessary to continue to push forward with concept and increased accountability. In 2010, the U.S. Department of Education recommended a revision that focused on four key areas: 1) improving teacher and principal effectiveness, 2) providing information to families to help them evaluate schools and to teachers to help them improve student learning, 3) implementing college- and career-ready standards with improved and aligned assessments, and 4) improving student learning and achievement in the lowest performing schools (U.S. Department of Education 2010).

One key direction in the *Blueprint* was a call for all students to be prepared for college and career readiness. States were expected to upgrade the standards they had in place or to work collaboratively with other states to develop and adopt common standards. This is now being played out, as 46 states have agreed to adopt the Common Core State Standards (CCSS). The CCSS were developed with a collaboration of the Gates Foundation, the Council of Chief State School Officers and the National Governors Association. Although the CCSS were not developed by the federal government, it does support their use and incentivizes it through the Race to the Top initiative (Phillips and Wong 2010).

Two consortiums are now developing a standardized assessment aimed at measuring progress on the CCSS. One is the Partnership for Assessment of Readiness for College and Careers (PARCC). PARCC will be developing a summative end-of-year assessment, in addition to aligned formative assessments intended for classroom use. Members of this included twenty-six states (including Illinois). The other consortium is the Smarter Balanced Assessment Consortium (SBAC). SBAC will also be developing a summative end-of-the year assessment aligned to a series of interim and formative assessments to be used in the classroom. There are currently 31 member states in this consortium (some overlap with PARCC). Both of these consortia will be using performance-based and computer-adaptive assessments as key components of their design.

The expectations set forth in the *Blueprint* (and reflected in the CCSS) is that these assessment systems will be better at assessing higher-order thinking skills, provide more accurate measures of student growth, and better inform classroom instruction to

respond to the identified student needs. At the same time, while the focus is on language arts and mathematics, it also states that a focus is on ensuring a complete education for students and developing them into well-rounded students (U.S. Department of Education 2010). However, the accountability tied to the *NCLB* assessments, where the focus was also on language arts and mathematics, showed that the idea of a well-rounded education did not result. With *NCLB*, accountability was tied to the state assessment results. The implementation of the CCSS is not intended to be used for accountability purposes (Phillips and Wong 2010). Whether this plays out is another question.

The development and implementation of high quality assessments tied to the CCSS is a desirable outcome for the schools. The concern comes with the possibility that these assessments, though perhaps well-designed and providing valuable information, could continue down the path of serving as high stakes tests. This already seems likely. States will likely be asked to identify definitions of “effective teacher” and “effective principal,” using growth measures tied to the CCSS (U.S. Department of Education 2010). This will again, cause many teachers and principals to place a lot of emphasis on the standardized assessments. The concern will again arise about the narrowing of the curriculum and efforts by teachers and principals to game the system. While the planned revision of *NCLB* has stalled, the CCSS have moved forward and will be affecting the education system over the next several years.

An ever-increasing reliance upon standardized assessment results, concerns with the state of schooling, and a constant move to improve our educational system have caused standardized assessments to be used in ways that do not always support good education. Due to these influences, educators have placed unwarranted value on the

standardized assessment results, rather than focusing on the actual meaning of the results and how they can be used to improve teaching. Practices such as drilling students with practice tests, narrowing the curriculum, and teaching to the test have prevailed (Gunzenhauser 2003). Gunzenhauser explains that such reactions become a ‘default’ philosophy of teaching for educators, which results in a lack of reflective, engaged dialogue about teaching goals and practices. Such a default limits the educators’ opportunities to innovate, and inhibits educators from developing and establishing their own priorities and visions (2003). Elmore and Fuhrman (2001) found that the schools that were under the most pressure to perform on a high stakes assessment tended to have teachers who worked harder and spent more time preparing students for particular tests (Gunzenhauser 2003).

### **Role of Assessment in Education**

#### *Types and Purposes of Assessment in Education*

The level of assessment knowledge that teachers possess has rarely been addressed in the national dialogue on teaching. Teachers should be able to assess the achievement level of their students- both individually and as a class. They need to be able to use assessment data, to improve instruction and design assessment procedures to understand student progress.

Historically, educational assessments have been viewed as belonging to the domain of statisticians, and not of teachers. This was in part due to the technical requirements and knowledge needed to be able to work with the standardized tests. The *Handbook of Research on Teaching*, for instance, did not include a chapter on classroom assessment, or even tests and measurement. Research shows that traditional educational

measurement courses offered to pre-service teachers focus on measurement skills. These are important topics, but for many pre-service teachers, they find learning these skills as not relevant to their situation (Corr-Bremme 1983; Gullickson and Ellwein 1985; Gullickson and Hopkins 1987; Salmon-Cox 1981; Stiggins and Bridgeford 1985; Stiggins and Conklin 1988; Stiggins, Conklin, and Bridgeford 1986; as cited by Wise 1983).

The assessments that are used by teachers can be broken into different categories. Two broad categories of critical importance include formative and summative assessments. The terms formative and summative were fashioned in 1967 by Michael Scriven. He brought up the terms in a discussion on two key roles that evaluation has in schools. One role, formative assessment, is to address the “on-going improvement of curriculum” and the other, summative, to “enable administrators to decide whether the entire finished curriculum,..., represents a sufficiently significant advance on the available alternatives to justify the expense of adoption by a school system,” (Dylan 2006, 283). In 1969, Benjamin Bloom stated that the same characteristics could be applied to the evaluation of student learning. Today this would be referred as to student assessment. Bloom noted the traditional role that tests were typically used for were judging and classifying students. But he also noted that formative evaluation would provide “feedback and correctives at each stage of the teaching-learning process,” (Dylan 2006, 283). He also stressed that formative assessment would be more beneficial if it were a separate entity from grading and used mostly as an aid for teaching. Dylan (2006) goes on to stress that assessment is formative only if the information that a teacher collects is used to make changes that would not have been otherwise made. Thus assessments are formative only if something is contingent on their outcome.

Assessments can further be broken down to three levels- the classroom, school, and institutional-levels. Classroom assessments should provide a continuous loop of information that allows teachers, students, and parents to understand where a child is in reference to their learning in particular topic areas. These assessments vary and are presented in a variety of ways. The school-level assessments provide an opportunity for administrators, teachers, and even parents to have a better image of what the school population is learning. They provide a snapshot as to how well students are mastering certain topics or skills, as well as to allow for analysis on how well educational programs within in a school are working. School-level assessments are also beneficial in that they can provide an opportunity for collaboration between teachers to design, implement, and analyze common assessments (Stiggins and DuFour 2009). At the institutional level there is less of a focus on the individual student, or even individual classrooms, and more focus on the proportion of students who master identified standards to show institutional progress. The institutional assessments, often the state standardized assessment, are sometimes used to examine individual students or teachers. Each of the three levels of assessment can be considered a formative or summative assessment, depending on the purpose and use of the results. The important factor to consider is that each of these levels of assessment is equally important, and when used in coordination, they can bring good insight to the school (Stiggins and DuFour 2009).

The accountability push has led to an ever-increasing reliance on standardized tests. Rick Stiggins asks: “Are our current approaches to assessment improving student learning? Might other approaches to assessment have a greater impact?” (Stiggins 2002, 758) Assessments have improved over time, including the essential component of



developing methods to generate reliable and valid test scores, but we are still not providing students with assessments that are truly improving their learning. While the annual tests taken are reliable, they do not provide teachers with all of the information needed to make the instructional decisions within their classrooms. With the amount of focus put upon these standardized tests and the constant concerns of reaching the necessary scores, not enough focus has been put on the classroom test and other assessments (2002).

Stiggins highlights the fact that the United States has spent billions of dollars over the years to ensure that the standardized test scores we view are accurate assessments of student learning. This ranges from the launching of the statewide testing programs in the 1960s, to the national assessments in the 1970s and 1980s, and then to focus on the international testing in the 1990s. All of this has culminated with *NCLB* and the mandate to annually test all students in the U.S. in grades 3 through 8 in reading and mathematics. While all of these assessments do provide useful information on what different students have learned and how schools, districts, states, and nations compare to each other, we have seemingly not invested enough to ensure that our educators are able to effectively use assessments for teaching and learning (Stiggins 2002).

The devotion to high-stakes assessment results and the standards movement stems from a behaviorist philosophy that relies heavily upon results that can be measured quantitatively. Psychometricians and teachers both realize that these results, or measurements, are approximations of student achievement and knowledge. The use of these results should be done with caution. Measurement experts counsel educators to think carefully before using test results to make critical individual decisions such as grade

promotion or instructional diagnoses (Gunzenhauser 2003). The public scrutiny, and in turn political pressure, to see high test scores helps the trend towards teaching ‘to the test’ to prosper, despite the evidence showing that it is not necessarily beneficial to the students’ widest educational interests. As teachers and schools ‘align’ their curriculum to what is on the standardized assessments, students may see an increase in their scores, but the general knowledge and mastery of curriculum domains are not necessarily improved. A study of 18 states with high-stakes testing by Amrein and Berliner (2002) found that the learning levels of students in most of the states were at the same level as before the high-stakes testing took place (Meisels, et al., 2003).

Assessments ‘of’ and ‘for’ learning are both instrumental in helping educators to understand what students are learning. Stiggins provides a series of steps that should be included to do this:

- Understanding and articulating in advance of teaching the achievement targets that their students are to hit;
- Informing their students about those learning goals, in terms that students understand, from the very beginning of the teaching and learning process;
- Becoming assessment literate and thus able to transform their expectations into assessment exercises and scoring procedures that accurately reflect student achievement;
- Using classroom assessment to build students’ confidence in themselves as learners and help them take responsibility for their own learning, so as to lay a foundation for lifelong learning;
- Translating classroom assessment results into frequent descriptive feedback for students, providing them with specific insights as to how to improve;
- Continuously adjusting instruction based on the results of classroom assessments;
- Engaging students in regular self-assessment, with standards held constant so that students can watch themselves grow over time and thus feel in charge of their own success; and
- Actively involving students in communicating with their teacher and their families about their achievement status and improvement (Stiggins 2002, 762).

The proper use of formative assessment, or assessment for learning, supports substantial learning gains for students. However, it is important for educators to realize it is not just the use of formative assessments that leads to increased learning. The key component is for educators to have a deep understanding of the uses, types, and benefits of various forms of formative assessments. Teachers need to receive support in the planning and implementation of this type of assessment, which can be provided through pre-service teacher education programs or professional development opportunities, (Black, Lee, Harrison, and Black, 2004). Unfortunately, the current emphasis on accountability is not compatible with advancing deep knowledge of assessment for teachers (Black, et al, 2004).

To maximize the power and effectiveness of formative assessments, Stiggins and DuFour lay out four essential conditions. First, clear learning targets must be identified. These learning targets are to focus on what the most critical information or skills the students should be able to learn. They should be within the developmental reach of students, manageable in terms of available time and resources, and integrated into the learning progressions within and across grades (2009). Margaret Heritage (2008) also stresses it is important that the information from formative assessments is not treated as a series of 'ad hoc' events, but rather is seen as a systematic process. By developing a system of constant information, one would expect that the instructional modifications made to meet the identified needs would impact student performance. Thus, teachers should be able to see how each student's learning evolves (Heritage, 2008). Assessment development requires clarity of learning goals and objectives that students are to master,

which allows teachers to recognize where changes need to be made (Shavelson, et al., 2008).

Stiggins and DuFour (2009) recommend that an assessment program should be committed to standards-based instruction, and that its results are meaningfully communicated to students and parents. Teacher preparation programs must be prepared to teach young professionals how to work with learning targets and progressions, standards-based instruction, and how to communicate the meaning of the results to parents and students. Student success can be advanced by aligning formative assessments with summative assessments, even standardized assessments, where possible. Teachers will be more familiar with the expectations that they face on the high stakes tests, and thus better able to help their charges to succeed with them (Shavelson, et al, 2008). This is not to imply that teachers should ‘teach to the test.’ It does mean, however, that they should be aware of the critical skills and knowledge covered on the standardized assessments.

### **Assessment Standards**

The question of what teachers should know about educational assessment has been asked for many years. There have been different answers to this question depending upon who is asked, the context of those who are assessing students, and at what stage in one’s experience this is asked. The American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) worked together to develop the *Standards for Educational and Psychological Testing (Standards)*. They have been cooperatively developing versions of the *Standards* since 1955 with new versions in 1966, 1985, 1999,

and one currently being developed (<http://teststandards.org/>). While the standards are not without detractors, they are largely seen by most measurement specialists as a “statement of technical standards for sound professional practice”. Experts also agree that these standards can be used to begin to define the assessment literacy required by teachers prior to their use of educational tests (Frisbie and Friedman 1987). While it is acknowledged that these standards will not be carefully read by all educators, they do provide educators with criteria to help them understand what is expected when one evaluates a test, partakes in various testing practices, or needs to be aware of the effects of test use. The *Standards* take time to promote a rational basis for test use in the classroom setting. At the same time, the *Standards* are not all encompassing in what they cover in relation to educational testing. Key areas, such as, grading, observational techniques, and the writing of different types of test items were not included (Frisbie and Friedman 1987). While the *Standards* do not cover all of the testing issues that arise, they do provide a solid base of understanding and guidance for educators.

The *Standards for Teacher Competence in Educational Assessment of Students* were published cooperatively by the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) (1990). These professional organizations were concerned that the educational benefits that student assessment can provide were not being realized and felt that developing standards for teacher competence in student assessment was an important new step. The document was meant to guide the training of both pre-service and in-service teachers, the accreditation of preparation programs, and the future certification of all educators (*Standards for Teacher Competence in Educational Assessment of Students*

1990). The development of these standards reiterates the claim that properly used student assessments are critical to the teaching and learning process. The intended uses of these standards include:

- guiding teacher educators as they design and approve programs for teacher preparation;
- guiding self-assessment by teachers to identify needs for professional development in student assessment;
- guiding workshop instructors to design professional development experiences for in-service teachers; and,
- an impetus for educational measurement specialists and teacher trainers to conceptualize student assessment and teacher training in student assessment more broadly than has been the case in the past.

The seven standards that have been identified as those that a teacher needs to be competent in the educational assessment of students include:

- choosing assessment methods appropriate for instructional decisions;
- developing assessment methods appropriate for instructional decisions;
- administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods;
- using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement;
- developing valid pupil grading procedures which use pupil assessments;
- communicating assessment results to students, parents, other lay audiences, and other educators; and,
- recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (*Standards for Teacher Competence in Educational Assessment of Students 1990*).

The standards help frame concepts related to both formative and summative assessments and provide an enhanced view of the assessment skills that are typically included in traditional measurement textbooks and courses. They present assessments to teachers as a process involving planning, decision-making, communicating, and creating while stressing the importance of fairness, clarity, and ethics (Whittington 1998).

**Assessment: Are teachers prepared to assess students?**

Despite the existence of assessment standards that describe expectations for teachers' understanding of assessments, not all pre-service teachers receive explicit instruction on assessment in their teacher education programs. Stiggins (1998) identified that fifteen states possess teacher certification standards that require competence in assessment, ten states hold a requirement that assessment course work is taken, and the remaining twenty-five states have no expectation of competence in assessment (Stiggins 1998).

Over fifty years ago Goslin and Mayo (1967) independently found teachers did not exhibit a high-level understanding of different assessment skills. These findings have been substantiated in more recent studies as well. Surveys, tests of assessment knowledge, and reviews of teachers' assessments have all shown that teachers have low levels of assessment knowledge (Gullickson and Hopkins 1987; Brookhart 2001; Campbell, Murphy, and Holt 2002; Stiggins 2001; as cited by Mertler 2003).

Marso and Pigge (1987, as cited by Wise 1993) conducted a study that collected ratings by classroom teachers, building principals, and supervisors of classroom teachers that identified needs for a variety of testing competencies. Teachers reported a high need for competencies involving the instructional use of test results such as grading and scoring activities, re-teaching, and identifying pupil strengths and weaknesses, as well as test validity-related competencies including matching questions with objectives, writing questions that measure higher order thinking skills, making tests that reflect what was taught, and measuring true progress of pupils. They reported little need for measurement statistics or for competency in selecting good test questions from sources such as teacher

manuals. Collectively, these teachers' ratings of needed testing competencies suggest little teacher concern for question structural quality while a direct analysis conducted of these teachers' self-constructed tests revealed frequent violations of common question writing guidelines (Wise 1993).

The results provided by the teachers were similar to those identified by principals and teacher supervisors. Principals and teachers differed in their perceptions of teachers' needs for essay testing, classroom observation, and pupil grading-related competencies. The teachers rated their need for competencies related to classroom observations and pupil grading considerably higher than principals or supervisors did; whereas principals perceived a greater need for teachers' essay testing skills (Wise 1993).

The increased emphasis on achieving higher test scores impacts the role of the day-to-day assessments. According to Campbell, Murphy, and Holt (2002; as cited by Mertler and Campbell 2005) teachers must develop classroom assessments aligned to state standards and there is a positive relationship between the quality of classroom assessments and student performance on standardized assessments. In fact, Stiggins (1999) found improving the quality of classroom assessments can increase average scores on the standardized assessments by as much as  $\frac{3}{4}$  of a deviation (Mertler and Campbell 2005). However, teachers report they do not feel sufficiently prepared to assess student performance and do not believe the assessment training provided in their teacher education programs prepared them to make sound decisions based on the assessments they give. A study by Popham (2003) showed that teachers' sense of discomfort and inadequate preparation are evident in their limited assessment literacy (Mertler 2009).



Assessment can be categorized into two categories- formative or assessments for learning and summative or assessments of learning. Popham aligns assessments as classroom assessments and accountability assessments. This is an important distinction as classroom assessments can be considered those that teachers will use for the purpose of grading and driving their curriculum and teaching (Popham 2009). Accountability assessments tend to be standardized assessments often tied to district, state, or national (*NCLB*) accountability processes. These assessments can be used for purposes such as identifying school- or district-wide areas of strengths or weaknesses, trends across sub-groups, or whether schools, administrators, and teachers are successful or not (Popham 2009). To use any assessment properly, one must be knowledgeable of the assessment types. For assessment of learning, teachers should be skilled at:

- interpreting and using percentile scores and standard deviation values, standard error values;
- understanding theoretical principles governing standardized and classroom-based testing;
- conducting item analysis;
- understanding the different uses of criterion versus norm-referenced assessments;
- understanding the specific uses and limitations of the various forms of summative assessments (DeLuca and Klinger 2010).

For assessment for learning, teachers should be skilled at:

- implementing self- and peer-assessment;
- modeling various learning levels;
- offering continuous descriptive feedback;
- establishing expectations and assessment criteria for the students (DeLuca and Klinger 2010).

Numerous surveys have looked at teacher testing practices. Teachers report a heavy reliance on teacher-made tests during day-to-day instruction; and report little

reliance on standardized test results for making instructional decisions. Salmon-Cox (1981) and Brookhart (2001, as cited by Mertler 2003) found only minor use of the results of standardized tests in classroom instruction. Borg, Worthen, and Valcarce (1986) found unfavorable and indifferent classroom teacher attitudes toward the use of standardized tests compared to a highly positive attitude toward the use of teacher-made tests. Stiggins and Bridgeford (1985) found classroom teachers tended to use teacher-made tests for pupil diagnosis, grouping, grading, evaluation, and reporting pupil progress in their classrooms. These teachers relied more on self-constructed tests than publisher-produced tests, structured performance assessments, or spontaneous observations of pupils in making instructional decisions. Further, Hall, Carroll, and Comer (1988), Lazar-Morrison, Polin, Moy, and Burry (1980), Stiggins and Bridgeford (1985) all found teachers reported utilizing the results from tests created by themselves more than any other resource (Wise 1993).

Studies used measurement theory to judge published classroom tests, portfolios, teachers' knowledge, teachers' reported beliefs and practices, and student teachers' lesson plans. Issues existed across the board (Brookhart, 2004). Frisbie, Miranda, and Baker (1993) found a textbook series set of unit tests averaged about half of identified objectives being assessed, two thirds of the items were phrase matches with the text, and nearly 90% of the items involved simple recall. Impara, Plake, and Fager (1993) completed an assessment of teachers on assessment knowledge and teachers averaged only 23.18 out of 35. Mertler (2001) completed a study that identified the steps teachers take to ensure reliability and validity of classroom assessments and noted steps incomplete or incorrect. Campbell and Evans (2000) found that only 25 out of 237 lesson

plans reviewed contained objectives that even contained partial evidences that the objectives would be assessed reliably and validly (Brookhart 2004).

A study by B. S. Plake (1993) found that teachers reported their perceived lack of preparation was primarily due to inadequate pre-service training they received in educational measurement. In one statewide survey (1999) over 85 percent of the teachers stated that they were not well prepared to assess students (Mertler 2009). Studies by Mertler (1999), Plake (1993), and Campbell, Murphy, and Holt (2002) focused on the assessment knowledge of pre- and in-service teachers and constructed around the *Standards*. They consistently found participating teachers were getting less than 70 percent of the questions correct. A study by Campbell and Mertler (2005) showed pre-service teachers who had recently completed a course in applied classroom assessment correctly answered only 68 percent of the items (Mertler 2009). Teachers perform poorly on surveys focused on assessment knowledge and also hold the perception that they lack critical assessment skills. Stiggins (1999) states there are a degree of assessment illiteracy among teachers that has led to inaccurate assessment of students, negatively impacting students' education (Mertler and Campbell 2005).

The lack of assessment knowledge teachers acknowledge and exhibit raises concerns on how these teachers are prepared to assess students. Only about one half of pre-service students report participating in a course focused on assessment, while the other half receives assessment knowledge integrated within another coursework. Compounding this issue is that when pre-service teachers take an assessment course, it tends to occur prior to student teaching, thus leaving students without the ability to apply the knowledge they learn in actual classroom context (Gullickson and Hopkins 1987).

Additionally, the issue has been raised that the content of these courses does not match the skill sets teachers identify as what they actually need and use in the classroom. There is often an overly heavy focus placed upon statistical analysis and standardized tests (1987). Gullickson's work has been replicated by others, including by Plake (1993) and Brookhart (2001). Brookhart cited literature that noted too much time was given to large-scale or standardized assessments and not enough focus was placed on classroom assessments (Mertler 2003). Studies on how teachers ensure validity or reliability found that in many cases teachers did not follow any steps. At other times they took efforts that were not appropriate for ensuring validity or reliability, including half of teachers in one study saying that a test was automatically reliable if it was teacher-created (Mertler 2003). Regardless of the method chosen to look at teachers' assessment practices and their levels of preparation, the findings consistently came back as the teachers possessing generally weak assessment knowledge.

Additional evidence that teachers do not feel that their pre-service education provided enough support in the area of assessment comes from a study by Mertler (1999) that found teachers felt only 'slightly prepared' to assess student learning prior to beginning their teaching career. On a follow-up survey with the same teachers after they began their careers, now felt 'somewhat prepared.' Though not an ideal response, it indicates teachers acquire more knowledge once they are actually on the job than during their teacher education program (Mertler 2003). A study by Volante and Fazio (2007) looked at the assessment literacy development of pre-service teachers in an education program for teachers in grades 1-6. These pre-service teachers were given a survey each of their four years in the program and consistently reported low levels of confidence each

year. The teacher candidates all expressed a desire to have more practical knowledge about assessment and would have been very supportive of a specific course focused on classroom assessment. This case found that the most relevant assessment education for these teacher candidates occurred through their guidance from the mentorship of a paired associate teacher rather than in their coursework (DeLuca and Klinger 2010).

Teachers are also expected to assess their students' progress through the use of a variety of assessment formats. Assessment literacy, or the understanding and appropriate use of assessment practices, is important in ensuring that teachers are able to validly use the assessment results they obtain from the students. Teachers with this knowledge will be able to use assessment results to make accurate inferences about student learning, communicate this information, and use it to guide their instruction. At the same time, teachers not skilled are more likely to make invalid uses of assessment results and ultimately make poor and inappropriate instructional decisions (DeLuca and Klinger 2010).

Assessments guide learning. DeLuca and Klinger (2010) advocate that teachers learn about assessment practices at the same time they learn their pedagogy, curriculum, and cognitive development. Brookhart (2003) stated that teacher candidates should have opportunities to learn various measurement principles collectively in order to expose them to other complexities. Not only that, but teachers should have a sense of the different theoretical principles that support various assessment practices. DeLuca and Klinger (2010) state that assessment education should provide support in a manner in which assessment education integrates practice, theory, and philosophy across both assessments for and of learning.

Judith Hollenweger (2011) found that teachers who were trained in using these assessment procedures still must have training in how to apply the information taken from these assessments into classroom practice. The informal decisions that are made from the classroom assessments typically work well for ranking students. However, the problems arise with overestimation of achievement occurs for low achieving students. They overestimate students' levels of achievement and underestimate the standard deviation of achievement (Hollenweger 2011). Additionally, secondary teachers were more likely to use teacher-developed assessments than elementary counterparts, while the elementary teachers were more likely to use performance assessment (Brookhart, 2004).

Teachers tend not to rely completely on standardized assessment scores to determine the success of their students, and in fact typically turn to the results of teacher-created assessments. This is where we can see intersections of the benefits and issues with the *NCLB* legislation. On the positive side, teachers are beginning to look closer at the data related to student scores on the standardized tests. While this is not the only piece of information, it is data that exists and helps inform teachers of student achievement. On the other hand, this would raise the question that if teachers place a heavy reliance upon classroom assessments, there should be more focus placed here- both in helping the teachers to understand how to better understand how to create and use classroom assessments and associated results.

No matter how skilled or knowledgeable teachers are with assessment practices, they will use them as key teaching instruments. Assessments are ingrained into the educational process – whether formally or informally. Many teachers know little about assessment as studies have shown. The onus of this issue is shared between teachers,

policy makers, teacher education programs, and assessment specialists. Work is being done to bring this issue more to the forefront and the call to hear more about the need for teachers to know more about assessments and how they can and should be used (Popham 2009).

## **CHAPTER III**

### **METHODOLOGY**

#### **Introduction**

This study used parallel self-report instruments that yielded data on the preparation level of pre-service teachers in the area of assessment knowledge. Pre-service elementary teachers at a Midwestern university were asked to report on their exposure and understanding rates of assessment knowledge. Elementary school principals also provided feedback by noting what they wanted beginning teachers to know about assessment and what they feel new teachers actually know. The development of two self-report instruments provided a basis to compare and contrast how a teacher education program prepared elementary pre-service teachers to use assessments against the desired levels of understanding that principals expected from young teachers.

The surveys were developed with a ‘forced choice’ format so that the results could be quantified. The responses to these surveys were organized and coded on EXCEL and then entered into the ‘Statistical Package for the Social Sciences (SPSS)’. This generated descriptive statistical data.

#### **Research Questions**

- I. To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skills, including:
  - a. exposure to creating and using different assessment methods;
  - b. exposure to understanding the strengths and limitations of different assessment types;
  - c. exposure to learning how to construct or evaluate tests and test items;
  - d. exposure to the interpreting different types of standardized test scores; and,
  - e. exposure to key assessment terms.



- II. Among elementary school principals, what assessment principles and practices are most important to know among beginning teachers working in classrooms?
- III. How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure?
- IV. What are the self-reported levels of understanding among pre-service teachers of various assessment concepts and skills?
- V. How do elementary school principals conceive of assessment understanding for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers?
- VI. What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?
- VII. What are the attitudes of elementary school principals toward standardized achievement results in the classroom? How congruent are these attitudes with the attitudes of pre-service teachers?

### **Instrument Development**

The study used two surveys developed specifically to collect the data that address the research questions in this study. The surveys both contained statements in which the participants had to indicate agreement to some degree, on a Likert Scale. The statements to which the respondents had to reply were designed specifically for the different population of participants – pre-service elementary teacher candidates and elementary school principals. By comparing the results of these two populations, I was looking to compare the knowledge and skills that elementary school principals hoped to see among beginning teachers in the area of assessment against what pre-service teachers reported being exposed to and what they reported to have learned in their specific teacher education program. Additionally, the survey addressed the perceptions that both populations had towards the usefulness of standardized achievement assessments in elementary schools.

The self-report items were developed to address three key components:

- I. exposure to assessment concepts and skills;
- II. reported assessment understanding; and,
- III. perceptions of the usefulness of standardized achievement assessments in elementary-based educational settings.

The training teachers receive during their pre-service education is a crucial source for the introduction of beginning teaching knowledge. Assessment is a critical component of teaching, comprising between 30 – 50% of all instructional time (Schafer (William) 1993). In this way, assessment knowledge is critical to the early preparation of teachers. Research shows that the assessment skills of teachers are not strong and that pre-service teachers report their lack of confidence in assessing students (Mertler 2003). Teacher education programs have an opportunity to rectify the problem by exposing pre-service teachers to these skills. With the exception of student teaching and a short practicum experience, the only opportunity most pre-service teachers have for this exposure is through the coursework within their teacher education program.

The *Standards for Teacher Competence in Educational Assessment of Students (Standards)* were referenced to guide the development of the surveys. The *Standards* consist of seven key statements that identify critical skills to be held:

1. Choosing assessment methods appropriate for instructional decisions;
2. Developing assessment methods appropriate for instructional decisions;
3. Administering, scoring and interpreting the result of both externally-produced and teacher-produced assessment methods;
4. Using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement;
5. Developing valid pupil grading procedures which use pupil assessments;
6. Communicating assessment results to students, parents, other lay audiences, and other educators;
7. Recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (*Standards for Teacher Competence in Educational Assessment of Students* 1990).

All of the research questions contain items that speak directly to the statements contained in the *Standards*. Questions I, II, and III consider the amount of exposure pre-service teachers reported on the different skills and concepts, as well as the importance levels principals place on the different topics. Questions IV and V consider the knowledge and skills pre-service teachers report and principals perceive beginning teachers to have when they enter the profession. Questions VI and VII address the attitudes that both pre-service teachers and principals place on the role of standardized achievement in elementary school curriculum.

To cover Component I, exposure rates to assessment concepts and skills, research questions I, II, and III all spoke to the relation between the exposure rates that pre-service teachers reported in their teacher education program and the desired knowledge principals expected beginning teachers to possess. The instrument measured perceptions connected to exposure to:

- various assessment methods,
- construction and evaluation of assessment items,
- evaluation of published assessments,
- understanding of strengths and limitations of various assessment types,
- interpretation and communication of different assessment results,
- knowledge of key assessment terminology, and;
- exposure to a variety of assessment formats and using the results to gauge student progress.

Component II, attained assessment knowledge, provides information on the perceived understanding of key assessment concepts and terminology, including:

- knowledge of various assessment methods and procedures,
- the ability to write sound items,
- interpreting assessment results,
- grading,
- interpreting results for parents and students, and;

- for the use of the results to guide instruction

Question IV provided the pre-service teachers perceptions of attained knowledge.

Question V provided an examination of the congruence that existed between the pre-service teachers' perceptions of achieved knowledge and the principals' perceptions of beginning teachers' knowledge.

Component III, perceptions of the importance of standardized achievement results in an elementary educational setting, provided an examination of prevailing attitudes toward assessment, including how well it:

- serves useful purposes,
- reflects students' educational achievement,
- provides evidence of teachers' effectiveness,
- provides information to monitor yearly academic progress, and;
- helps identify students' academic strengths and weaknesses.

Question VI addresses the attitudes that pre-service teachers possess about this topic, while research Question VII addresses the attitudes of principals and how those compare with the pre-service teachers' responses.

Table 3.1. Research Questions – Component alignment

<b>Research Question</b>	<b>Research Components</b>	<b>Standards</b>	<b>Item Numbers</b>
<b>Question I:</b> To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skills?	I	1, 2, 3, 4, 6	1.1 – 1.25
<b>Question II:</b> Among elementary school principals, what assessment principles and practices are most important to know among beginning teachers working in classrooms?	I	1, 2, 3, 4, 6	1.1 – 1.25
<b>Question III:</b> How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure?	I	1, 2, 3, 4, 6	1.1 – 1.25
<b>Question IV:</b> What are the self-reported levels of understanding among pre-service teachers of various assessment concepts and skills?	II	1, 2, 3, 4, 5, 6	3.1 – 3.7
<b>Question V:</b> How do elementary school principals conceive of assessment understanding for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers?	II	1, 2, 3, 4, 5, 6	3.1 – 3.7
<b>Question IV:</b> What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?	III	3, 4, 6, 7	2.1 – 2.5
<b>Question VII:</b> What are the attitudes of elementary school principals toward standardized achievement results in the classroom? How congruent are these attitudes with the attitudes of pre-service teachers?	III	3, 4, 6, 7	2.1 – 2.5

### **Selection of the Participants**

The selection of the participants for this study was done to try to capture the pre-service teachers that principals might potentially hire. The focus on a specific teacher education program allowed the study to identify associations between pre-service teachers' perceptions of the assessment knowledge learned in their teacher education

program and what school principals potentially felt was needed in the areas of assessment among the teachers they would hire. The focus on one program also allows for a closer look at how one particular population has prepared its students.

The pre-service teachers who participated in this study attended an elementary education teacher preparation program and were in the process of completing their student teaching experience. This population would have been exposed to all of the courses required in the program at the time of the study. Additionally, a vast majority of participants were assumed to be applying for teaching positions immediately following the student teaching experience, and would be entering their career with the knowledge they possessed at the time of the survey.

Elementary education majors were chosen as the focus for this study because they all took a core set of classes in common. While there are specializations for elementary education majors, they nevertheless have a common set of courses they take - unlike high school teachers where majors are drawn from various disciplines. Additionally, the pre-service teachers likely applied for elementary teaching positions and were potentially interviewed or hired by elementary school principals.

Table 3.2. Distribution of the Pre-service Teacher Sample Population

	<b>Number of pre-service teachers responding</b>	<b>Total population of pre-service teachers contacted</b>	<b>Response Rate</b>
Overall	55	150	37%
Spring 2008	33	87	38%
Fall 2008	23	63	37%

Table 3.3. Principal Sample Population

	<b>Number of principals responding</b>	<b>Total population of principals contacted</b>	<b>% responding</b>
Overall	62	1,039	6%

The final analysis included fifty-five elementary pre-service teachers from one elementary education program. All elementary pre-service teachers who were in the process of student teaching in this program were invited to participate in the study during each of two different semesters. The survey responses for the pre-service teachers came during the spring semester of 2008 and the fall semester of 2008. Thirty-three out of eighty-seven pre-teachers (38%) responded during the spring semester, while twenty-three of sixty-three (37%) responded in the fall semester. Of the thirty-three students responding during the spring semester, one of them was dropped from the analysis because that individual did not respond to all of the prompts. A total of fifty-five out of one hundred fifty pre-service teachers that were available for this study participated, providing a 37% response rate.

Responses from pre-service teachers were taken from two semesters to provide a larger sample size. Of the sample population, twenty-two (40%) entered the teacher education program in 2006 or earlier and thirty-three (60%) after 2006. The self-identified content specialization of the teachers was another variable identified. Of the respondents, twenty-six (47%) of the pre-service teachers in the sample self-identified as literacy specialists (reading and/ or language arts), nine (16%) as math and science specialists, six (11%) as social science specialists, and the remaining fourteen (25%)

identified their specialty as art, special education, early childhood, or a combination of specialties.

Elementary school principals from across the state of Iowa and the Chicago Public Schools (CPS) were invited to participate in this study. Sixty-three principals submitted the survey, although one of these was excluded because it was not complete. The responses from sixty-two principals were analyzed. The survey was sent electronically (via Websurveyor) to one thousand thirty-nine principals. The response rate was only 6%. The lists of schools and the respective principals were obtained from the Iowa and Illinois State Board of Education sites and included districts across the state of Iowa and elementary schools within the Chicago Public Schools district. The principal survey was completed during the spring of 2009. This would have been at the time when the pre-service teachers were completing their first year of actual teaching, although it is not known if any of those teachers interviewed at or received a job in one of these schools.

### **Survey Development and Collection**

The surveys used in this study were developed specifically for this research project. The purpose of the items within the survey were developed in order to provide a sense of the assessment concepts and skills pre-service teachers are exposed to and understand, and to compare them with the expectations and perceptions held by principals on the same topics. The responses to these items were intended to provide insight into whether the preparation of pre-service teachers was adequately preparing young teachers for what they need in the area of assessment once they enter the classroom. As mentioned earlier, the items were developed to address three core



components - exposure to assessment concepts and skills; attained assessment understanding; and perceptions of the usefulness of standardized achievement assessments in elementary-based educational settings.

The items were developed and in turn reviewed by curriculum expert Professor Hlebowitsh (University of Iowa – Teaching and Learning Professor and Departmental Executive Officer) and assessment expert Professor Frisbie (former professor at the University of Iowa and Director of Iowa Testing Programs). Items were reviewed and revised in order to ensure that the subject matter was covered in depth. A test survey was sent out in November of 2007 to pre-service teachers. Based on a low-response rate and advice from experts, the survey was revised to ensure that the directions and desired response types were clear to potential respondents. One example was to ensure that the scale used had a common factor of the highest number representing the most agreement and the lowest number the least agreement.

The surveys that were used to collect the data for this research project were developed using a Likert Scale format. The surveys were built into Websurveyor and sent to the identified audience in an electronic format from my university account. The survey was sent to two types of respondent: elementary pre-service teachers and elementary principals. The respective survey sent to principals and pre-service teachers was written for the specific population (see Appendix C - E and Appendix F – H respectively). For example, with the first section of the survey (instructional exposure to assessment knowledge), the directions were written in the following manner for each population.

- **Pre-service teachers:** Please indicate your opinion of how much you have been exposed to the following assessment topics during your pre-service education courses (5 being the MOST exposure, 1 the LEAST).
  - 1.1 – Learning how to use different assessment methods/ approaches for a variety of purposes.
- **Elementary Principals:** How important is it for the beginning teachers to understand the following:
  - 1.1 - To understand how to use different assessment methods/ approaches for a variety of purposes.

The portion of the survey that was used to elicit the self-report was in three sections aligned to the three components (for both pre-service teachers and principals). The sections included:

- 1) instructional exposure to assessment knowledge (twenty-five items),
- 2) perceptions of standardized achievement assessments in elementary schools (five items), and;
- 3) perceptions of assessment knowledge (seven items).

Elementary school pre-service teachers that were in the process of student teaching during the spring and fall semesters of 2008 were selected to participate in the study. The pre-service teachers were all sent a link to the Websurveyor site along with an invitation to participate in the study. An email reminder was sent to the pre-service teachers two weeks after the initial request.

Principals from the state of Iowa and the Chicago Public Schools were sent email requests to participate in the study during the spring of 2009. The principals invited to participate in the study were identified from the Iowa and Illinois state education sites. The principals were sent the Websurveyor link embedded in an email along with the invitation to participate.

### **Data Analysis**

The surveys were designed to highlight any differences or similarities that exist between the perceptions of what assessment knowledge or skills school administrators would like beginning teachers to enter their career with and what these teachers are being taught or exposed to in their teacher education program.

The survey responses were collected directly through the Websurveyor device. The responses were put into EXCEL to be organized and coded. The ‘Statistical Package for the Social Sciences’ (SPSS), was used to manage and analyze the collected data. SPSS allowed for the generation of associated frequencies, percentages, and means of the acquired data. The data were organized and analyzed around the three components identified earlier including, exposure to assessment concepts and skills, attained assessment understanding, and perceptions of the place standardized achievement assessments in elementary-based educational settings. Analyses used for each of the three components are identified in Table 3.5.

Table 3.4. Component Analysis

Component	Item Numbers	Analysis Completed
<p><b>Component I:</b> Exposure to assessment concepts and skills (Each question was further broken into five categories) <i>Questions I, II, and III</i></p>	1.1 – 1.25	<p><i>Questions I and II:</i></p> <ul style="list-style-type: none"> <li>• Comparison of overall category means</li> <li>• Comparison of item means</li> <li>• Comparison of response rates within each item and across items</li> </ul> <p><i>Question III:</i></p> <ul style="list-style-type: none"> <li>• Pre-service teacher – Principal comparison of means on each category</li> <li>• Pre-service teacher – Principal comparison of response rates within each item and across items</li> </ul>
<p><b>Component II:</b> Reported assessment understanding <i>Questions IV and V</i></p>	3.1 – 3.7	<p><i>Question IV:</i></p> <ul style="list-style-type: none"> <li>• Comparison of item means</li> <li>• Comparison of response rates within each item</li> </ul> <p><i>Question V:</i></p> <ul style="list-style-type: none"> <li>• Comparison of item means</li> <li>• Pre-service teacher – Principal comparison of item means</li> <li>• Pre-service teacher – Principal comparison of response rates within and between items</li> </ul>
<p><b>Component III:</b> Perceptions of the usefulness of standardized achievement assessments in elementary-based educational settings <i>Questions VI and VII</i></p>	2.1 – 2.5	<p><i>Question VI:</i></p> <ul style="list-style-type: none"> <li>• Comparison of item means</li> </ul> <p><i>Question VII:</i></p> <ul style="list-style-type: none"> <li>• Pre-service teacher – Principal comparison of item means</li> <li>• Pre-service teacher – Principal comparison of response rates within and between items</li> </ul>

### Limitations of the Study

The method for selecting the sample did not likely produce a representative sample. Limiting the sample to pre-service teachers from one university did not allow for a true sample of what all pre-service teachers are exposed to - other than at that one location. Therefore the results from this study are not generalizable across schools of education or other states. Significant differences, for instance, exist between teacher

education programs that carry a specific required course on assessment and programs that carry a less visible commitment to assessment. Only elementary principals who worked either in the state of Iowa or the Chicago Public Schools (CPS) were contacted. This also provided some limitation. All principals on the state of Iowa's education website were emailed. Approximately one-half of the elementary principals from CPS were contacted. Those were selected by being in the first half of the contacts listed on the Illinois State Board of Education site.

Elementary school educators were the focus of this study. This leaves out a significant portion of the K-12 education community, including high school teachers and administrators. Assumptions cannot be made that the responses of elementary pre-service teachers and principals would correlate with high school pre-service teachers and principals. Significant differences exist in the practices and learning that occurs within the two environments, and therefore the results of this study are not generalizable to that population.

Comparisons were made between the responses of the pre-service teachers and the elementary principals. While principals from Iowa and CPS were chosen because there is a chance that the graduates from the university would seek employment at one of the schools represented, it is not known if these pre-service teachers either applied or were hired by any of the participating principals.

The method of allowing both pre-service teachers and principals to self-select their responses introduces potential respondent bias. Self-reported data are not supported by any observational records or tested against some outcome of reality. Additionally, the subject of standardized achievement assessments and the use of the related results within

public schools is an issue that many educators have strong biases towards. The use (and at times misuse) of assessment results has certainly generated strong feelings among the principals and prospective teachers, especially in the context of *NCLB* routines.

Additionally, the pre-service teachers potentially have only been exposed to this topic through the opinions of others – including the public, their instructors, cooperating teachers, or other educators they may know.

The use of a survey may also present certain limitations. Respondents may have felt constrained to choose between a closed set of possible responses, without having the opportunity to either clarify or explain the reasoning behind their choices. While efforts were made to provide respondents with choice, there was not any option for more qualitative responses. Additionally, there is a possibility, especially with the pre-service teachers, that there are items containing assessment topics that they do not truly understand.

The survey was sent electronically to targeted respondents. The use of the electronic survey allows for ease of collection, flexibility in response time for respondents, and a common format. However, this method of collection also could introduce problems. The response rate may not have been as high as it might have been if the survey were given in person.

There may also be limitations to the study due to the timing of the surveys. The surveys were distributed and completed in 2008 and 2009. Since that time, public school reforms have shifted in focus. No Child Left Behind, while still playing a role in the educational landscape, does not look the same as it did only a couple of years ago. Today the adoption and impending implementation of the Common Core State Standards

(CCSS) along with its associated assessments offers a new landscape for classroom assessment. Moreover, Race to the Top legislation has brought forward a great number of teacher performance mechanisms across many states – a topic of considerable concern and weight to teachers and principals.

## CHAPTER IV

### RESULTS

The elementary pre-service teachers who participated in this study were student teaching at the time the surveys were given. The populations included participants that were student teaching in elementary schools during the spring semester of 2008 and the fall semester of 2008 at a Midwestern university. The principals that participated in the study were working in schools across a Midwestern state and a large Midwestern urban public school system.

Each research question was answered and addressed individually and in the order in which they appear in Chapter I. The items addressing Questions I, II, and III were each answered on a 5-point Likert scale, while Questions IV, V, VI, and VII were addressed with a 4-point Likert scale. The items were developed around three strands:

1. Exposure to assessment concepts and skills – (Questions I, II, and III);
2. Reported assessment understanding – (Questions I and V); and,
3. Perceptions of the place of standardized achievement assessments in elementary-based educational settings – (Questions VI and VII).

The research questions that served as the basis of this study include:

- I. To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skills, including:
  - a. exposure to creating and using different assessment methods;
  - b. exposure to understanding the strengths and limitations of different assessment types;
  - c. exposure to learning how to construct or evaluate tests and test items;
  - d. exposure to the interpreting different types of standardized test scores; and,
  - e. exposure to key assessment terms.
- II. Among elementary school principals, what assessment principles and practices are most important to know among beginning teachers working in classrooms?



- III. How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure?
- IV. What are the self-reported levels of understanding among pre-service teachers of various assessment concepts and skills?
- V. How do elementary school principals conceive of assessment knowledge for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers?
- VI. What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?
- VII. What are the attitudes of elementary school principals toward standardized achievement results in the classroom? How congruent are these attitudes with the attitudes of pre-service teachers?

**Question I - To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skills?**

The twenty-five items for Question I (Items 1.1 – 1.25) addressed a variety of assessment topics and were covered through five categories:

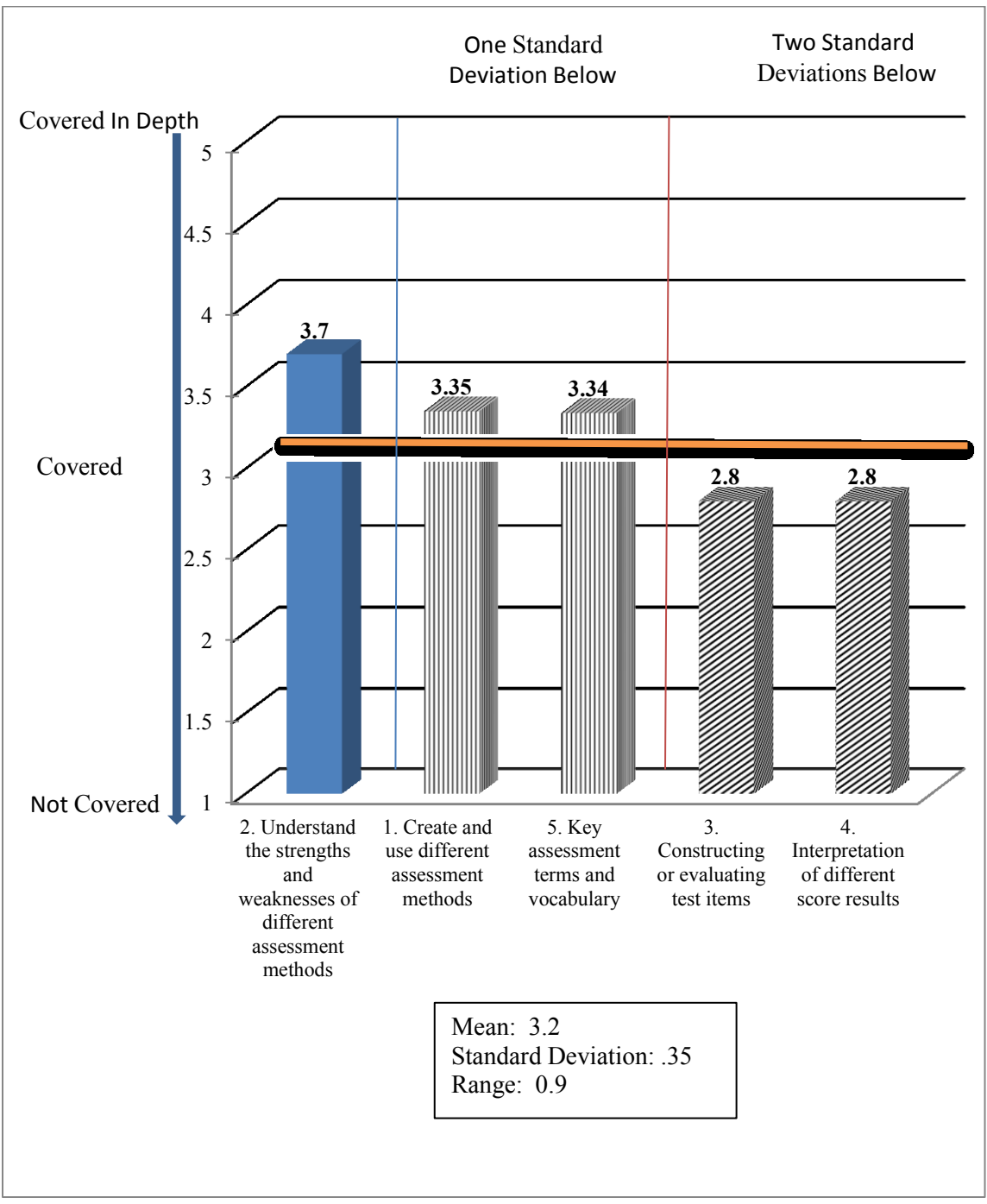
- (1) Creating and using different assessment methods,
- (2) Understanding the strengths and weaknesses of different forms of assessment,
- (3) Constructing or evaluating test items,
- (4) Interpretation of different score results, and
- (5) Key assessment terms and vocabulary.

The pre-service teachers indicated that, when considered collectively, they were exposed to a variety of assessment topics through their teacher education program. The average exposure for the items reported by self-service teachers for Question I was 3.2 and had a standard deviation of 0.35. They indicated that the topic, “understanding the strengths and limitations of different assessment types,” was covered the most within the coursework of the teacher education program with an exposure rating of 3.7. The categories covering the topics “creating and using different assessment methods,” and “key assessment terms and vocabulary,” had respective exposure ratings of 3.35 and 3.34 and fell one standard deviation below Category #1. “Constructing or evaluating tests and

test items,” and “Interpreting different score results,” had reported exposure levels that fell two standard deviations lower than Category #1 and a mean of 2.8. Figure 4.1 provides a visual of the comparisons of means and standard deviations for the five categories.

The pre-service teachers’ responses indicate that they receive exposure to a variety of assessment topics through their pre-service education. Considering that a three indicates that a topic was covered, three of the five categories were covered and the other two fell right below that rating. However, the ratings never approach a central tendency number representative of “covering the material in depth.” Additionally, according to these ratings, the exposure that was provided focused more on general assessment topics including creating and using different methods, looking at the strengths and weaknesses associated with different methods, and assessment-related terminology and vocabulary. Less focus, as indicated by the results being two standard deviations below the highest rated category, was given to the actual construction and evaluation of test items, as well as how to interpret the results that are gathered from the assessments given.

Figure 4.1. Question I - To what degree are pre-service teachers in elementary education exposed to various assessment concepts and skill?



The overall mean for exposure of Category #1, “the creation and use of different assessment methods”, was 3.35 with a standard deviation of .39. This indicates that the

category as a whole is covered (three representing 'covered') although there were individual items within the category that were significantly below the mean. Figure 4.2 provides an overview of the responses given to each item in within the category.

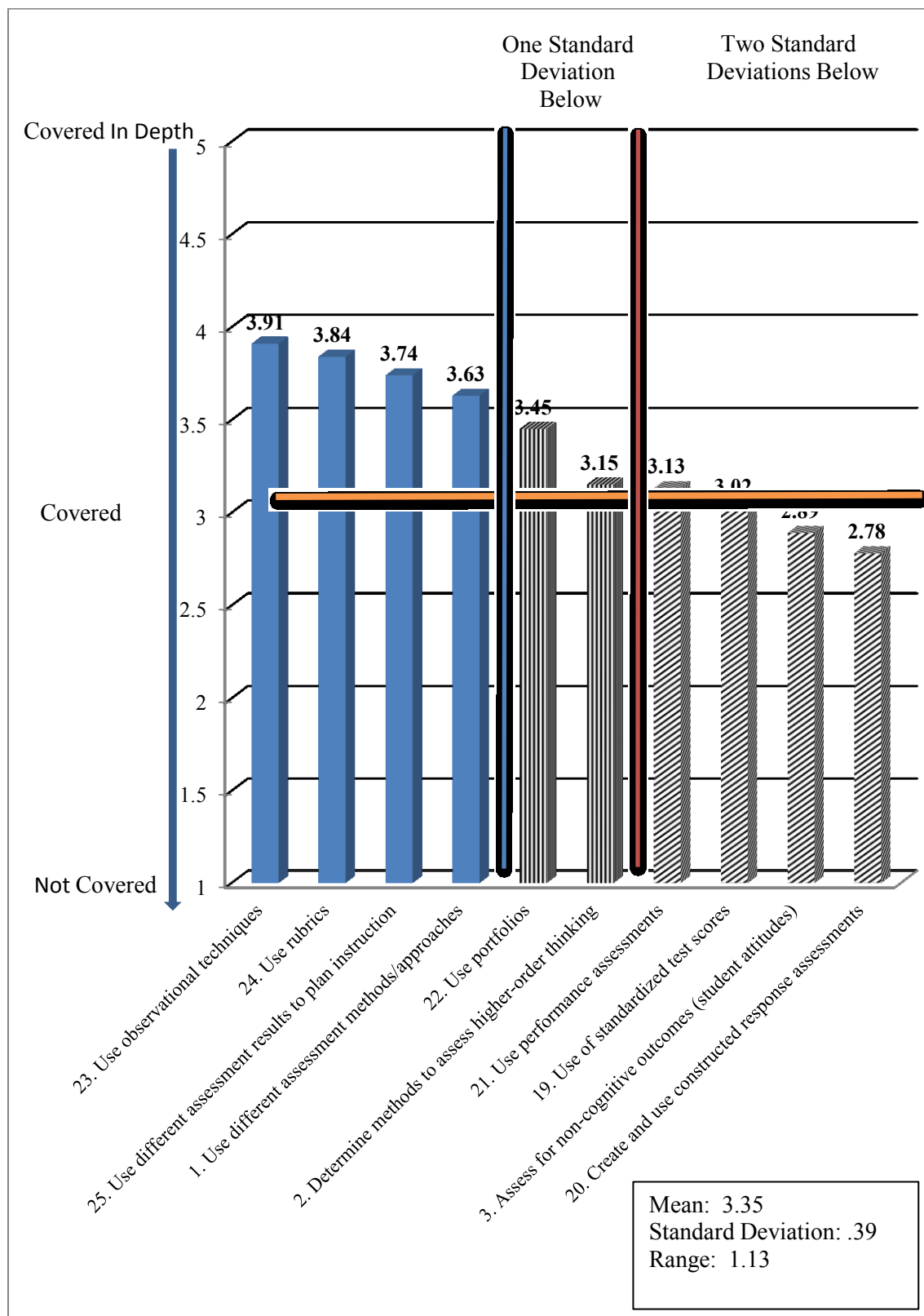
The pre-service teachers reported the highest levels of exposure in the use of item 1.23, "observational techniques," (3.91) and item 1.24, "the use of rubrics," (3.84). Meanwhile, the lowest levels of exposure, all at least two standard deviations away from item 1.23, were the topics "creating and using constructed responses," (2.78), "testing for non-cognitive outcomes," (2.89), "using standardized assessments to monitor student progress," (3.02), and "using performance assessments," (3.13). The two topics referring to the exposure of more general topics – "using different assessment methods/ approaches," and "using different assessment results to plan instruction," - received higher levels of exposure than all of the individual topics (with the exception of the two highest rated topics, using portfolios and observational techniques). The four highest rated items all fell within one standard deviation of each other.

Table II (APPENDIX I) provides a breakdown for the percentage of each possible response on the items in Category #1. Looking at the actual response rates for each item shows that the response, 'covered,' stayed fairly consistent across the items. However, the differences in the responses, 'covered in depth,' and 'not covered' ('5' and '1' respectively) provided substantial difference starting with "determining methods appropriate to assess higher order thinking skills," and "using portfolios," even though both were within one standard deviation of each other. The means are 3.15 and 3.45 respectively, but a look at the actual response rates makes the difference more apparent. For determining methods appropriate for assessing higher order thinking skills, 15% of

the respondents provided a rating of '5' compared with 9% indicating the same level of exposure for exposure to the use of rubrics. Meanwhile, 26% of respondents provided a rating of '2' for coverage on determining methods appropriate for assessing higher order thinking skills, compared to only 13% for using rubrics. This trend continues as the lower-rated items in the category had pre-service teachers shifting '4' and '5' ratings to more '1' and '2' ratings.

The topics two standard deviations away from the highest rated item, including using performance assessments, using standardized test scores, assessing for non-cognitive outcomes, and creating and using constructed response tests all had a minimum of twenty-four percent of respondents reporting little to no coverage compared with using observational techniques where only 6% of respondents reported little to no coverage. With the onset of the adoption and implementation of the Common Core State Standards (CCSS) and the associated assessments being developed, some of the lower rated items represent a concern. The CCSS will likely put a larger focus on the use of performance assessments in the future.

Figure 4.2. Question I - Category #1 – Exposure to the creation and use of different assessment methods



Category #2, “understanding the strengths and limitations of different assessment types,” had the highest exposure rating of the five categories in Question I with an average response rate of 3.7 and a standard deviation of .38. Figure 4.3 and Table J1 (APPENDIX J) provide a closer look at the response rates within the category. The two topics with the highest exposure levels within the category were item 1.7, “understanding the difference in uses of formative and summative assessments,” (4.07) and item 1.10, “understanding the strengths and limitations of results from standardized tests” (3.98) and were within one standard deviation of each other. The topic “exposure to the strengths and limitations of results from teacher-made tests,” (3.11) was one standard deviation from item 1.7, and the topic, “understanding the strengths and limitations from performance assessments,” (3.64) the lowest of the category, was two standard deviations away.

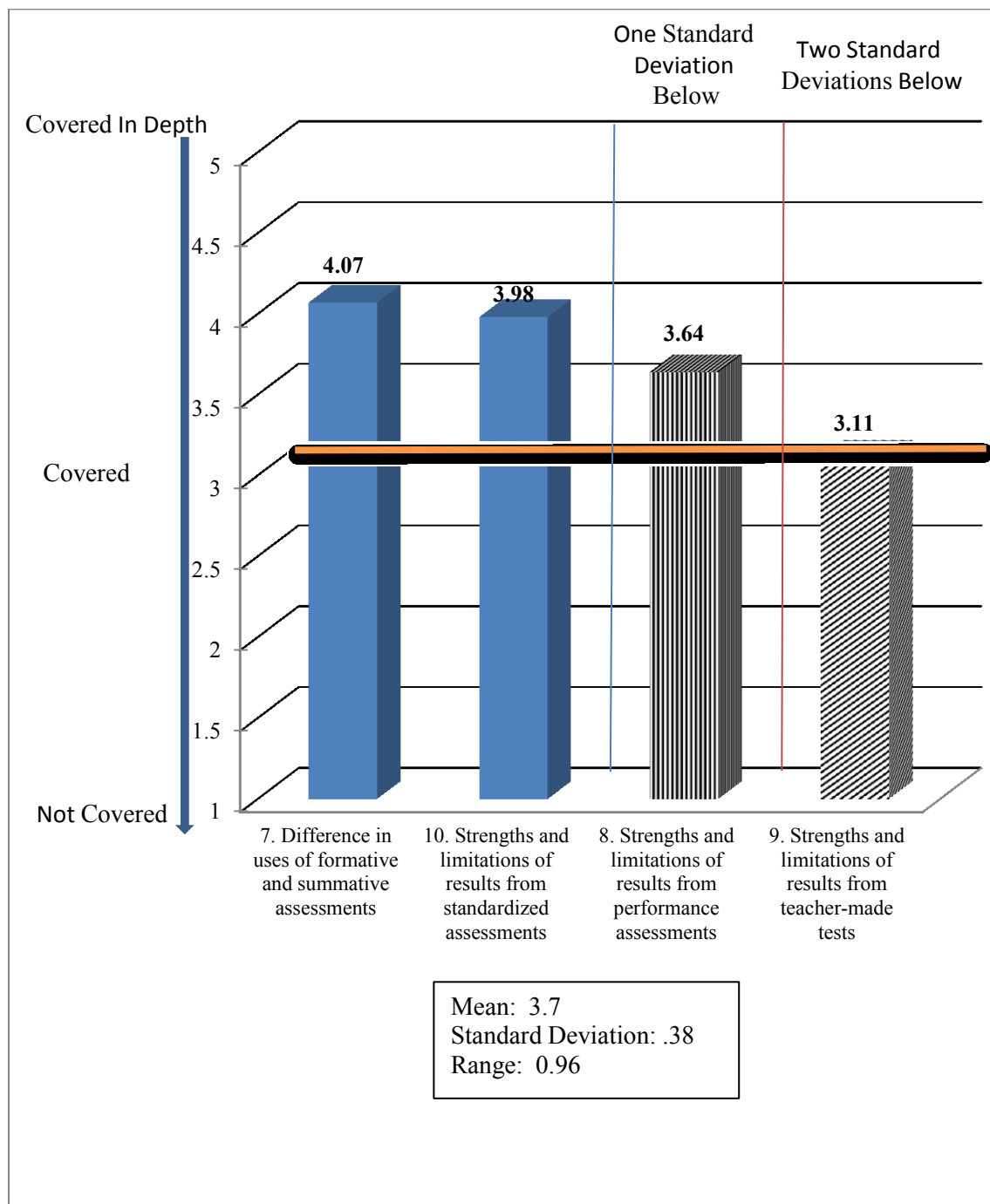
Again, as with Category #1, there was a difference of two standard deviations and close to one point differential in means between the highest and lowest rated items, indicating different assessment topics were covered more than others. The high-level of coverage reported for the topic, “the uses of formative and summative assessments,” is important as teachers need to understand the differences between the purposes of each type, as well as to be able to use them properly within their instruction to ensure students are learning. As seen in Table A2, 94% of the respondents reported that both this topic and item 1.10 were covered to covered-in-depth during their teacher training. The relatively high response rate for “understanding the strengths and limitations of standardized test results,” contradicts the Category #1 response, where they indicated that only 67% of the respondents had received some form of coverage on the use of

standardized test scores. This puts into question whether standardized tests and their accompanying results are covered within the coursework. It is important that teachers not only understand that standardized assessments are going to play a role in their teaching experience, but how they can and should be able to use the results to guide their instruction. The topic “understanding the strengths and limitations of results from teacher-made tests,” provided a rating one standard deviation below understanding the difference between formative and summative assessments. 1/3 of the respondents reported little to no exposure on the topic, which is a concern due to the fact that teacher-made tests or assessments will play a prominent role in their teaching experience.

The responses provided within Category #2 had some contradictions. The pre-service teachers reported high levels of exposure to the differences of formative and summative assessments and the strengths and limitations of results from standardized tests but they simultaneously reported lower levels of exposure to the strengths and limitations aligned to results from teacher-made tests. Teacher-made tests include formative and summative assessments, as well as performance assessments. This leads to questions of why this particular topic does not receive more coverage.



Figure 4.3. Question I - Category #2 – Understanding the strengths and limitations of different assessment types

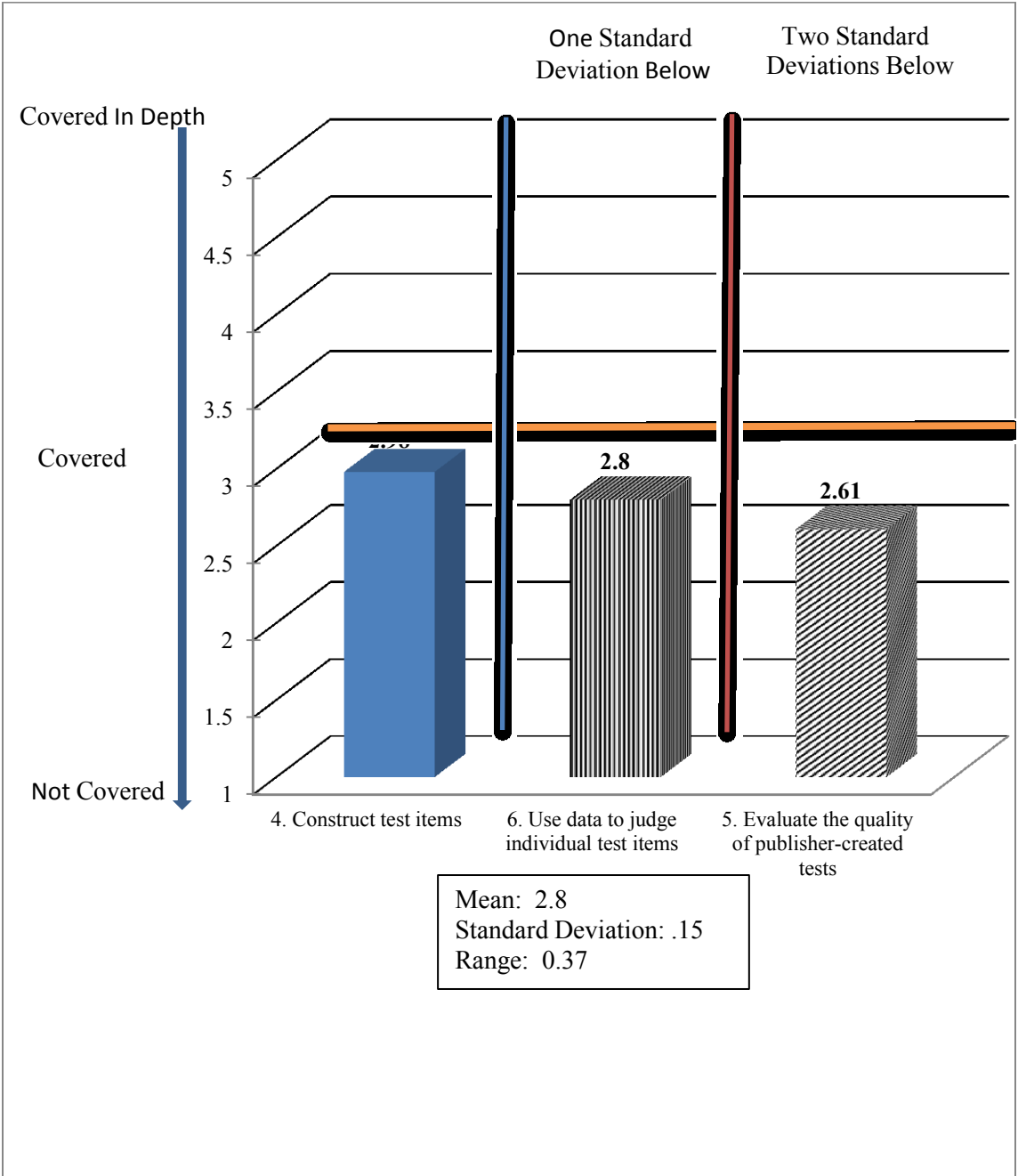


Category #3 focused on pre-service teachers' levels of exposure on learning how to construct or evaluate tests and test items. With an overall mean of 2.8 and a standard

deviation of .15, this category received the lowest ratings of the five categories addressed in Question I (along with Category #4). Figure 4.4 and Table K1 (APPENDIX K) provide a detailed look at the item response rates. Each of the three items received an average rating less than '3', or 'covered.' Pre-service teachers identified item 1.4, "exposure to constructing test items," (2.98) as the topic with the most coverage, followed by "using data to judge individual test items," (2.8) which was one standard deviation below item 1.4. The response rate to "evaluating the quality of publisher-created tests," (2.61) was two standard deviations lower than the exposure to the construction of test items. Each of the three items had 31% to 50% of the respondents report little to no coverage. Teacher- and publisher-created tests make up a significant proportion of the assessments used by teachers. Thus the data from this survey bring up a point of concern if they are not (or at least do not realize they are) receiving instruction on these topics. To ensure that teachers are receiving reliable and valid data from the classroom assessments they are using, they will need to understand how evaluate whether the tests and test items they are using are providing meaningful information. And, at least for the participants of this study, a significant proportion is not receiving exposure to these skills.

The topics covered in this category are not receiving a lot of exposure according to the pre-service teachers. Between one third (items 1.4 and 1.6) and one half (for item 1.5) of the responses indicated little to no coverage on the topics, which indicates that the topics in this category were not extensively covered. This would indicate that teachers were not fully prepared (as a whole group) to gauge whether the assessments and items within those assessments were actually measuring what the students were learning.

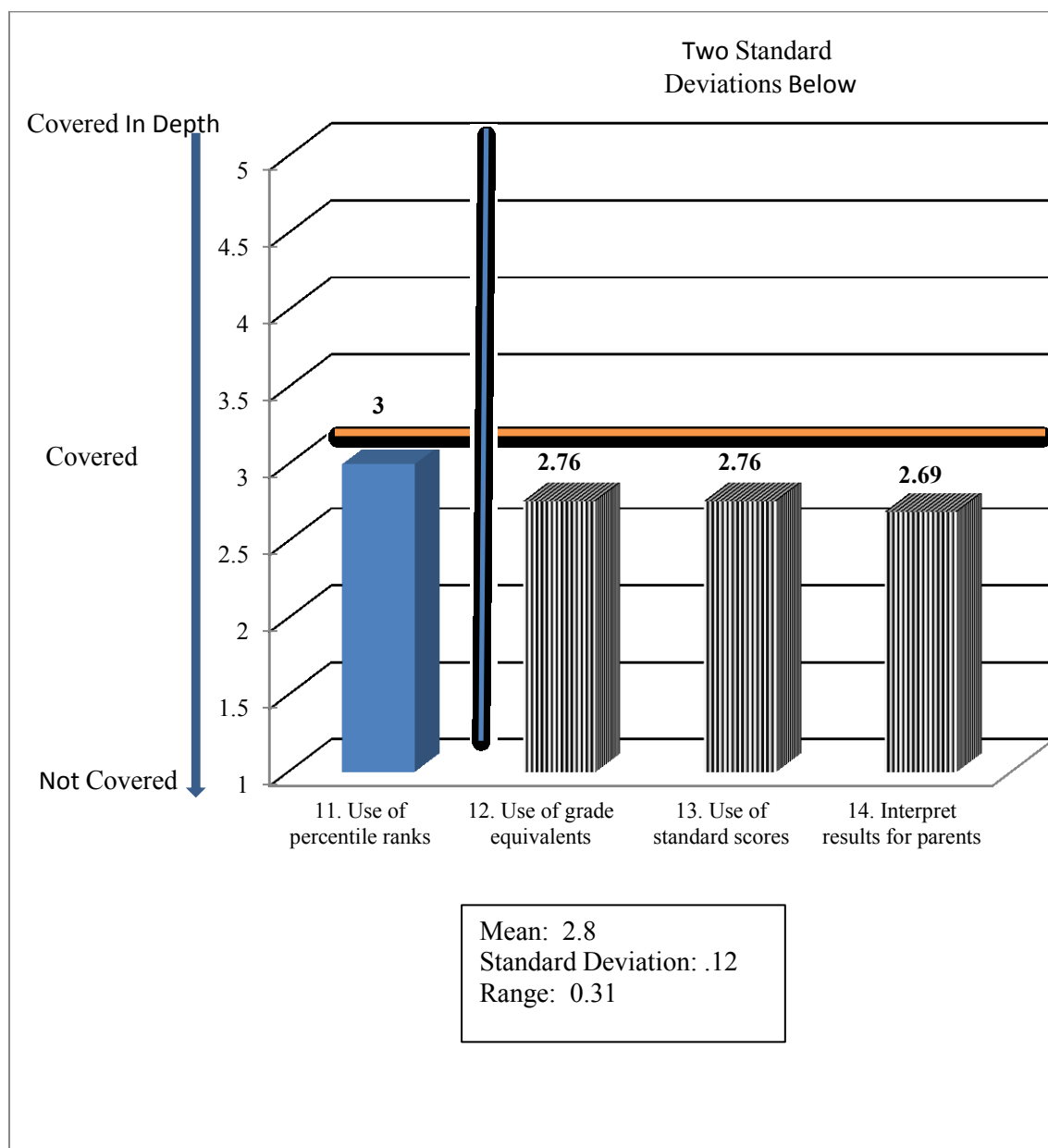
Figure 4.4. Question I - Category #3 – Learning how to construct or evaluate test items



Category #4, “exposure to the interpretation of different types of standardized score results,” had an average response rate of 2.8, the lowest mean exposure rate of the categories, (along with Category #3) and a standard deviation of .12. Figure 4.5 and Table L1 (APPENDIX L) provide a detailed look at the response rates of the items

covered in this item. The differences between the individual items in this category were not large, as evidenced by the highest and lowest ranked items being separated by only one standard deviation. The highest levels of exposure were reported for item 1.11, “using percentile ranks,” (3) and the lowest levels of exposure were for, “interpreting results for parents” (2.69). The lowest response rates came for interpreting results for parents with 47% stating they had little to no exposure with this skill. 31% of the respondents reported little to no exposure to the use of standardized test score results in item 1.19 of Category #1, which, along with the response rates to these items, indicates that approximately 1/3 of the teachers in this study entered their teaching career with exposure to the idea of standardized tests, yet not much exposure on how to interpret the results or use them in a meaningful way. Interpreting scores requires teachers to understand different analysis and some statistics, and this appears to be an area that is not given a lot of focus during their preparation.

Figure 4.5. Question I - Category #4 – Exposure to the interpretation of different types of standardized score results



Category #5, “exposure to key assessment terms”, had an overall mean exposure level of 3.34 and a standard deviation of .33. The terms covered in this category are common terms teachers will likely encounter as they interact with different forms of assessments. Figure 4.6 and Table M1 (APPENDIX M) provide a closer look at the

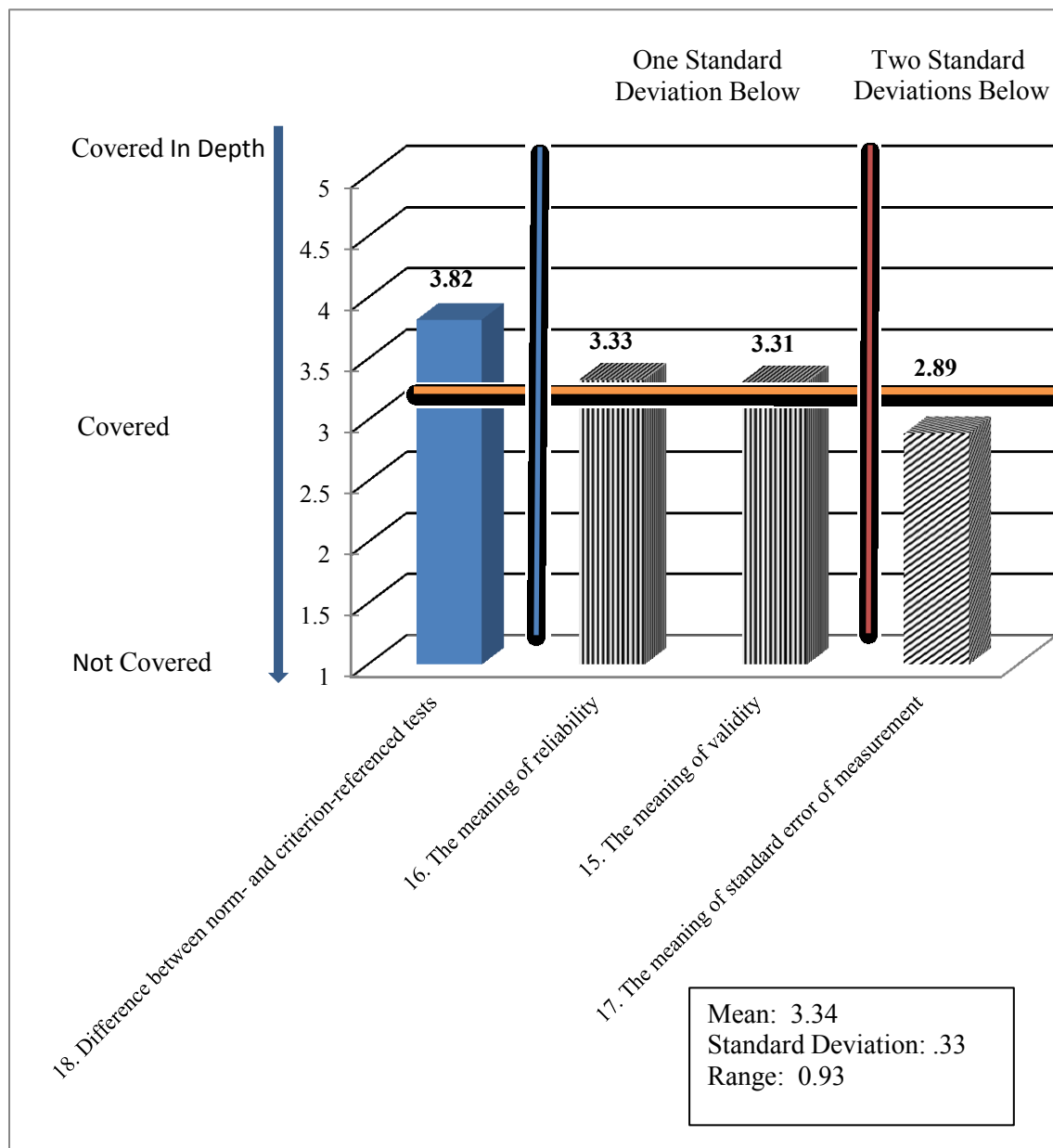
response rates that pre-service teachers provided in response to these items. Item 1.18, “understanding the difference between norm- and criterion-referenced assessments,” had the highest reported levels of exposure (3.82). It was followed by the topics, “the meaning of validity,” and the “meaning of reliability,” (each at 3.33) both one standard deviation below item 1.18. “The meaning of standard error of measurement,” (2.89) had the lowest exposure ratings in this category. Items 1.17 and 1.18 had almost a full point differential in reported exposure levels and were two standard deviations apart.

A closer look at the response rates show that on the topic of understanding the difference between norm- and criterion-referenced assessments resulted in only 6% of respondents noting that the topic received little to no coverage while 35% reported in-depth coverage. This compares to the topics of reliability and validity which had nearly 25% report little to no coverage and only 16% report in-depth coverage. The topic of standard error of measurement had 37% of respondents reporting little to no coverage and only 7% reporting in-depth coverage. While this represents a fairly large difference, learning about the definition standard error of measurement differs from the other items in the category. It is more of a technical term that requires more statistical knowledge. The others items are used more commonly and can be applied to a variety of assessment types.

Prospective teachers should know what norm- and criterion-referenced tests are, as well as to understand the differences between them. Some teachers reported coverage on the meaning of reliability and validity, but almost one-fourth of the respondents reported little to no coverage on these terms. Teachers have to be able to understand whether an assessment they are giving is reliable. They also should understand the

importance of using the results in valid manner. Without this understanding, a possibility remains that the assessments given to students will not provide usable information about the students' academic growth or achievement.

Figure 4.6. Question I - Category #5 – Exposure to key assessment terms



The topic with the highest exposure ratings for Question I was “understanding the difference between formative and summative assessment” with an average exposure

rating of 4.07 and 94% of responses indicating it had been covered to covered in-depth. Every other topic covered in Question I had an exposure rating under a 4.0 (see Table 4.1). The topic rated with the lowest exposure level ratings was “evaluating the quality of publisher-created tests available with textbooks,” with an average rating of 2.61 and was separated from item 1.7 by three standard deviations. The responses for this item indicated that 50% of the pre-service teachers participating received little to no exposure on the topic.

The five topics identified as having the least amount of exposure included interpretations of standardized test scores and creating and evaluating tests or test items. The interpretation of standardized test score topics included “using grade equivalents,” (2.76), “using standard scores,” (2.76), and “interpreting results for parents,” (2.69). Evaluating the quality of publisher created tests had the lowest rating (2.61) and was three standard deviations below the topics with the highest levels of exposure.



Table 4.1. Question I - Assessment topics exposure ratings – 1 – 25

Item	Items rankings	Average Rating Mean – 3.25 SD - .45
1.7	Difference in uses of formative and summative assessments	4.07
1.10	Strengths and limitations of results from standardized tests	3.98
1.23	Use observational techniques	3.91
1.24	Use rubrics	3.84
1.18	Difference between norm- and criterion-referenced tests	3.82
1.25	Use different assessment results to plan instruction	3.74
1.8	Strengths and limitations of results from performance assessments	3.64
1.1	Use different assessment methods/approaches	3.63
1.22	Use portfolios	3.45
1.16	The meaning of the term reliability	3.33
1.15	The meaning of the term validity	3.31
1.2	Determine methods appropriate to assess high order thinking skills	3.15
1.21	Use performance assessments	3.13
1.9	Strengths and limitations of results from teacher-made tests	3.11
1.19	Use of standardized test scores	3.02
1.11	Use of percentile ranks	3
1.4	Construct test items	2.98
1.3	Assess for non-cognitive outcomes	2.89
1.17	The term standard error of measurement	2.89
1.6	Use data to judge individual test items	2.8
1.20	Create/use constructed response tests	2.78
1.12	Use of grade equivalents	2.76
1.13	Use of standard scores	2.76
1.14	Interpret results for parents	2.69
1.5	Evaluate quality of publisher-created tests	2.61
	Yellow - +1 SD from highest ranked item	
	Green - +2 SD from highest ranked item	
	Blue - +3 SD from highest ranked item	

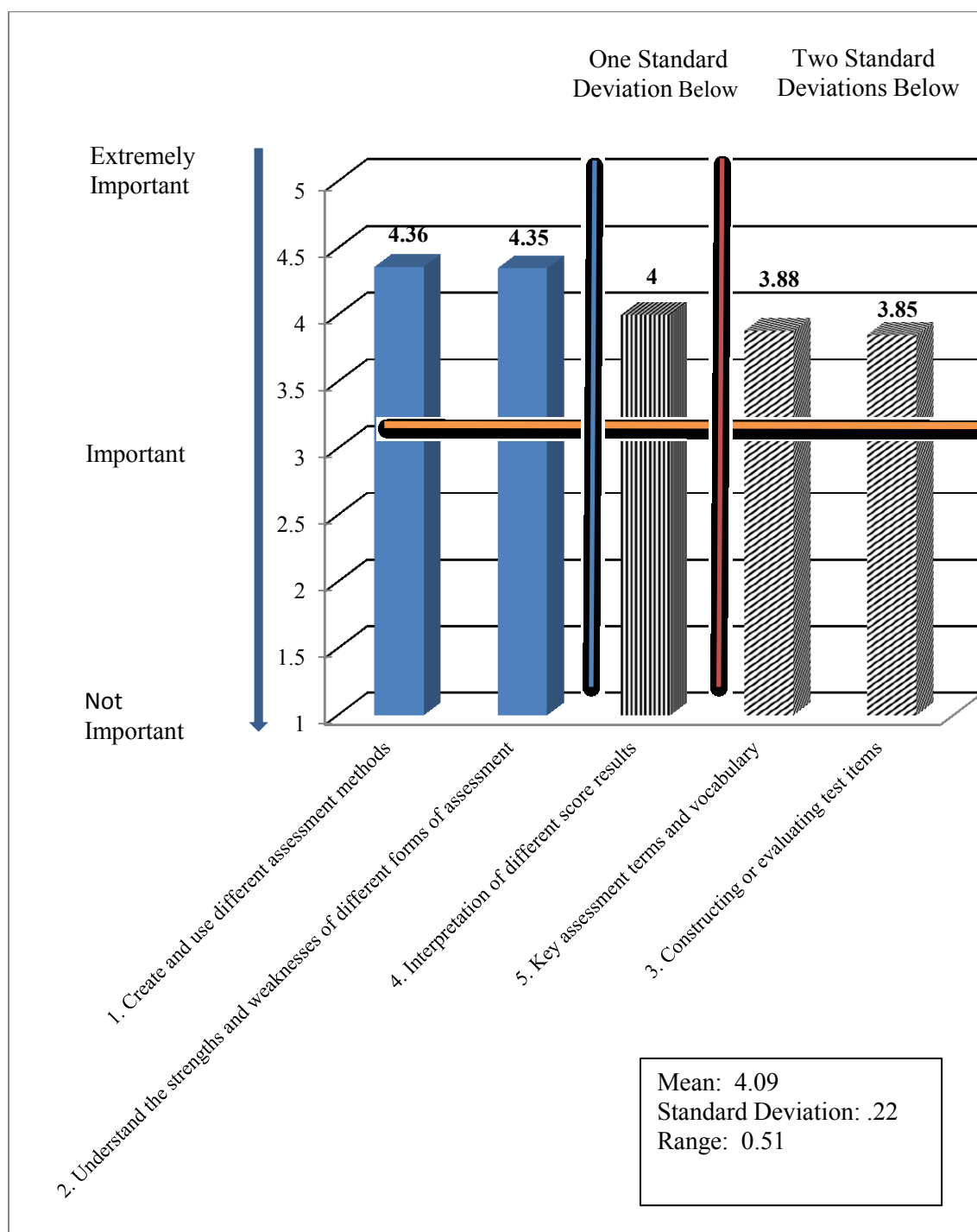
**Question II - Among elementary school principals, what assessment practices are most important to know among beginning teachers working in classrooms?**

Question II addressed the perspectives of the participating principals on what assessment practices they considered to be important for beginning teachers to

understand, thereby providing an indication of the amount of coverage that would be necessary for pre-service teachers to have on the various practices as they leave their teacher education program.

Category #1, “create and use different assessment methods,” received the highest ratings from principals (4.36), followed closely by Category #2, “understand the strengths and weaknesses of different forms of assessments,” (4.35) and were within a standard deviation of each other. Category # 3, “constructing or evaluating test items,” and Category #5, “key assessment terms and vocabulary,” were the two lowest rated items and rated similarly (3.85 and 3.88) and were both two standard deviations below the highest rated categories. Category #4, “the importance of being able to interpret different types of standardized test score results” had a rating of 4.0 and was one standard deviation from the higher rated items. Although there was a 0.51 differential between the high and low category ratings, all of them had high results indicating that the principals that participated in this study felt that assessment is a key component of the educational process and that it is important for new teachers to understand how to use a variety of assessment practices. See Figure 4.7 to view the ratings given to each category.

Figure 4.7. Question II – Among elementary school principals, what assessment principles and practices are most important to know among beginning teachers working in classrooms?

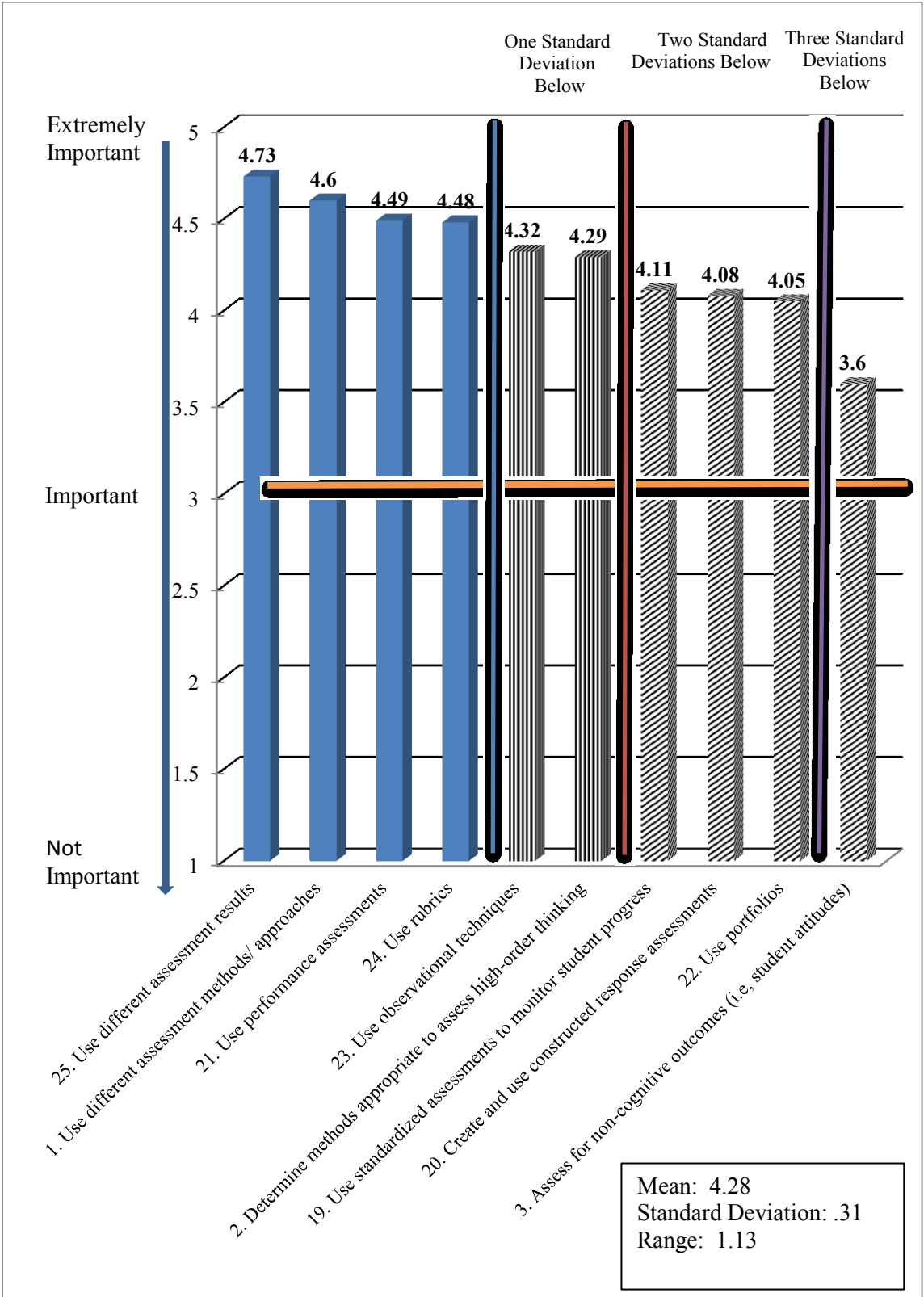


Elementary principals provided an average response rate of 4.28 for the items in Category #1, “creating and using different assessment methods,” and there was a standard

deviation of .31. Figure 4.8 and Table N1 (APPENDIX N) provide an overview of the responses to the items in Category #1. The four highest rated topics were all located within one standard deviation of each other and included “use different assessment results,” (4.73) “use different assessment methods/ approaches,” (4.6) “use performance assessments,” (4.49) and “use rubrics,” (4.48). Topics that had ratings two standard deviations below item 1.24 included - “use standardized student assessments to monitor student progress,” (4.11) “create and use extended response assessments,” (4.08) and “use portfolios,” (4.05). “Assess for non-cognitive outcomes,” was the lowest rated topic and was three standard deviations below item 1.25, but even it had average response rate of 3.6 indicating principals feel it is an important topic for teachers to know about as they begin their career.

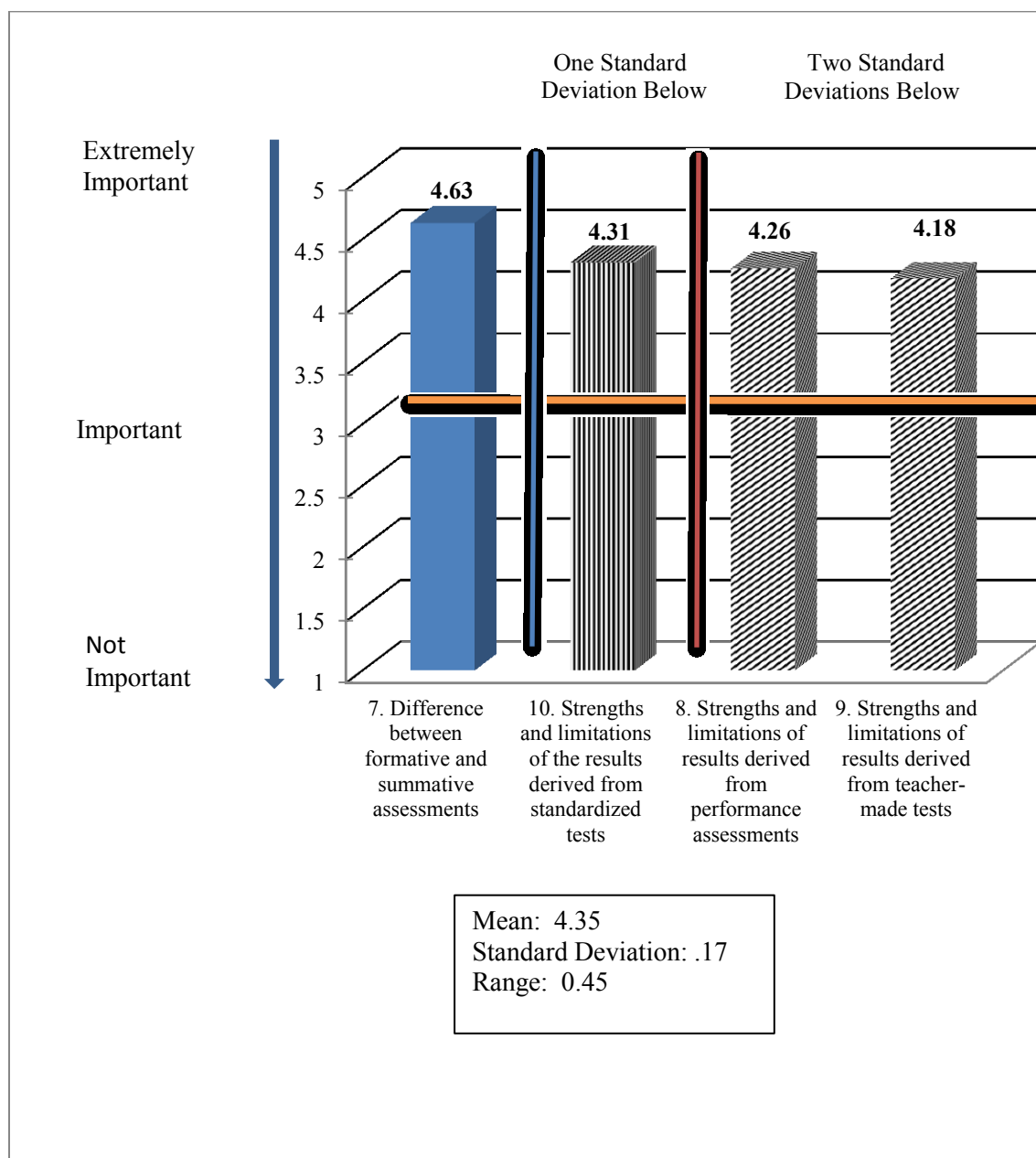
The importance of these topics is obvious when looking at the break down of the individual item responses (see Table N1). Only 2% of this population rated any of the topics with a ‘1’ (Not important), and that was the use of standardized test scores. Further, using standardized assessment to monitor student progress, creating and using constructed response assessments, and using portfolios (all two standard deviations below item 1.25) only had 3% of respondents assigning a ‘2’ (somewhat important) and assessing for non-cognitive outcomes (three standard deviations below item 1.25) had 8% of respondents providing a ‘2’. No other item in this category had a ‘1’ or a ‘2’ assigned.

Figure 4.8. Question II - Category #1 - Create and use different assessment methods



Category #2, “understanding the strengths and limitations of different assessment types” had an overall mean of 4.35 and a standard deviation of .17. See Figure 4.9 and Table O1 (APPENDIX O) for breakdowns of the individual response rates. The highest rated topic was item 1.7, “learning the difference between the uses of formative and summative assessments,” (4.63). “Learning the strengths and limitations of the results derived from teacher-made tests,” (4.18) received the lowest ratings and was two standard deviations from item 1.7. Also two standard deviations below item 1.7 was the topic “learning the strengths and limitations of results derived from performance assessments,” which had an average rating of 4.26. All of these topics had an average rating over 4.0 indicating that; again, the principals in this study consider each item to be important for new teachers to know about. Only the topic, “understanding the strengths and limitations of results derived from teacher-made tests,” received any ‘1’ ratings and that was only 2% of this study’s population. Additionally, the rating of a ‘2’ (somewhat important) was indicated for only 3% of the topic, “strengths and limitations of results from standardized tests,” 7% of, “strengths and limitations of results from teacher-made tests,” and 2% of, “strengths and limitations of results derived from teacher-created tests.” Otherwise it was a collective acknowledgement that these were all, at the very least, important assessment concepts for teachers to know.

Figure 4.9. Question II - Category #2: Understanding the strengths and limitations of different assessment types



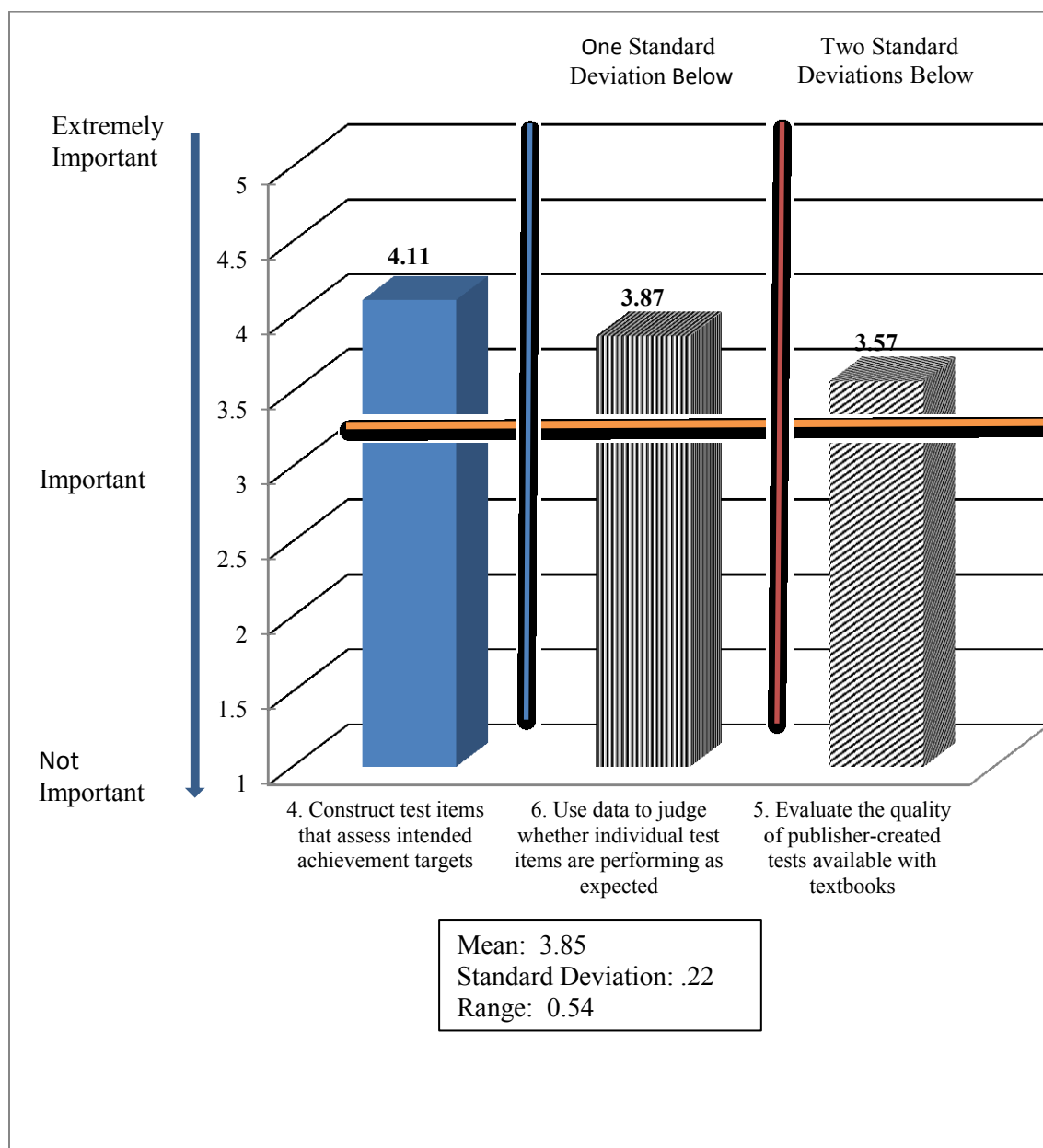
Category #3, “learning how to construct or evaluate test items,” had an average rating of 3.85 and a standard deviation of .22. Figure 4.10 and Table P1 (APPENDIX P) provide an overview of the ratings attributed to the various content knowledge items covered. Item 1.4, “learning how to construct test items,” received the highest ratings

(4.11) of the topics in this category. “Learning how to use data to judge whether individual test items are performing as expected,” (3.87) was one standard deviation below item 1.4. “Learning how to evaluate the quality of publisher-created tests available with textbooks” (3.57) was the lowest rated topic in the category and was two standard deviations below item 1.4.

This category had the lowest overall ratings of Question II, particularly the items which referred to the evaluation of test items and publisher-created tests. While the ratings are low relative to the ratings of other items within Question II, they do not indicate that the principals consider these assessment topics to be unimportant. The lowest average was a 3.57, higher than a ‘3’ or important rating. At least 85% of the responses for evaluating the quality of publisher-created tests, which was two standard deviations below item 1.4, indicated that the topic was important or extremely important.



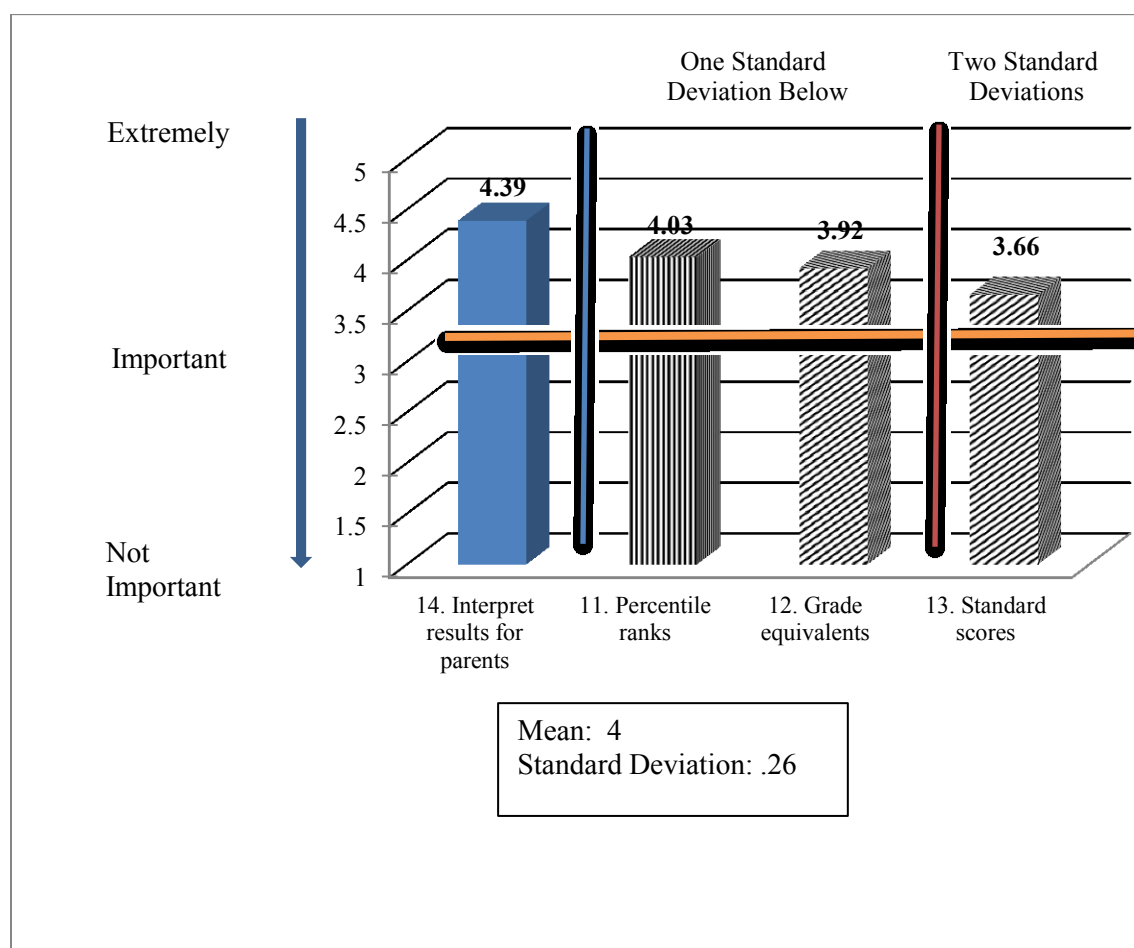
Figure 4.10. Question II - Category #3 - Learning how to construct or evaluate test items



Category #4, “the importance of being able to interpret different types of standardized score results,” had an overall mean of 4.0 and a standard deviation of .26. Figure 4.11 and Table R1 provide a breakdown of the principals’ ratings for each item in this category. Item 1.14, “learning how to interpret the results of standardized tests to parents in a meaningful manner,” (4.39) was the highest rated item in the category. The

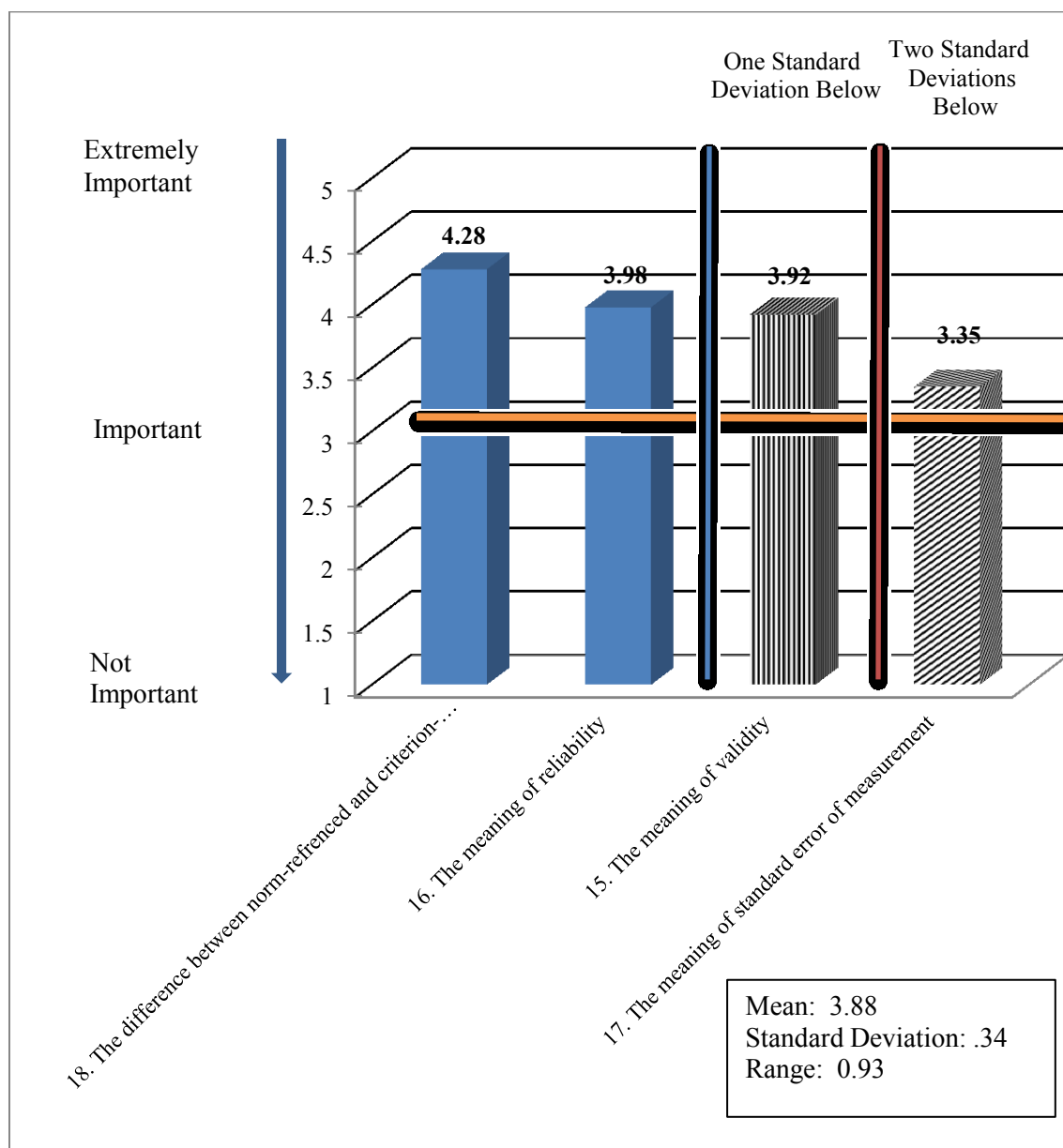
topic, “standard scores,” (3.66) received the lowest rating and was two standard deviations below item 1.14. With a ‘4’ indicating that a topic is very important, the principals identify this category as containing assessment topics that are valued by principals. Only one topic, understanding standard scores, received a ‘1’ (not important) rating and this was only 5% of the respondents. That same item had 85% of the respondents indicate that the topic was either important or extremely important. The interpretation of results for parents was a topic that principals placed a lot of value on; with 57% indicating it as an extremely important topic (97% rated it as a ‘3’ or higher).

Figure 4.11. Question II - Category #4 - The interpretation of different types of standardized score results



Category #5, “knowledge of key assessment terms,” had an overall mean response rate of 3.88 and a standard deviation of .34. Figure 4.12 and Table R1 (APPENDIX R) provide details on the response rates of the individual items. Item 1.18, “the difference between norm- and criterion-referenced tests,” (4.28) was the highest rated topic and was the only item with an average rating over ‘4’. 57% of the respondents indicated that it is extremely important for teachers to know about this concept. The topic “the meaning of reliability,” (3.98) was within one standard deviation and “the meaning of validity,” (3.92) was slightly over one standard deviation away from item 1.18. It is interesting to note that these respondents indicated these two topics, which are critical if teachers are to use assessment results properly, were not deemed as important as understanding the difference between norm- and criterion-referenced tests. While the overall mean differences are not that great, the differences come from the breakdown of the actual responses. 57% of the principals rated understanding the difference between norm- and criterion referenced tests as extremely important, compared to just over 30% of them rating understanding the meaning of validity and reliability as extremely important, leaving a differential of 27%. The topic, “the meaning of standard error of measurement,” (3.35), while is also indicated as important, was rated much lower than the other topics in this category and fell two standard deviations below item 1.18. Only 18% of the respondents identified it as extremely important, and 23% indicated that it was either somewhat or not important.

Figure 4.12. Question II - Category #5 – Knowledge of key assessment terms  
It is important that teachers learn:



The elementary principals that participated in this study placed a high value on the assessment topics identified in Question II. Seventeen of the twenty five items (see Table 4.2 for the items ranked in order of importance as noted by the principals) received an average rating over 4.0 indicating that the topics are very to extremely important and

had a standard deviation of .34. The only item contained in Question II that received a rating under 3.5 was item 1.17, “the meaning of the term standard error of measurement,” (3.35) and it fell four standard deviations from the highest ranked items. The second lowest rated item, 1.4, “evaluate the quality of publisher-created tests,” (3.57) was three standard deviations from the highest rated items, yet it even had 85% of respondents indicating that it was either important or extremely important.

The ratings discussed with Question II provide evidence that this population of principals feel strongly that assessment knowledge on a variety of topics and skills is critical for teachers to possess as they begin their teaching career. This highlights the fact that assessment is not something that exists outside of the realm of teaching and learning, but rather something that is integral to its success or failure. And by virtue of identifying these topics as knowledge that is important for beginning teachers to possess, it also indicates that these are assessment concepts, knowledge, and skills that should be addressed in a teacher education program.

Table 4.2. Question II – Assessment topics level of importance ratings #1 – 25

Item	Items rankings	Average Rating Mean – 4.13 SD - .34
1.25	Use different assessment results to plan instruction	4.73
1.7	Differences in uses of formative and summative assessments	4.63
1.1	Use different assessment methods/ approaches	4.6
1.21	Use performance assessments	4.49
1.24	Use rubrics	4.48
1.14	Interpret results for parents	4.39
1.23	Use observational techniques	4.32
1.10	Strengths and limitations of results from standardized tests	4.31
1.2	Determine methods appropriate to assess high order thinking skills	4.29
1.18	The difference between norm- and criterion referenced tests	4.28
1.8	Strengths and limitations of results from performance assessments	4.26
1.9	Strengths and limitations of results from teacher-made tests	4.18
1.4	Construct test items that assess intended achievement targets	4.11
1.19	Use of standardized test scores	4.11
1.20	Create and use constructed response assessments	4.08
1.22	Use portfolios	4.05
1.11	Use of percentile ranks	4.03
1.16	The meaning of the term reliability	3.98
1.12	Use of grade equivalents	3.92
1.15	The meaning of the term validity	3.92
1.6	Use data to judge whether individual test items are performing as expected	3.87
1.13	Use of standard scores	3.66
1.3	Assess for non-cognitive outcomes (i.e., student attitudes)	3.6
1.5	Evaluate the quality of publisher-created tests available with textbooks	3.57
1.17	The meaning of the term standard error of measurement	3.35
	<b>Yellow</b> - +1 SD from highest ranked item <b>Green</b> - +2 SD from highest ranked item <b>Blue</b> - +3 SD from highest ranked item <b>Purple</b> - +4 SD from highest ranked item	

**Question III - How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure?**

This question provides a comparison between the perceived amount of exposure on various assessment topics given to the pre-service teachers that participated in this study through their teacher preparation program and the importance the principals put upon each corresponding topic for beginning teachers to possess. The exposure levels the pre-service teachers report provide a look at which assessment topics are being incorporated into their teacher education training and how much coverage they are perceived to be given. The principals' feedback further allows for a look at the assessment topics and the depth of these topics they believe beginning teachers should have been exposed to prior the beginning of their teaching careers.

The data from the principal and pre-service teachers have been compared to each other by looking at the means of the responses to the respective items by each sub-group. The ratings were done on the same scale of 1 – 5, however the meaning attached to each value differed by population (See Table 4.3).

Table 4.3. Question III - Pre-service Teacher and Principal Comparison, Values of corresponding scales

	Rating Key				
	1	2	3	4	5
Pre-service Teachers	Not covered	-	Covered	-	Covered in depth
Elementary Principals	Not important	-	Important	-	Extremely important

Category #1, “the creation and use of different assessment methods,” had an overall mean differential of 0.92 between the participating elementary principals and pre-service teachers. Figure 4.13 provides a comparison of the average response rates between principals and pre-service teachers. The principals provided a higher rating of importance for every item covered compared with the perceived levels of exposure identified by the pre-service teachers. The topics, “use constructed response

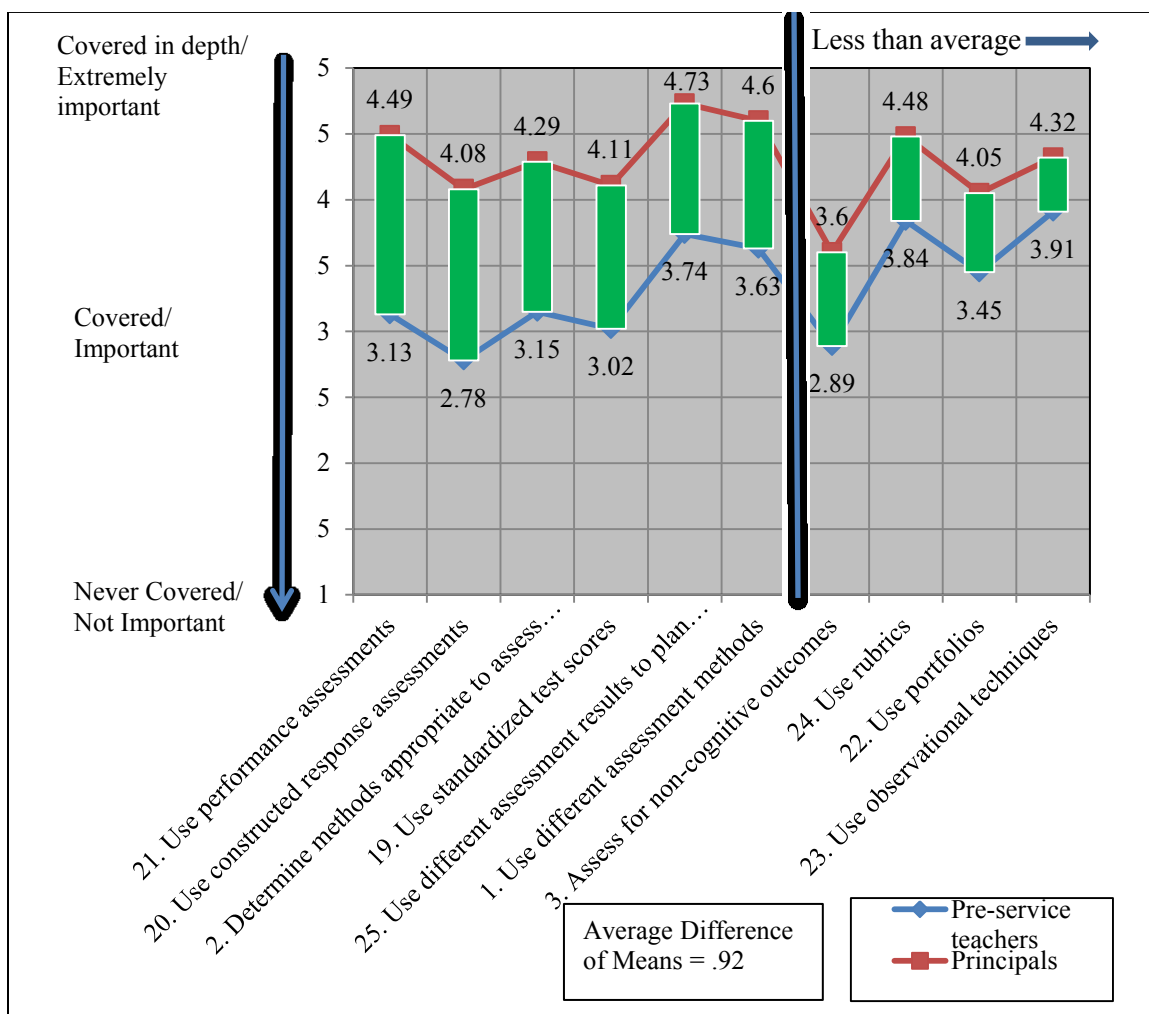
assessments,” (1.3) and “use performance assessments to gauge student progress,” (1.36) had the greatest differentials. Looking at these two topics as examples, it is easy to see how the patterns of responses differ. Tables I1 and N1 (APPENDICES I and N) provide details on the individual responses to the items compared in Question III. On the topic “create and use constructed response assessments,” 37% of the pre-service teachers provided a ‘1’ or ‘2’ rating compared to only 3% of principals. Conversely, 73% of the principals provided a response of ‘4’ or ‘5’ for this topic compared to 25% of pre-service teachers. On the topic of “using performance assessments,” 24% of pre-service teachers provided a response of ‘1’ or ‘2’, while none of the principals provided either of those responses. On the other hand, 91% of principals provided a response of ‘4’ or ‘5’ on this topic compared to only 33% of pre-service teachers.

The four topics rated closest were “assessing for non-cognitive outcomes,” (0.71 - differential) “learning how to use rubrics,” (0.64) “learning how to use observational techniques to gauge student progress” (0.41) and “learning how to use portfolios to gauge student progress” (0.6). While the differences are smaller than the other items, there is still a noticeable difference in the responses. On the topic of learning to use portfolios, 17% of pre-service teachers provided a ‘1’ or ‘2’ compared with 3% of the principals. The difference with this item is also seen on the other end of the spectrum where 39% of principals rated this as extremely important compared to only 15% of pre-service teachers. For using observational techniques, 62% of the pre-service teachers provided a ‘4’ or ‘5’ compared to 90% of the principals providing the same rating. While this topic was apparently covered in the pre-service teachers’ education training, the amount of coverage does not correlate with the level of importance placed upon it by principals.



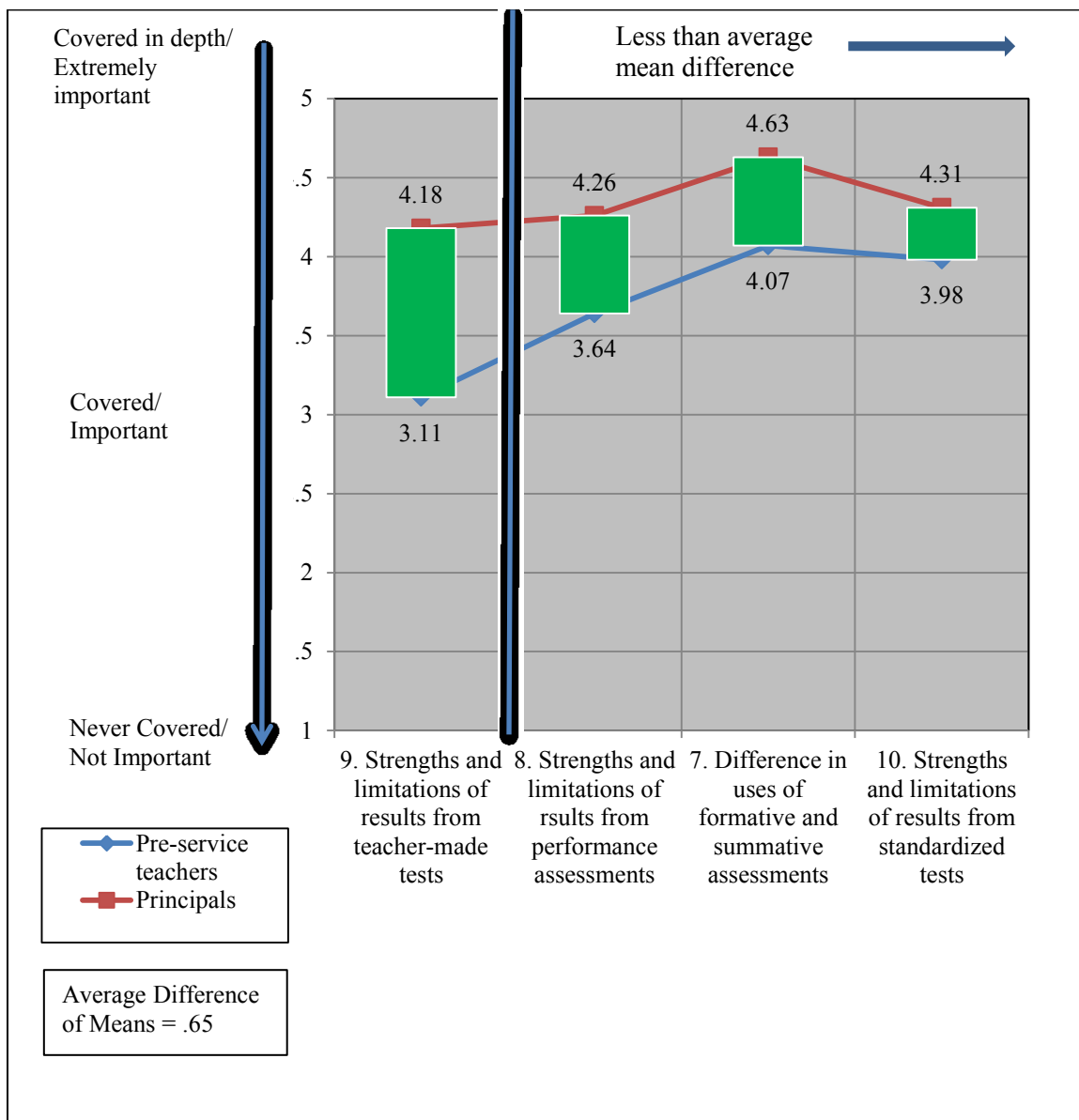
Looking at those four items, as well as the overall differential of 0.92, it indicates that within the participants of this study, there is a misalignment of the priority which is placed on the teaching of the assessment topics contained in Category #1. Pre-service teachers consistently report a lower level of exposure to the items within the category as compared to the respective levels of importance principals identified for beginning teachers to know about them.

Figure 4.13. Question III - Category #1 - Pre-service teacher and principal comparison - Create and use different assessment methods



Category #2, “understanding the strengths and limitations of different assessment types,” had an overall difference in means of 0.65. Figure 4.14 shows the overall differences in ratings provided to each item in the category and Tables J1 (APPENDIX J) and O1 (APPENDIX O) provide a breakdown of the individual item responses. The only item in this category with a difference greater than one was the topic “understanding the strengths and limitations of results from teacher-made tests,” (1.07). This topic received the lowest average ranking for the category from both pre-service teachers and principals. “Understanding the strengths and limitations of results from standardized tests,” had the closest alignment with a difference of only 0.33. Both pre-service teachers and principals had the topic “understand the difference between formative and summative assessments,” rated the highest and had a difference of 0.56 (4.07 and 4.63). This indicates, while not a perfect correlation, that there is congruence between the coverage of the different assessment topics and the perceived importance by the principals.

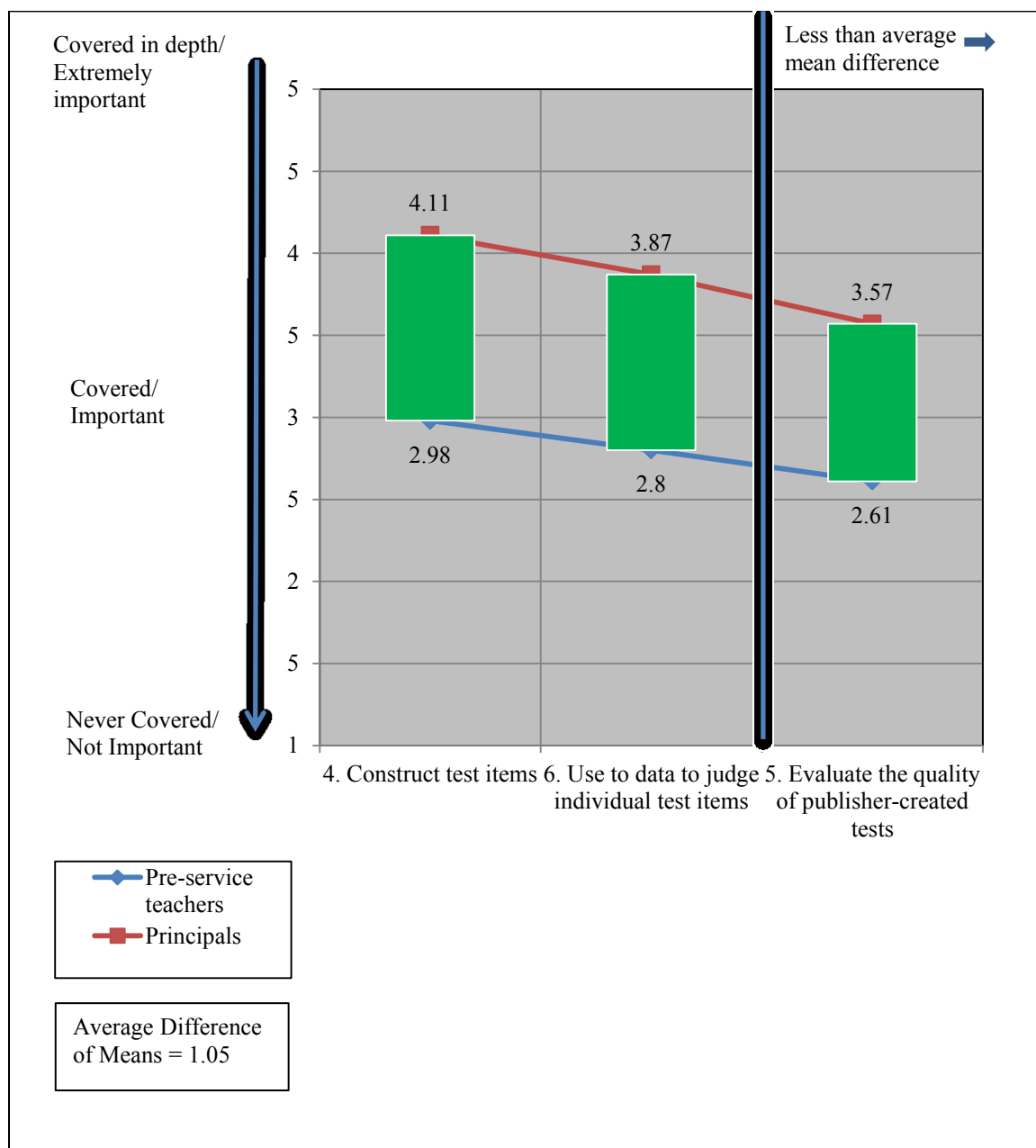
Figure 4.14. Question III - Category #2 - Principal and Pre-service teacher comparison - Understanding the strengths and limitations of different assessment types



The response rates provided by the principals and pre-service teachers for Category #3, “learn how to construct or evaluate test items,” had an overall mean differential of 1.05. The average principal rating was 3.85 and pre-service teachers had an average reported exposure level of 2.8. Figure 4.15 reflects the overall differences between the items while Tables K1 (APPENDIX K) and P1 (APPENDIX P) reflect the

individual response rates for each item. Item 1.4, “learning how to construct test items that assess intended achievement targets” had the greatest differential at 1.13. It was the highest rated item from this category for both populations with the average principal rating of 4.11 compared to 2.98 for the pre-service teachers. 50% of the principals rated this item as extremely important, compared with just 6% of the pre-service teachers reporting that the topic was covered in depth. Conversely, 10% of the principals rated this item as either a ‘1’ or a ‘2’, while 31% of the pre-service teachers did. The topic “using data to judge whether individual test items are performing as expected,” also had a difference greater than the average mean difference. It had a difference of 1.07, and principals placed a great deal more emphasis on this topic than the coverage received by the pre-service teachers. 70% of the principals rated this topic as either very or extremely important and only 21% of the pre-service teachers indicated exposure levels that correspond to those responses. The topic “learning how to evaluate the quality of publisher-created tests available with textbooks,” had the smallest differential - 0.96, and was the lowest rated item within the category for both principals and pre-service teachers. For this topic, 55% of the principals felt that this topic was either very or extremely important compared to only 21% of the pre-service teachers indicating the corresponding amount of coverage.

Figure 4.15. Question III - Category #3 – Principal and Pre-service teacher comparison - Learning how to construct or evaluate test items



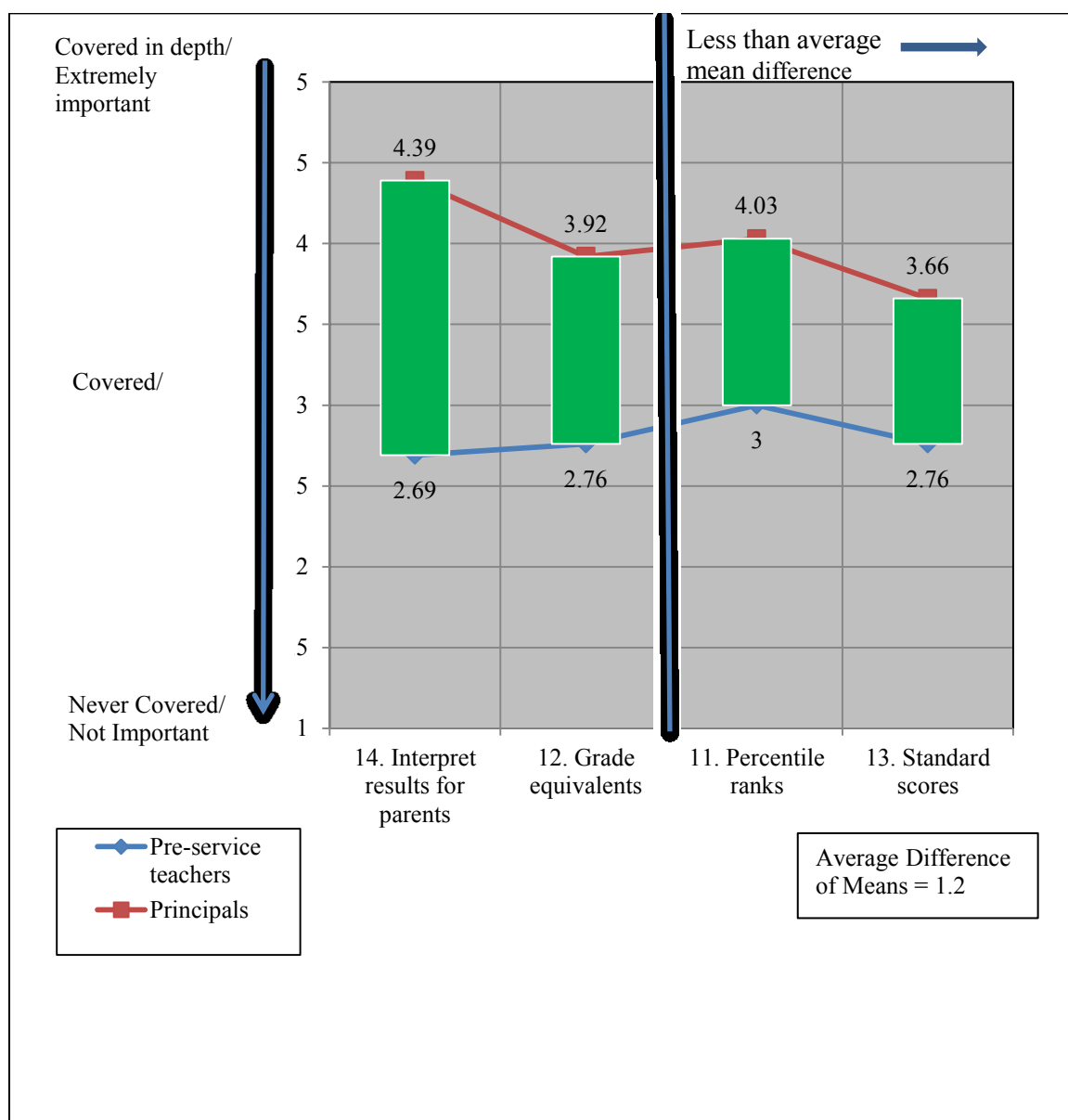
Category #4, “the importance of being able to interpret different types of standardized score results,” had an overall difference in means of 1.2 ( $4 - 2.8$ ) between principals’ and pre-service teachers’ ratings. Figure 4.16 provides a comparison of the items within the category and Tables L1 (APPENDIX L) and Q1 (APPENDIX Q)

provide a look at the individual response rates of the items. The difference of the rates indicates that with the populations participating in this study there is misalignment in the amount of coverage given to this topic area. The largest differential between the two groups was for the topic, “learning how to interpret the results of standardized tests to parents in a meaningful manner.” Principals rated this at 4.39 compared to the pre-service teachers’ rating of 2.69, providing a substantial difference of 1.7. The principals place a great deal of emphasis on this particular item indicating that they realize it is important for parents and students to understand the results that come along with the annual standardized assessments given to all students. Additionally, with the amount of emphasis placed on these assessments by the public, press, and policy makers, there are going to be questions associated with the scores. A look at the response rates show how different the respective ratings are for this topic. 57% of the principals rated this as a topic extremely important for new teachers to know about while only 4% of the pre-service teachers provided the same rating in terms of exposure. At the same time, 47% of the pre-service teachers reported that they had received little to no coverage on the topic. Not only was this the largest difference in ratings, but within the category the principals had this rated as the most important topic to know and the pre-service teachers identified it as the least covered.

The other three items each addressed specific methods of interpreting scores by using percentile ranks, grade equivalents, or standard scores. Learning how to interpret results using grade equivalents represented the topic with the second greatest differential of the category, and along with interpreting the results for parents, were the items with a differential greater than the overall mean differential. For this category, the smallest

difference indicated by the survey results was learning how to interpret standardized test results using standard scores. Principals rated this at 3.66 in comparison with the pre-service teachers' rating of 2.76, providing a difference of 0.9. While this is the item within the category with the smallest differential, that difference is still almost one full point lower than the rating provided by the principals.

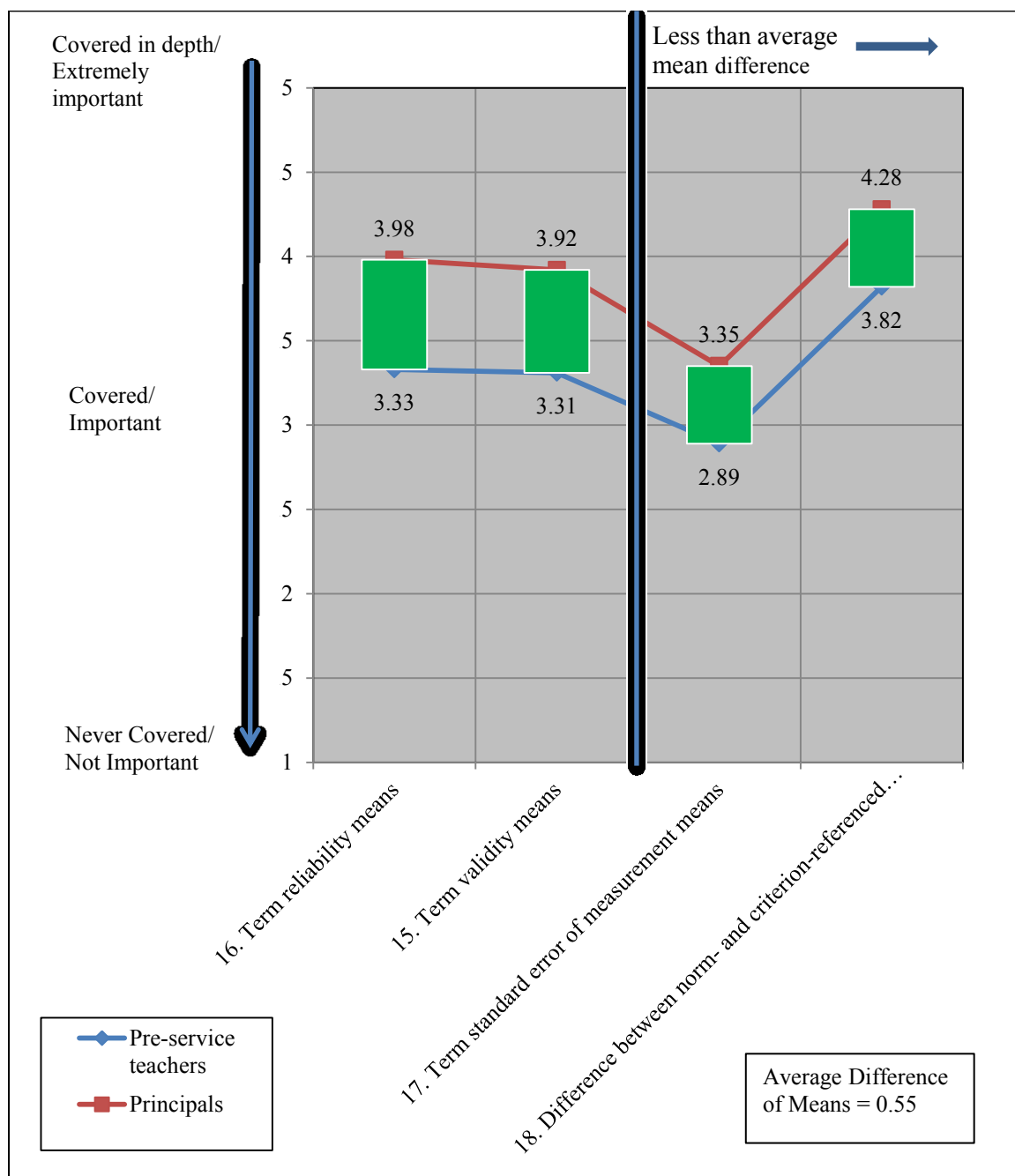
Figure 4.16. Question III - Category #4 - Principal and Pre-service teacher comparison - The importance of being able to interpret different types of standardized score results



Category #5, “exposure to key assessment terms,” had an overall differential between principal and pre-service teacher ratings of 0.54. The principals provided a mean of 3.88 compared to the pre-service teachers’ mean of 3.34. Figure 4.17 provides an overview comparing the differences between populations while Charts M1 (APPENDIX M) and R1 (APPENDIX R) look at the differences in the individual response rates on each of the items in the category. The largest differential for this category was 0.65 on item 1.16, “understanding what the term reliability means.” Knowing what the term validity means within the assessment process was similar and had a mean difference of 0.61. These means of these two items were both more than the overall average. The similarity between these two items makes sense as reliability and validity are topics that are interrelated. Nearly  $\frac{1}{4}$  of the pre-service teachers reported that this topic either received little to no coverage. Conversely, nearly  $\frac{3}{4}$  of the principals identified these two topics as very to extremely important for new teachers to know. The topics, “understanding what the term standard error of measurement means,” and “understanding the difference between norm- and criterion-referenced tests,” both had a differential of 0.46. Both populations provided the lowest ratings for understanding the meaning of standard error of measurement. They also had the highest ratings for understanding the differences between norm- and criterion-referenced tests. Only 13% of the pre-service teachers identified this as a topic with little to no coverage, indicating that this is a topic that is covered at some point during their education training. And while this is an important understanding for teachers to have, it is interesting to note that the ratings were higher for it than for understanding the role of reliability and validity in the assessment process.



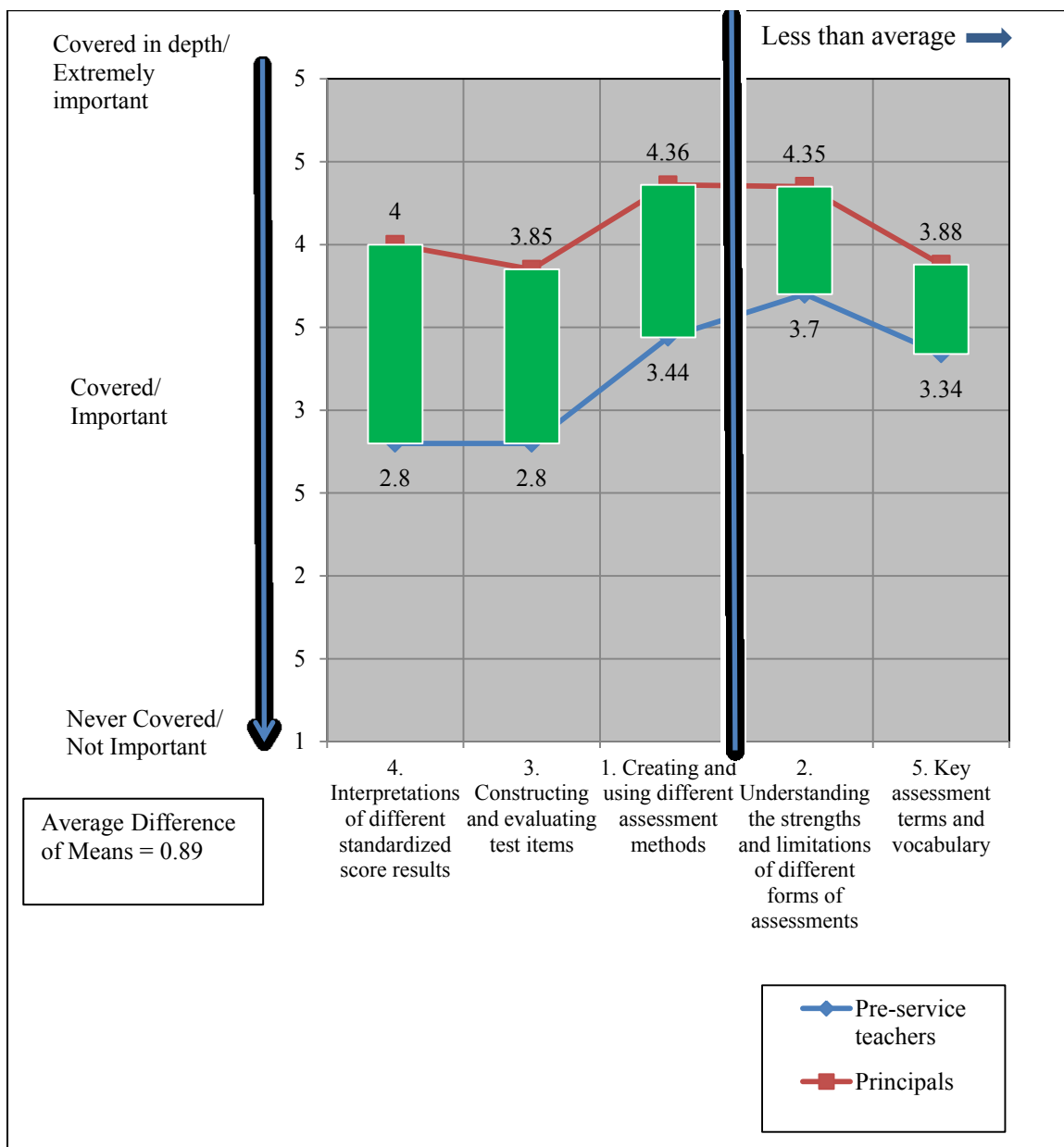
Figure 4.17. Question III - Category #5 - Principal and Pre-service teacher comparison - Exposure to key assessment terms



The overall comparison of the responses for Question III, “How do elementary school principals’ expectations for teacher assessment knowledge compare with self-reported pre-service teachers’ levels of exposure?”, indicates that the elementary

principals in this study place a higher level of importance on beginning teachers' knowledge on the surveyed items than the corresponding pre-service teachers' perceptions of how much coverage the topics received during their time in their teacher education program. See Figure 4.18 for a comparison of the means for each of the five categories covered in Question III. There was not one item in which the pre-service teachers provided a higher average rating than the principals. The overall average of all five categories for principals was 4.09 compared to an average of 3.20 for pre-service teachers. Category #4, "interpretations of different standardized score results," had the largest differential at 1.2 and Category #5, "key assessment terms and vocabulary," had the smallest differential at .54. Pre-service teachers rated understanding the strengths and limitations of different forms of assessment as the category most covered, while principals had it rated as the second most important topic behind creating and using different assessment methods, (the difference between the two was only 0.01). Pre-service teachers had constructing and evaluating test items and interpretations of different standardized tests as the lowest rated (both at 2.8), while principals rated constructing and evaluating test items and key assessment terms and vocabulary as the lowest.

Figure 4.18. Question III - How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure?



**Question IV: What are the self-reported levels of understanding among pre-service teachers of various assessment concepts and skills?**

Question IV addresses the knowledge that the participating pre-service teachers perceived they had towards different assessment concepts and skills. This question was

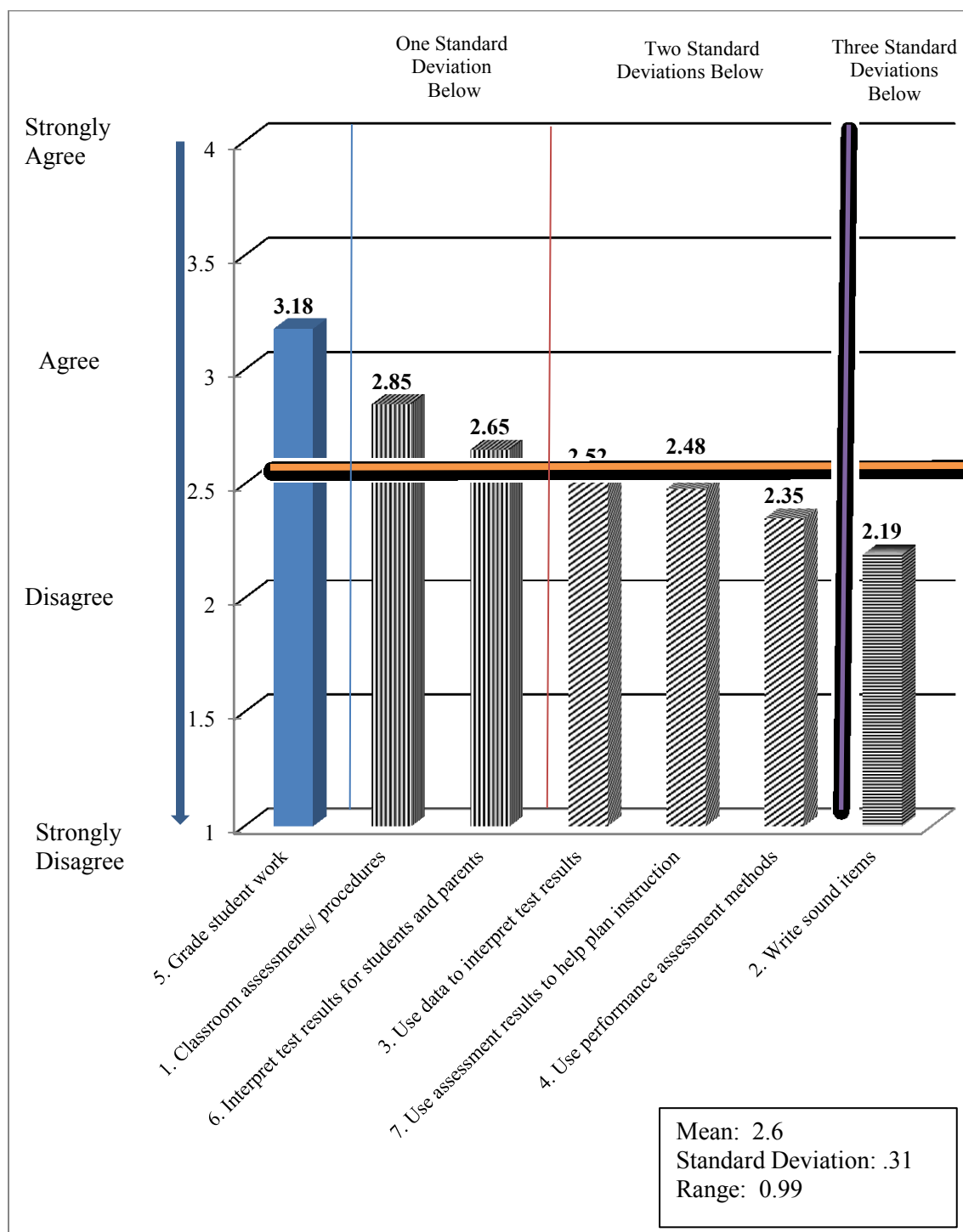
addressed by seven items that included pre-service teachers' self-reported levels of understanding on:

1. classroom assessments and procedures,
2. writing sound test items,
3. using data to interpret test results,
4. using performance assessment methods,
5. grading student work,
6. interpreting test results for students and parents, and
7. using assessment results to help plan instruction.

The overall mean for the items covered in Question IV was 2.6 (4-point scale) and a standard deviation of .31. Item 3.5, "grading student work," had the highest level of reported understanding at 3.18 (see Figure 4.19 for a comparison of the averages for the items). There was a range of 0.99 separating the topic with the highest levels of self-reported knowledge, grading student work, with that of the lowest levels, writing sound test items. The lowest rated item fell three standard deviations below the highest rated item. Topics rated one standard deviation below item 3.5 were, "knowledge of classroom assessment and procedures," with an average rating of 2.85 and "interpret results for students and parents," with a rating of 2.65.

"Writing sound test items," was the assessment skill with the lowest level of reported understanding with a rating of 2.19 and was three standard deviations lower than item 3.5. This was followed by "using performance assessment methods," with a rating of 2.35, "using assessment results to help plan instruction," with a rating of 2.48, and "use data to interpret test results," with a rating of 2.52. All three of these topics fell two standard deviations below item 1.5.

Figure 4.19. Question IV - What are the self-reported levels of understanding among pre-service teachers of various assessment concepts and skills?

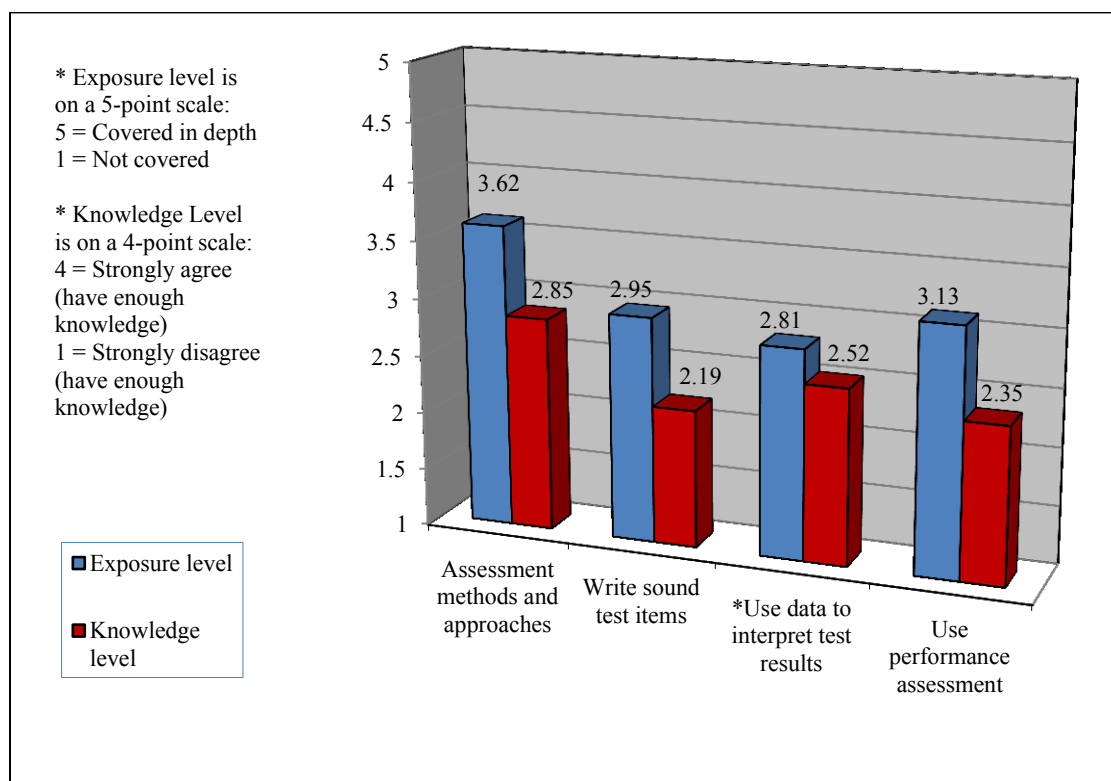


- Items #2, #3, #4, #6, and #7 were all written as “I need additional help with....”

- These items were recoded to run the SPSS to allow them to scale with #1 and #5
- Items #1 and #5 were written as “I have sufficient knowledge of...”
- Item 3.7 was added to the survey and only provided to the teachers in the cohort that was surveyed during the fall of 2008.

A comparison was made between four of the items covered in Question IV, including item 3.1, “assessment methods and approaches,” item 3.2, “write sound test items,” item 3.3, “use data to interpret test results,” and item 3.4, “use performance assessment methods,” and the responses of the corresponding topics in Question I where the exposure levels to the topics were identified (see Figure 4.20). One key difference in the item responses is because the items in Question I was completed on a 5-point scale and those in Question IV were completed on a 4-point scale. The items addressing the use of different assessment methods and approaches were the highest rated for both exposure and knowledge levels. However, the alignment level of exposure compared to knowledge levels does not line up as neatly for the other three topics. Using data to interpret test results was the lowest rated topic in regards to exposure levels, yet it is the second highest rated in regards to knowledge levels. The topics understanding how to write sound test items and use performance assessment methods have approximately the same difference between exposure and knowledge levels at just over a 0.75 differential. While there is some sign that there is an alignment between the reported exposure levels and knowledge levels, it is too difficult to draw any real conclusions to the differing scales and small number of matching items and topics.

Figure 4.20. Question IV - Pre-service teacher perceptions of assessment knowledge compared to levels of exposure



**Question V - How do elementary school principals conceive of assessment understanding for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers?**

Question V provides principals' perceptions of new teachers' understanding of assessment practices and compares that with the pre-service teachers' perceptions of their own knowledge. Figure 4.21 contains the ratings that indicate the principals' perceptions of new teachers' understanding of assessment concepts and Figure 4.22 contains the comparison of ratings between the principals and pre-service teachers.

The overall rating provided by principals was only 1.9 on a 4-point scale where a '2' indicated disagreement. There was a standard deviation of 0.47. On average, across the items, the participating principals indicated that they did not agree that new teachers have sufficient knowledge of the assessment topics addressed in this question. Even the

highest rated item, item 3.5, “new teachers possess enough knowledge to grade student work accurately and fairly,” had a rating of only 2.77 which indicates that they don’t agree that new teachers possess enough assessment knowledge. Falling within one standard deviation was the topic, “new teaches have sufficient knowledge with classroom assessments/ procedures,” with a mean of 2.43. The remaining five topics are all fell two standard deviations below item 3.5. “New teachers have enough knowledge to use assessment results to plan instruction,” was the lowest rated topic with an average of only 1.41. “New teachers have sufficient knowledge to use data to interpret test results,” and “new teachers have sufficient knowledge to use performance assessment methods,” each had an average rating of only 1.59. These results, as indicated in Figure 4.20 reflect that the elementary principals participating in this study do not have confidence in the knowledge levels that new teachers bring with them to their career in regards to these assessment topics.

Pre-service teachers rated their knowledge level higher on every item compared to the corresponding ratings of knowledge the principals indicated for new teachers. The difference in the overall means for the seven items was 0.7 (on a four point scale), where the pre-service teachers provided a mean of 2.60 compared with the principals’ mean of 1.90. Three of the topics had a difference less than the average mean difference of 0.7, including the topics, “grading student work,” “classroom assessments/ procedures,” and “writing sound test items.” Grading student work received the highest ratings by both groups, and general knowledge about classroom assessments and procedures was the second highest rated item by both. This was the only topic which had either group provided a rating over ‘3’ and that was provided by teachers (3.18). Writing sound test



items was the item the pre-service teachers reported the least amount of knowledge on, while principals indicated it was the third most known topic by new teachers. But, the principal still provided a rating of 1.79 and that was 0.40 less than the pre-service teachers rating.

The topic with the greatest differential in ratings, 1.07, was “using assessment results to help plan instruction.” Pre-service teachers rated their knowledge at 2.48 compared to the principals’ perceived knowledge level of new teachers rated at 1.41. Figure S1 (APPENDIX S) provides a closer look at the responses and the ratings provided. Nearly 96% of the principals indicated that they either disagreed or strongly disagreed that new teachers had sufficient knowledge on how to use assessment results. Only 48% of the pre-service teachers self-identified as not having sufficient knowledge, while nearly 53% indicated that they are confident in this knowledge. This is a large difference, especially when noting that basically none of the principals felt that new teachers had sufficient knowledge.

The other topics which had differences greater than the average mean difference included interpreting test results for students and parents, using data to interpret test results, and using performance assessment methods. The second largest differential between the perceptions of knowledge was item 3.6, “sufficient knowledge of how to interpret test results for students and parents,” which had an average differential of 0.95. Figure T1 (APPENDIX T) provides a closer look at the ratings chosen for this item. 59% of the pre-service teachers agreed or strongly agreed that they had sufficient knowledge of how to interpret the results from parents. Once again the perceptions held by principals do not align with the pre-service teacher perceptions. Nearly 90% of the

principals indicate that new teachers do not have sufficient knowledge in this topic. In fact, 48% of them strongly disagreed with the statement that new teachers have sufficient knowledge in this area.

Item 3.3, “sufficient knowledge to use data to interpret test results,” had a 0.93 difference. Figure U1 (APPENDIX U) shows the item response comparison for the item. Nearly 94% of the principals indicated that they disagree with the idea that new teachers have sufficient knowledge on this topic. Pre-service teachers had nearly 48% of the responses indicating agreement with the principals, while nearly 42% felt they did have sufficient knowledge.

Item 3.4, “sufficient knowledge to use performance assessment methods,” had a differential of 0.76 separating the two means. While the difference here is still fairly large, the responses to the items begin to even out to some degree. Figure V1 (APPENDIX V) shows the breakdown of the principals and pre-service teachers’ responses to this item. 43% of the pre-service teachers responded that they agreed or strongly agreed that they had sufficient knowledge on the topic, however, 57% responded that they did not have sufficient knowledge. There continued to be large misalignment because 47% of the principals strongly disagreed that new teachers had sufficient knowledge, compared to 11% of the pre-service teachers with the same rating.

Items 3.1, “sufficient knowledge of classroom assessments and procedures,” and 3.5, “grade student work,” have differentials between 0.4 and 0.41 between the means. As Figure W1 (APPENDIX W) and Figure X1 (APPENDIX X) indicate, the participating principals show more confidence in new teachers’ levels of assessment knowledge on these topics. 52% of the principals still disagreed or strongly disagreed

that new teachers had sufficient knowledge of classroom assessments and procedures. However, of that number, only 8% strongly disagreed, which is lower than the response rate given to previous assessment topics that were covered. And 47% of the principals agreed or strongly agreed that new teachers had sufficient knowledge. Pre-service teachers had only 24% of the respondents disagreeing or strongly disagreeing that they had sufficient knowledge and 65% agreeing that they had sufficient knowledge. What is noticeable with this item is that between 87% and 88% of both populations provided ratings of '2' or '3', and had less responses on either of the extremes.

Item 3.5, "grade student work accurately and fairly" received the highest pre-service teacher rating at 3.18 and principal rating at 2.77. The principals' confidence in new teachers' knowledge of this assessment was more obvious here than with the other items covered in Question V. 74% of the participating principals either agreed or strongly agreed that new teachers had sufficient knowledge of grading student work compared with 91% of pre-service teachers that rated their own knowledge in the same way. Once again the principals stayed away from the extreme responses and responded with either a '2' or a '3' with 87% of the responses. Although over  $\frac{1}{4}$  of the principals still believed that new teachers lacked sufficient knowledge on this topic, it is a much lower rate than the other topics within this question.

Pre-service teachers rated Item 3.2, "learning how to write sound test items" as the lowest within this question with a rating of 2.19. This item also had the smallest differential in means with a difference of 0.4. Figure Y1 (APPENDIX Y) provides the breakdown of the response rates within this item. Only 13% of the principals agreed that new teachers had sufficient knowledge of this topic, however, the reason that that average

is closer to the pre-service teachers rankings is because the pre-service teachers expressed their overall lack of knowledge with this item. Approximately 66% of the pre-service teacher respondents indicated that they disagreed or strongly disagreed that they had sufficient knowledge. This is a topic the pre-service teachers did not feel as comfortable with and that is why, despite 87% of the principals indicating that they disagreed or strongly disagreed that new teachers had sufficient knowledge with this topic, that there is a closer alignment in perceived knowledge levels. Unfortunately, the reason that the ratings are so similar is because not only did the principals lack confidence in teacher knowledge, but the pre-service teachers were not confident in their own knowledge levels.

And while there was not a perfect correlation in the respective order for the ranking of each item, there was a general trend that is noticeable in Figure 4.21. Topics the pre-service teachers identified they needed the most support with, were ranked in an order similar to the same items the principals identified that new teachers had the least understanding of.

Figure 4.21. Question IV - How do elementary school principals conceive of assessment practices for beginning classroom teachers?

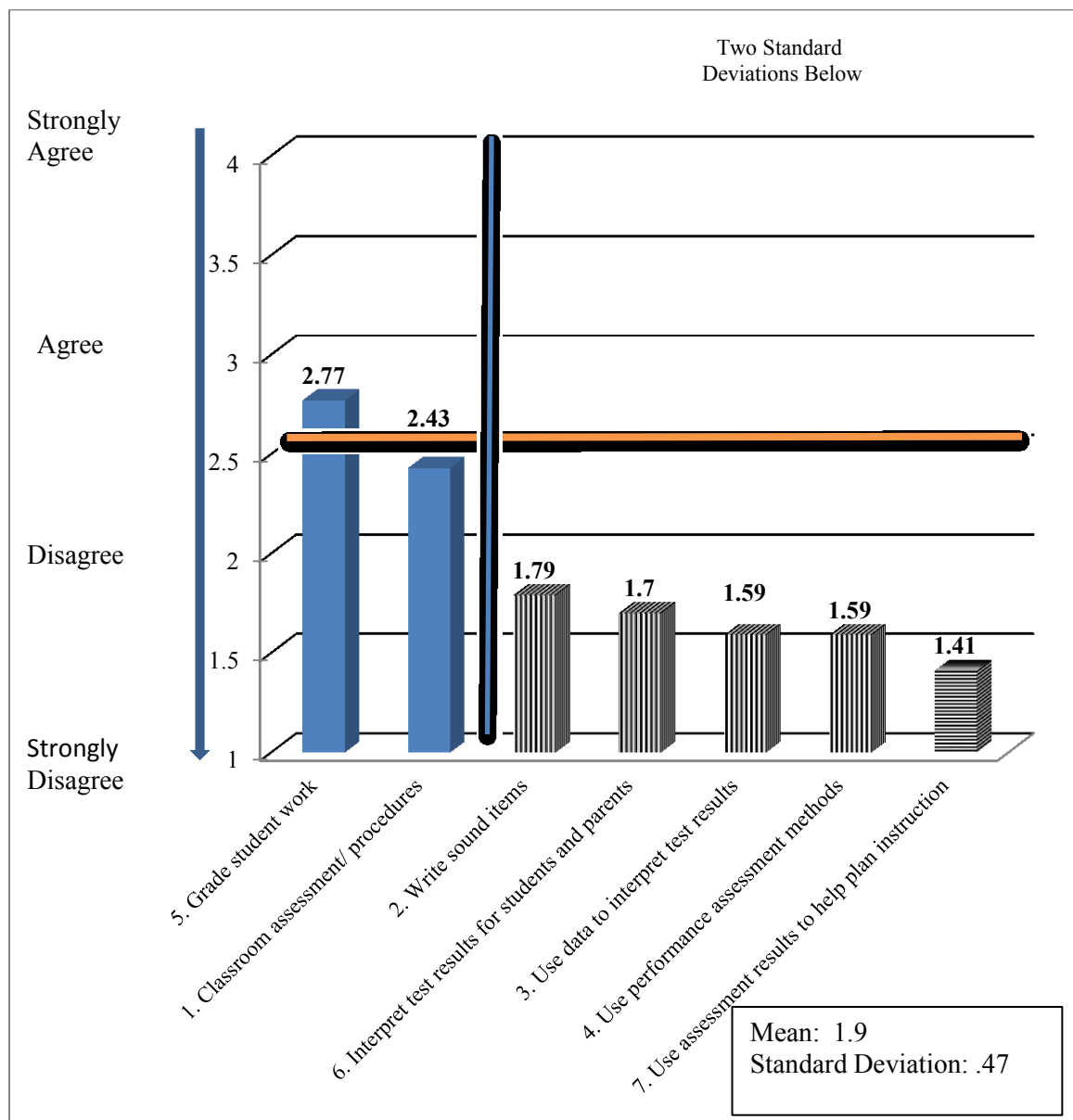
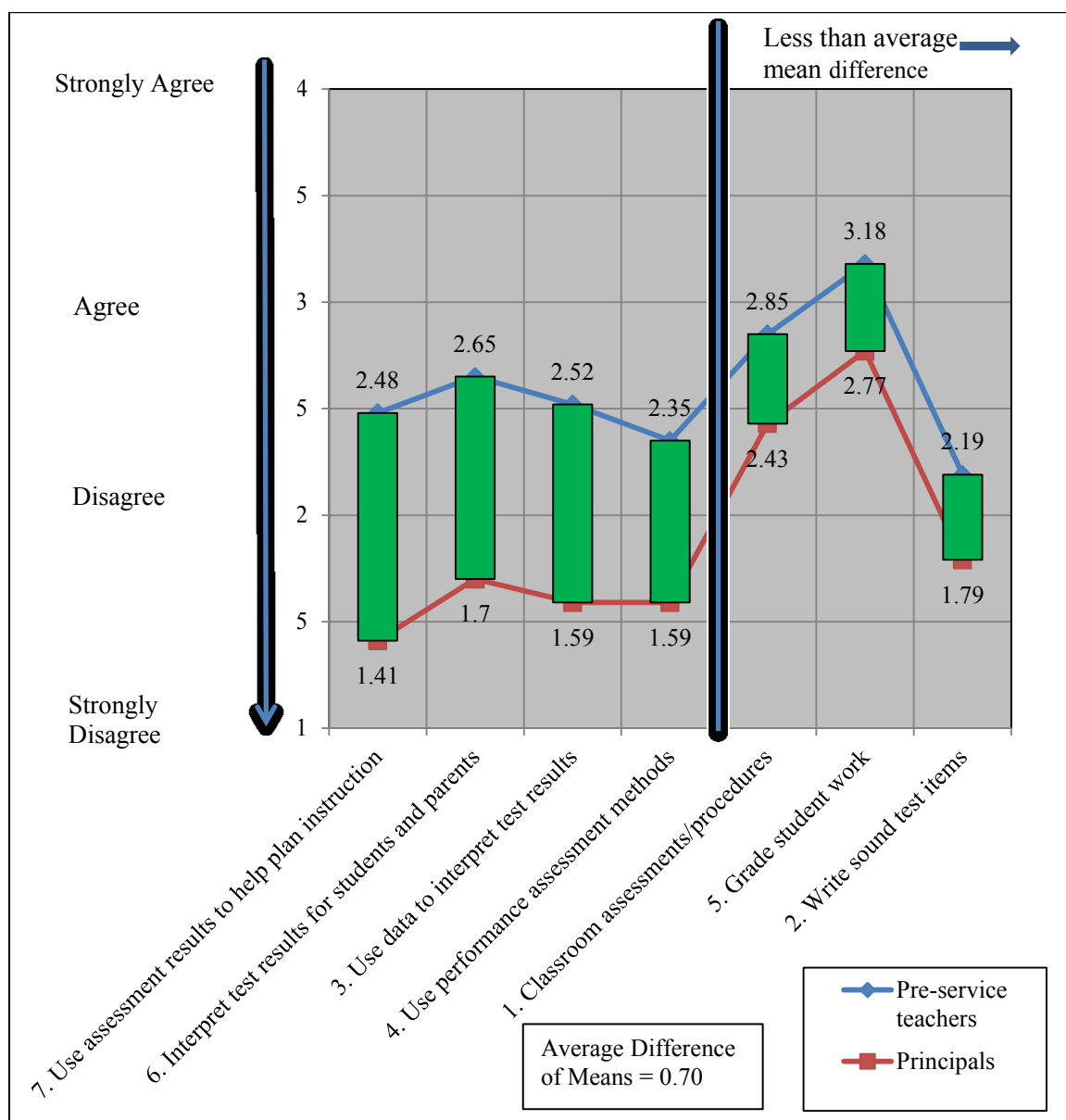


Figure 4.22. Question V - How do elementary school principals conceive of assessment understanding for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers?



Pre-service teacher and principal comparison of means for Item #1 - #7

- Questions 2, 3, 4, 6, and 7 were all written as “I need additional help with....”
- Questions 1 and 5 were written as “I have sufficient knowledge of....”
- Items 3.2, 3.3, 3.4, 3.6, and 3.7 were recoded to get them to scale

**Question VI - What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?**

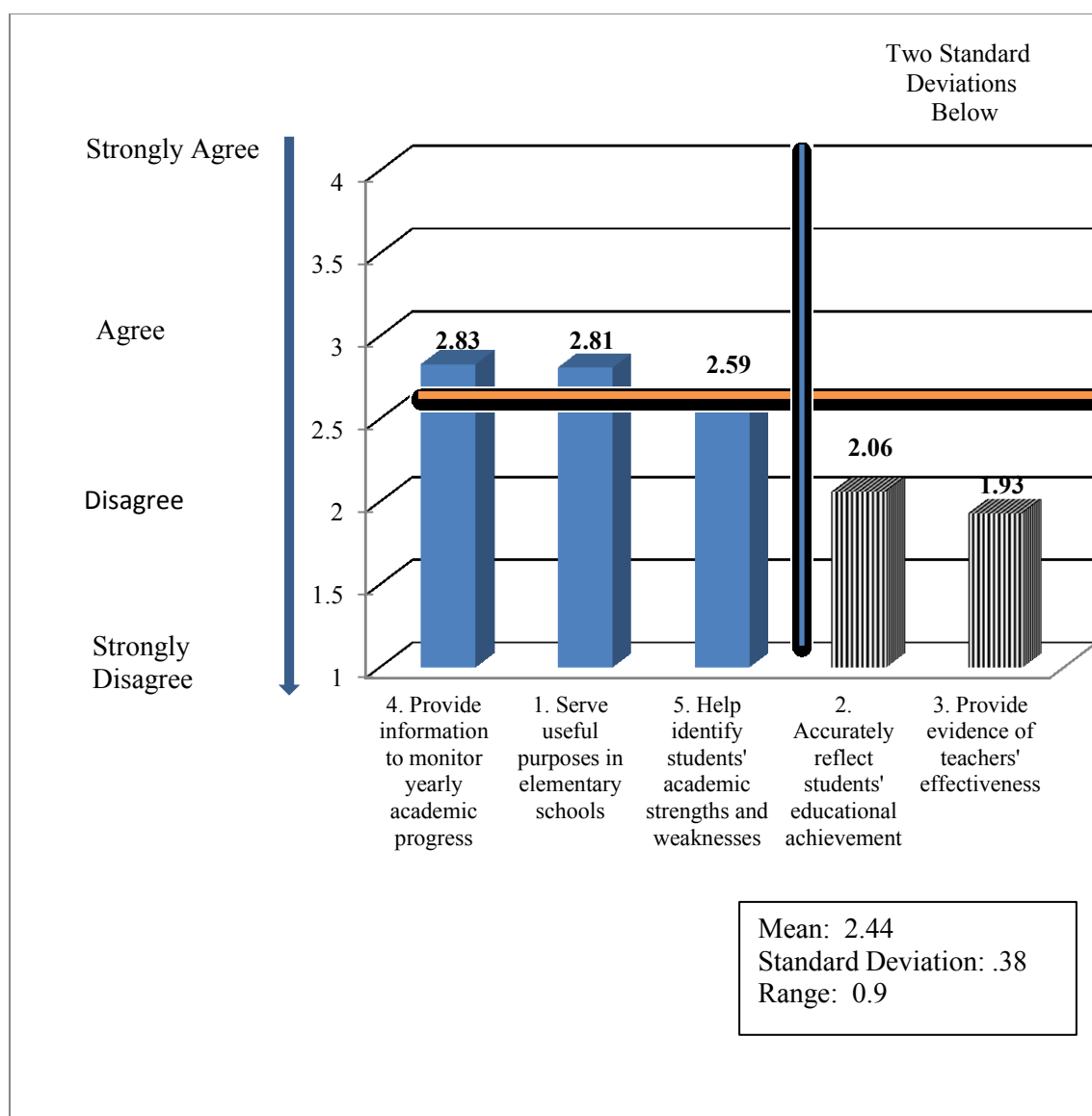
Standardized assessments and the accompanying results play a large role in public education. Depending on the audience they are viewed as the ‘gold standard’ of a method to judge academic achievement as well as be vilified as a serious problem within education. The five items covered by Question VI address the attitudes of pre-service teachers towards the usefulness of standardized assessments and whether the accompanying results to:

1. serve useful purposes in elementary schools,
2. accurately reflect students’ educational achievement,
3. provide evidence of teachers’ effectiveness
4. provide information to monitor yearly academic progress, and
5. help identify students’ academic strengths and weaknesses.

The overall mean of the items contained within this question was 2.44 on a 4-point scale with a standard deviation of .38. Figure 4.23 provides details on for Question VI. Pre-service teachers indicated that of the items included in the survey, item 2.4, “standardized achievement assessments can provide information that makes it possible to monitor year-to-year academic progress with elementary school students” was the highest rated with a 2.83. This was followed by the topic, “standardized achievement assessments can serve useful purposes in elementary school” with a rating of 2.81 and the topic, “help identify students’ academic strengths and weaknesses,” which had a rating of 2.59, both within one standard deviation of item 2.4. Given that a ‘3’ means agreement, the fact that the highest rating was under ‘3’ indicates that the pre-service teachers in this study are not fully supportive of the importance or usefulness of standardized assessment results within and across schools.

The topic, “scores from standardized achievement assessments can accurately reflect students’ educational achievement,” had the second lowest rating at 2.06 and was followed by the topic, “standardized achievement assessments provide an important source of evidence of teachers’ effectiveness in the classroom,” with a rating of 1.93. These two items received much lower ratings than the other three items, with a range of scores of 1.25, and were two standard deviations below item 2.4.

Figure 4.23. Question VI – What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?





**Question VII - What are the attitudes of elementary school principals toward standardized achievement results in the classroom? How congruent are these attitudes with the attitudes of pre-service teachers?**

Question VII addressed the principals' attitudes toward the use of standardized assessment results and compared those with the perceptions pre-service teachers had on the same topics. Figures 4.24 and 4.25 provide the details for the principal ratings and a comparison with the perceptions of the pre-service teachers. The principals' overall mean rating for the five items was 3.04 and had a standard deviation of .21, indicating that overall, they agree that these results can provide benefits to elementary schools. Principals rated item 2.1, "standardized achievement assessments can serve useful purposes in elementary schools," with the highest rating at 3.34. Falling within one standard deviation was the topic, "standardized achievement assessments provide information to monitor yearly academic progress." The topic, "standardized achievement assessments help identify students' academic strengths and weaknesses," was one standard deviation below item 2.1. The lowest rating provided by principals was for the topic, "scores from standardized achievement assessments can accurately reflect students' educational achievement in elementary schools," at 2.79, and along with the topic "provide evidence of teachers' effectiveness," was two standard deviations below item 2.1. The difference in the two items provides a range of 0.55.

The ratings provided by pre-service teachers followed a similar track, though the principals provided higher ratings on every item. Across the five items, the pre-service teachers had a mean of 2.44 compared to the 3.04 mean provided by principals with an average difference of 0.60. The principals had item 2.1 rated the highest with a rating of 3.34 while it was the second highest for pre-service teachers with an average of 2.81.

Principals rated the topic stating the standardized achievement results accurately reflect students' educational achievement the lowest (2.79), whereas pre-service teachers had it as the second lowest with an average response of 2.06. Pre-service teachers rated the idea "that standardized achievement assessments provide an important source of evidence of teachers' effectiveness in the elementary classroom," the lowest (1.93), and principals had that rating as the second lowest (2.85).

Figure 4.24. Question VII – What are the attitudes of elementary school principals toward standardized achievement results in the classroom?

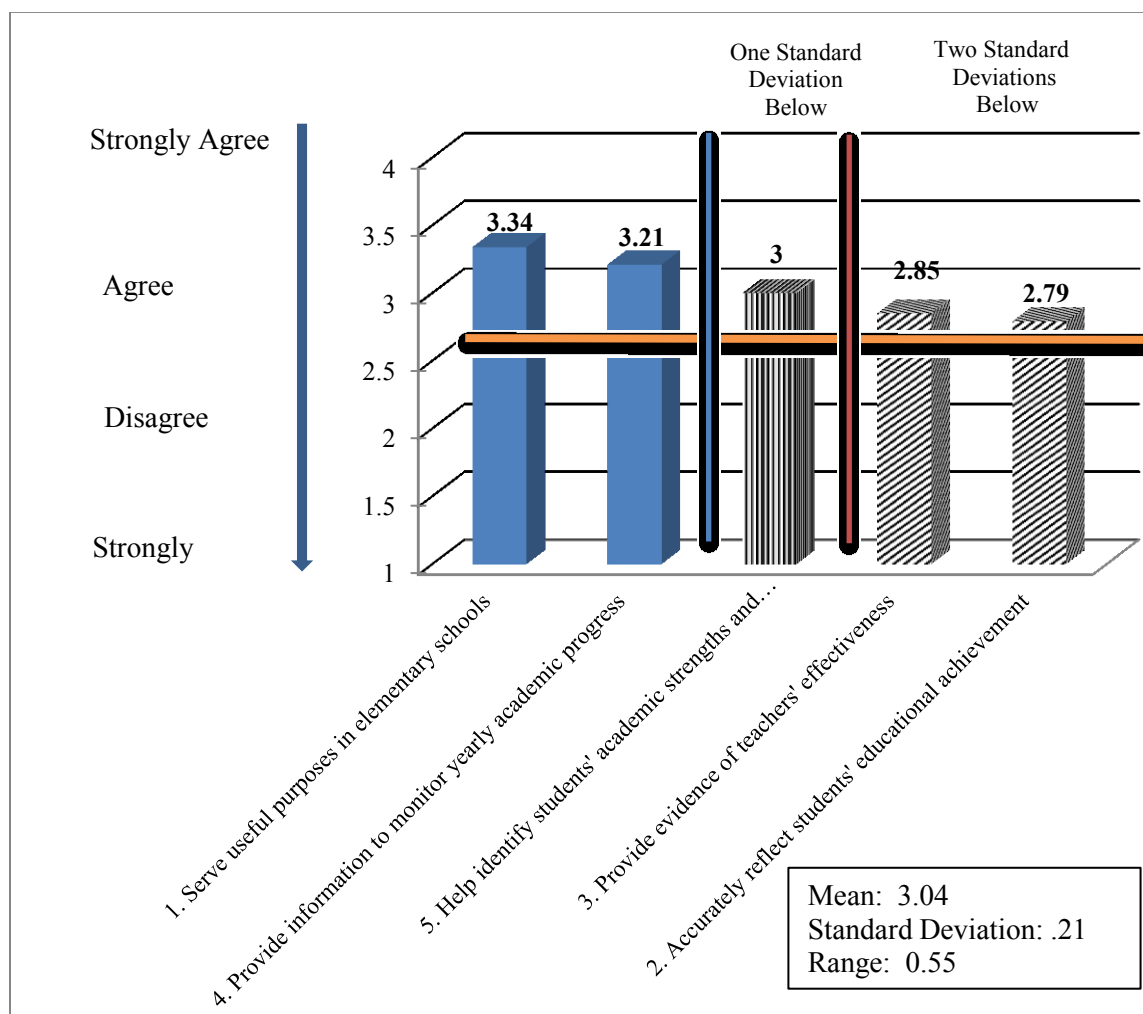
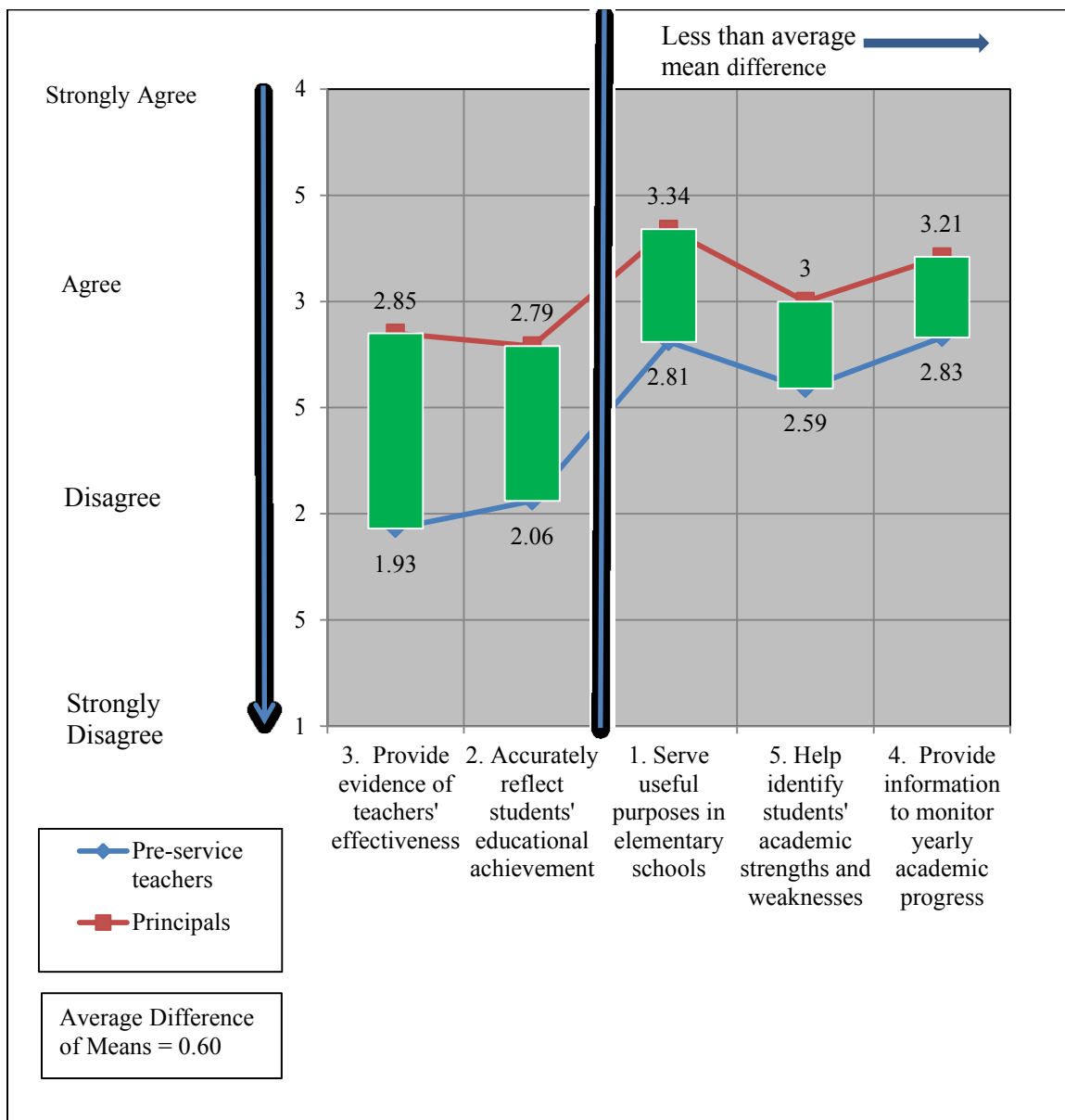


Figure 4.25. Question VII - Principal and Pre-service teacher comparison – Attitudes towards the use of standardized achievement results in the classroom



**Conclusions:** (See Table 4.15 for overall Question means)

The results of the surveys showed that there are differences between the perceptions of coverage and knowledge that pre-service teachers reported in relation to the perceptions and desires of elementary school principals on assessment skills and knowledge. Principals consistently reported higher levels of importance to assessment

topics than the identified levels of coverage on the same topics by the pre-service teachers. There was not one item in Questions I – III that the pre-service teachers rated higher than the principals. This indicates that the amount of coverage on assessment topics these pre-service teachers received did not match the amount of assessment knowledge principals want new teachers to possess.

Further, the perceived assessment knowledge levels that these pre-service teachers have for themselves, is higher on every item than the principals' perceptions of the assessment knowledge levels of new teachers. Despite the fact that the pre-service teachers indicated that the coverage they received did not match the levels of importance noted by principals, they still had higher levels of confidence in their knowledge levels.

Additionally, the pre-service teachers in this study do not hold standardized assessments in the same regard as the principals. Their ratings indicate that they have reservations about the uses of the results in elementary classrooms, while principals indicated that standardized assessments and their results can play a useful role in elementary schools and classrooms.

Table 4.4. Research Questions Comparison of Means and Range

#	Question	Pre-service Teacher Average Rating	Principal Average Rating	Range
I	To what degree are pre-service teachers exposed to various assessment concepts and skills?	3.20	-	-
II	Among elementary school principals, what assessment practices are most important to know among beginning teachers working in classrooms?	-	4.09	-
III	How do elementary school principals' expectations for teacher assessment knowledge compare with self-reported pre-service teachers' levels of exposure?	3.20	4.09	0.89
IV	What are the self-reported levels of understanding among pre-service elementary teachers of various concepts and skills?	2.6	-	-
V	How do elementary school principals conceive of assessment knowledge for beginning classroom teachers? How does this correspond with the self-report of pre-service teachers?	2.6	1.9	0.7
VI	What are the attitudes of pre-service teachers toward the use of standardized achievement results in the classroom?	2.44	-	-
VII	What are the attitudes of elementary school principals toward the use of standardized achievement results in the classroom? How congruent are these with the attitudes of pre-service teachers?	2.44	3.04	0.60

## CHAPTER V

### RESULTS DISCUSSION

#### Introduction and Description of the Study

The use of assessments and assessment results is a critical component of teaching, and when done properly it adds to the quality of instruction students receive. As mentioned earlier, the definition provided in the *Standards for Teacher Competence in Educational Assessment of Students* (1990) states that assessment is "the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his or her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy," (1). This is not a skill that one typically picks up without support. When a new teacher enters the teaching profession he or she will likely spend up to 50% of their time assessing students. Without adequate knowledge of assessment principles and practices, a novice teacher will struggle.

The current K–12 educational environment puts a great deal of emphasis on standardized assessments and their results. *NCLB* and other school reforms have elevated the stakes for testing, putting students, teachers, and principals at increased risk about being judged against a single standardized measure. The national media have followed along these same lines, presenting test score data as the dependent variables for school success. Such assessments, of course, provide useful information, but many teachers do not know how to interpret or use the results and tend to have a negative reaction against them.

The purpose of this study was to analyze what pre-service elementary education students know about the use of assessment methods in the context of teacher education. The study examined the perceptions of pre-service teachers in relation to what they were exposed to and what they subsequently learned about assessment. The study also claimed to relate pre-service teachers' knowledge of assessment against what principals expected and desired from young teachers in the area of assessment knowledge. Additionally, pre-service teachers and principals provided information on the role of standardized tests in the elementary school setting. On-line self-report surveys were prepared and given to pre-service teachers and principals to obtain and analyze these perceptions. The results are discussed in this chapter.

### **Analysis**

Teaching students is complex and teachers make many interconnected decisions every day that impact how they teach. To ensure that these decisions support student learning, teachers must rely on data they obtain through a variety of assessments, both formative and summative. Formative assessment, or assessment for learning, is the type of assessment that occurs daily and allows teachers to make those 'in the moment' decisions to change how a lesson or concept is taught or provide extra support to a student who struggles with a particular concept. To do this though, one must have a firm understanding of the uses, types and benefits of the different forms of formative assessments (Black, Lee, Harrison, and Black 2004).

Common sense would dictate that this would be an easy issue to address. Simply ensure that teachers are prepared to understand and use assessments. This can be done in teacher education programs and professional development sessions. The problem is that

this, inexplicably, is one area that has not received much focus. Meanwhile high-stakes testing and accountability have been front and center as our society has focused on the overall results of annual tests rather than pay attention to all of the information that could be extrapolated from the classroom and school-level assessments. Teachers do not have a firm grasp of how to properly develop or use different assessments properly or use the results to help guide their instruction. And while a plethora of money has been spent on developing standardized testing systems, there has not been a similar effort to improve teachers' ability to effectively use assessments for learning (Stiggins 2002).

The *Standards for Teacher Competence in Educational Assessment of Students* (1990) articulated the areas of educational assessment that teachers should be competent in. The results of this study addressed these seven standards and show that the participating pre-service teachers do not feel prepared to implement the different components identified in these standards. Additionally, the principals that participated had an even more negative perspective on how well prepared new teachers are in the area of assessment. Unfortunately, this is not an unusual trend or even one that has just recently been recognized. Noll, Thorndike, and Hagan (1955), Conant (1963), and Mayo (1967) all identified this as an area to focus on and as an area of concern (Wise 1993). More recent studies have also consistently found that teachers typically have low levels of assessment knowledge (Gullickson and Hopkins 1987; Brookhart 2001; Campbell, Murphy, and Holt 2002; Stiggins 2001; as cited by Mertler 2003).

### **Assessment Exposure**

Questions I, II, and III addressed assessment exposure through five categories that included:



- creating and using different assessment methods,
- understanding the strengths and limitations of different assessment types,
- learning how to construct or evaluate test items,
- exposure to the interpretation of different types of standardized score results, and;
- exposure to key assessment results.

#### Category #1 – Creating and using different assessment methods

Pre-service teachers and principals responded to statements about exposure to and the importance of the creation and use of different assessment methods. Assessment topics covered in the category included many that teachers use, often on a daily basis. They were all assessments that can be used as formative or summative (or as a component of those), and should be used in support of each other. Standardized test scores will be used, at a minimum, on an annual basis, while the others such as performance assessments, constructed response, and portfolios can be used throughout the instructional experience. The exposure to a variety of assessment types is important for teachers – particularly new teachers. Stiggins and DuFour stressed that the different types of assessment play unique and critical roles. Teachers should be prepared to understand and use the results of assessments at the classroom, school, and institutional level. The combination of the three information sources is much more effective and has a greater impact when they are used in conjunction. To focus one's (or a school, district, or nation's) attention on only one of these assessment types, will dilute the value of all (2009).

The pre-service teachers indicated that they covered most of these topics during their teacher education coursework (overall mean of 3.20). However, there is variation in the amount of coverage when looking at the individual topics. While the amount of exposure to the use of observational techniques and rubrics (3.91 and 3.84) were high, on

the other end of the spectrum creating and using constructed response assessments (2.78) and the use of standardized test scores (3.02) were much lower. These results are fairly consistent with earlier studies where teachers rated the need for competencies related to classroom observation and grading high while the need for essay writing was considerably lower (Wise 1993).

The ratings of the participating principals (average rating of 4.09) indicate that they felt it is very important for new teachers to have been well versed on how to use a variety of assessment methods. Each of the topics covered had a mean over four, with the lowest being the use of portfolios (average rating of 4.05). The topics rated as the most important by the principals were the use different assessment methods/approaches (4.6) and the use different assessment results (4.73). Each of the other topics was a specific assessment method, indicating that the participating principals prefer teachers use a variety of assessment methods and results to help guide instruction - which research, such as by Stiggins and DuFour (2009), shows to be effective. Of the specific assessment methods that the principals in the study found useful were the use of performance assessments and the use of rubrics. These topics can be seen as related as rubrics are often used as criteria for the performance assessment, though rubrics can also be used with a variety of other assessment methods, including constructed response, as well.

The principals' ratings were higher than the pre-service teachers' respective ratings for each item covered in the questions (overall means of 4.09 and 3.20). The largest differential (difference of 1.36) occurred with learning how to use performance assessments to gauge student progress. Performance assessment, though it has long been

a part of education and assessment, is now receiving a considerable amount of attention due to different factors that are coalescing. One is on-going research that is focused on how individuals read, think, and learn brings out the complexity of learning and the areas of related student growth and achievement. Next, new standards such as the Common Core State Standards (CCSS) require assessments that allow teachers to capture complex student achievements. Third, is an ever-increasing sense that traditional tests do not give a full account of how much students have learned and how they are able to use what they have learned. And, performance assessments, when used properly, can be integrated more directly into the teaching and learning process (Afflerbach 2012). Rubrics are an essential component of performance assessments. Performance assessments should be designed in relation to a logical rubric which will then serve three purposes:

- 1) identify the criteria to evaluate student work,
- 2) specify the differential quality of students work, and;
- 3) consistently and accurately score student work, (Afflerbach 2012).

The recognition by principals of the importance of performance assessments and rubrics is important, but the fact that the pre-service teachers in this study rated performance assessments as having relatively lower exposure is a concern. Not only are performance assessments an important tool, but as Popham has noted, they are difficult to develop when done properly. A lot of time and thought will have to go into their development and the creation or selection of rubrics that will allow for reliable scoring of the assessments (Afflerbach 2012).

Pre-service teachers also rated the use of constructed response assessments as low. The use of constructed response assessments was the item with the next largest differential in ratings of importance versus coverage (1.30). While the difference was

fairly substantial, there was consistency in the ratings as it was the lowest rated item for pre-service teachers and the second lowest item (only 0.03 higher than use of portfolios) for principals. At the same time both report fairly high levels of exposure to rubrics – a key aspect to performance assessments and constructed response items. This leaves an interesting question of why the exposure to the use of rubrics is so high, yet the assessment types that one would typically associate with the use a rubric are relatively low.

Overall, principals placed a high-level of importance for the topics covered in this category, and not just one or two of them, but all of them. The pre-service teachers on the other hand, reported levels of exposure that did not align with the principals' ratings. This provides a level of concern because, as noted earlier, assessment plays such a critical role in the teaching and learning process. If pre-service teachers are not being prepared to enter their career with enough exposure of assessments and their uses, then this will be something that will have to be learned on the job – a job that will already challenge them on many different levels. And the fact that these principals recognize the importance of these topics highlights the concern that these teachers would have benefited from additional exposure to these topics.

#### Category #2 – Understanding the strengths and limitations of different assessment types

The category explores the understanding of the difference between formative and summative assessments and understanding the strengths and limitations the results obtained from performance assessments, teacher-made tests, and standardized tests. The perception of participants of this study is that of the five categories in Question I, the

coverage of the strengths and limitations of different assessment types was the highest (3.7).

Summative and formative assessments serve different, but critical roles. Whereas summative assessments are used to provide information on what a student knows or can do at a set point, formative assessments are used to provide a continuous flow of evidence that will allow teachers to identify potential learning needs and make necessary adjustments to instruction (Wiliam 2006). Summative assessments will provide teachers with a final analysis of how much a student (or students) ended up being able to show they have learned at the end of a unit of study, program, or academic year. Taken together the use of results from formative and summative assessments provides a better overall experience for students by allowing the teacher to have the information they need to properly prepare the students (Shavelson, et al, 2008).

There was agreement between pre-service teachers and principals on understanding the differences in the uses of formative and summative assessment on the reported exposure and importance levels (average responses were 4.07 – 4.63). For teachers to enter their career able to understand the difference between them is important, though it is important that they not only know the difference, but are able to use the different assessments appropriately to help guide their instruction.

Understanding the strengths and limitations of results from standardized assessments (average response was 3.98) is another important understanding for teachers understand. As noted earlier in this research, standardized assessments play a critical role in our school system today, yet there is also a tremendous amount of misunderstanding in relation to the information that their results provide. For teachers to truly understand the

strengths and limitations of standardized assessment results could make a tremendous impact upon education. It is important that new teachers enter the profession with an understanding of how they can and should use these assessment results. When the high-stakes concerns take over educators' ideas of how to use the results it can lead to poor teaching practices such as drilling students with practice tests, narrowing the curriculum, and a focus of teaching only what is known to be on the test (Gunzenhauser 2003). Despite the responses indicating that the strengths and limitations of standardized assessments are covered, the same pre-service teachers indicated that actually using these scores was covered at a much lower level (respective average ratings of 3.98 – 3.02). This is a distinction that would require a closer look and could indicate some bias against standardized assessments and the uses of their results, or perhaps pre-service teachers are not fully grasping the place of these assessments in the educational landscape.

Similar to the responses with standardized assessment results, a similar pattern occurs with the exposure to the strengths and limitations of results from performance assessment compared with how to actually use this type of assessment (respective average ratings of 3.64 - 3.13). It is important that teachers are able to understand the strengths and limitations associated with the use of the different assessments methods and types, however, it would be a concern if that was a majority of the coverage given to the topics rather than a greater focus on how one would actually use the assessments.

Teacher-created tests are critical component of a teacher's repertoire, and research has shown that well-developed classroom tests will improve students' academic achievement (Mertler and Campbell 2005). Additionally it has been shown that teachers rely heavily on teacher-made tests to inform the instructional decisions in their

classroom, yet there was not much focus spent on this topic (especially compared to the time spent on statistical analysis and standardized assessments) in teacher education programs (Mertler 2003). The pre-service teachers of this study rated this as the least exposed topic in this category (average rating of 3.11). The lack of time spent on a topic that plays such a prominent role in teaching is a concern. Students will be assessed by and have decisions made based on a type of assessment that their teachers have had little formal training with.

### Category #3: Learning how to construct and evaluate test items

Teachers need to understand the types of assessments available and how to use the results they gather from them, but it is also imperative that the teachers will be able to ensure that the assessments they do choose are worthwhile. Teachers need to have a reliable stream of information about their students in order to effectively tailor their instruction (Heritage 2008). Without reliable assessments, the information received by the teachers will not paint an accurate image of what the students actually know or can do. Noted earlier, but important to know, is that a positive correlation exists between the quality of classroom assessments and academic achievement (Mertler and Campbell 2005). And unfortunately, there is evidence that classroom assessments often do not meet the standards of reliability and validity expected of other assessments, and that the assessments used by teachers are often low quality (Stiggins 2001). Further, other studies found that teachers had little concern for question structure quality and analyses of their self-constructed tests revealed frequent violations of common question writing guidelines (Wise 1993).

Given the importance of this topic, and the fact that teachers spend such a high percentage of their time assessing students, formally or informally, the idea that the pre-service teachers in the study report a low level of exposure to the understanding of whether the items used to assess students are appropriate or not is disconcerting. And while research has shown that pre-service teachers typically are not high on the assessment components that rely upon statistics and statistical analysis, it is important that teachers are able to comprehend whether or not the assessment items (and the assessments themselves) are reliable or the results are used in a valid manner (Mertler 2003).

The survey results show that this topic is not considered as important as the other topics in this study for pre-service teachers or principals (respective average ratings of 2.8 and 3.85). Both populations indicate the most exposure came with the construction of test items and the lowest was the evaluation of the quality of publisher-created tests. The pre-service teachers' ratings were all under '3' or 'covered'. This indicates that this population of aspiring teachers will likely enter the classroom with little exposure on how to either construct or evaluate the assessment items that they will utilize.

#### Category #4 – Exposure to the interpretation of standardized score results

Teachers should understand the value, purposes, and meanings of standardized test results. This is not because they will tell you all that you need to know about your students, or because some believe that they provide the best way to measure academic achievement. Even experts in the measurement field share that the results of these assessments must be used with caution and that they are not designed for certain uses (grade promotion or instructional diagnosis) (Gunzenhauser 2003). The reality is that



standardized tests are prominent in our public education system and will continue to be so in the foreseeable future. Pre-service teachers in this study did not feel that they had a lot of exposure to learning how to interpret the standardized test results using the different techniques (average response was 2.8). There was only a difference of 0.31 between the item with the highest perceived amount of exposure, use of percentile ranks (3.0) and the lowest, interpreting results for parents (2.69). The results of the items in this category remain constant with the reported levels of exposure to the use of standardized test scores from Category #1 (3.02). The interpretation of standardized score results is important and should not be overlooked. Standardized assessments are prevalent throughout the educational landscape, with some of the results having a consequential impact on students, teachers, and school districts. The results provide useful information to teachers and when used properly, a good supplement to other assessment results collected.

The one topic in this category with the largest differential and that indicated misalignment between the pre-service teachers and the principals is the exposure to be able to interpret these results to parents in meaningful manner. This item was the least exposed for pre-service teachers (2.69) but the highest rated for principals (4.39). Regardless of the concerns that schools of education may have towards standardized assessments and the uses of the results, there is a need for teachers to fully understand the benefits and concerns associated. Unless teachers are able to effectively communicate to parents and the public in general what the results mean, the opportunity to help the public to understand the pros and cons will continue to rest with policy makers and the media. Add to that the results of these assessments do provide some meaningful snapshots of student achievement, teachers need to be able to accurately interpret the results to help

identify (for teachers and parents) areas where students are successful and where they may need some additional support.

Given that the use of standardized assessment results is so prevalent in our public education system, the results to these items seem to align with the need for teachers to understand how to interpret those results. Principals will feel the pressures and impacts tied to the results more acutely than most other stakeholders. School report cards, public media, and even school closings all focus on these assessment results, and typically in a negative fashion. The responses provided by the principals add insight from this population that standardized test results are important – whether that is because of the information they provide or because it is important for teachers to understand the tests and results that do have such a deep impact upon them.

#### Category #5 – Exposure to key assessment terms

The results of the respective surveys in this study showed that the population of pre-service teachers in this study had more exposure to the difference between norm- and criterion-referenced tests than for learning about either validity or reliability. Similarly, the principals also indicated that it was more important for teachers to be exposed to learning the difference between norm- and criterion-referenced tests. This may be an indication of the overall concerns of assessment knowledge that exists not only with teachers, but with principals as well. Assessment results that are not valid or reliable, whether they are from norm- or criterion-referenced assessments, will not provide meaningful information. Educators have to understand what they are assessing, why, and what instrument will provide the best information. Understanding how reliability and

validity work, will allow that process to happen and without that one cannot truly know what the assessment results mean.

How teachers develop, implement, and use assessments (and their results) has a deep impact on students' learning. It is the teachers' responsibility to ensure they are doing this process in a reliable and valid manner. Issues pertaining to these concepts have existed for over a century. Giordano provided details discussing the concerns that were raised over grading practices and how they were determined arbitrarily or based on personal factors (2005). Reliability and validity concerns continue to exist today with grading practices of teachers as research shows that grading decisions are still based on issues such as student behavior and effort (Remesal 2010). Research has shown that assessments used in the classroom, whether teacher- or publisher-created, are often times not reliable and the uses are not valid (Mertler 2001, Brookhart 2004). Further studies showed that teachers, when trying to ensure reliability and validity did not follow any set steps, and teachers in one study even said a test was reliable if it was developed by the teacher (Mertler 2003).

### **Assessment Knowledge**

Understanding the coverage provided on different assessment concepts and skills within the coursework of a teacher education program provides a view of what pre-service teachers have been exposed to, but it is also important to know if that translates to actual assessment knowledge. In this study the pre-service teachers identified their perceived knowledge levels on different assessment concepts and skills and principals provided their perceptions on how much understanding new teachers possessed about assessments.

Teachers should have a deep understanding of the uses, types, and benefits of different assessments, and be provided with the support on planning and implementing these processes. This support would come through either teacher education programs or professional development opportunities (Black, Lee, Harrison, and Black 2004).

This coincides with past research, where teachers (already in the classroom) report they do not feel sufficiently prepared to assess students. Further, research has shown that teachers will attribute this lack of preparation and understanding to the pre-service training they received in regards to assessment. One state-wide study that asked teachers of their perceived level of preparedness to assess student learning resulting specifically from their pre-service teacher preparation program resulted in over 85 percent of the respondents stating they were not well prepared (Mertler 2009). Mertler completed another study (1999) that found that teachers leaving their teacher education program felt only 'slightly prepared' to assess student learning. Supporting that, Volante and Fazio completed a study (2007) that showed pre-service teachers reporting low levels of a confidence in their ability to properly assess students and that a large number of those students would have liked a course specifically designed to cover classroom assessment (DeLuca and Klinger 2010).

Similar to previous studies, the pre-service teachers that participated in this study are not comfortable with their knowledge levels of these topics (mean rating of 2.6 on a 4-point scale). And across the items here, the participating principals indicated they had even less confidence in new teachers' assessment knowledge and skill-base (average rating of 1.9). In regards to the perceived understanding of the general use of classroom assessments and procedures pre-service teachers and principals expressed reservations

(respective average ratings of 2.85 and 2.43). The responses for this item provide a more general view of overall assessment understanding in comparison to the other items which focus more on specific assessment knowledge. As the other items all combine to compromise a significant portion of classroom assessments it is interesting that both groups rated the general topic so much higher. Again, as noted earlier, this might be the effect of a topic not seeming so difficult until one begins to consider the actual processes involved.

The pre-service teachers reported high levels of understanding with their ability to grade student work (3.18) and even the principals (2.77) expressed some confidence that new teachers were capable of grading. The reported level of comfort with grading is important and is a process teachers practice every day. The question that can be brought up around this, despite the high rating, is what it is that teachers actually know about grading. Grading has been an issue that was identified as problematic in the early 20<sup>th</sup> Century and continues to be so today. Concerns expressed one hundred years ago are similar to the concerns that are identified of current grading practices, including teachers allowing grades to be impacted by arbitrary reasons including such things as student personalities and effort (Giordano 2005 and Remesal 2010). Grading is a something that everyone who has attended school has been around. However, it is not just something that one can do validly without understanding the reason they are grading an item or project in a particular way. One has to consider the objectives being measured, the degree to what those objectives need to be met, and then identify the best method to measure that. The concern I have is that there is a potential that aspiring teachers do not

think of the complexities involved with grading, but are focused on a more simplistic action of marking something wrong or right and using a percentage to assign the grade.

Principals also have very low confidence in new teachers' ability to interpret assessment results and use those results to inform or plan instruction (1.41). If this does not happen, then the whole point of using assessments is not being met. 96% of the principals responded that beginning teachers do not have sufficient knowledge with this, while 53% of the pre-service teachers indicated that they do have sufficient knowledge.

There were also large differences in the perceptions of beginning teachers' knowledge of their understanding of how to use data to interpret results (94% of principals say beginning teachers do not have sufficient knowledge compared to 52% of pre-service teachers saying that they do have sufficient knowledge), use performance assessment methods (93% of principals say beginning teachers do not have sufficient knowledge compared to 43% of pre-service teachers saying they have sufficient knowledge), and interpret results for students and parents (90% of principals say beginning teachers do not have sufficient knowledge compared with 59% of pre-service teachers saying they have sufficient knowledge). These are all critical components of assessing students and ideally would be used to support their academic achievement. Whether young teachers feel more confidence because of their overall lack of exposure to assessments and don't really understand what is really needed to be able to do it well, or they have not had an opportunity to apply their knowledge of assessment in a real, classroom context needs to be explored further.

### **Perceptions of Standardized Assessments**

Public school education is moving through an era where accountability to standardized test scores is at a premium. Conversations are too often centered on the state of these scores and the need for our students to score higher. Critical decisions such as student promotion, teacher evaluation ratings, school closures, and many others are based on these scores. This comes despite warnings from measurement experts and developers that those are not appropriate uses of the results (Gunzenhauser 2003). The pressure to rely so heavily on these results is not a recent phenomenon and has been developing since early in the 20<sup>th</sup> century. The introduction of standardized assessments was used by some to address issues with subjective grading (Giordano 2005), by others to allow for the differentiation of students and the academic tracks that they would follow (Shepard 2000 and Embretson 2003), as well as to serve as tools for determining college admissions (Giordano 2005). Speaking to the prevalence of the place of standardized assessments in the public education realm is how the list of fifteen standardized test applications identified by Symonds (1928) are all still relevant today. The ones that have been brought to the surface in our current accountability phase include promoting competition, determining quality of instruction, rating teachers, and studying the efficacy of the schools (Giordano 2005). Beginning with *A Nation at Risk* (1983) and President Reagan's administration call for increased school accountability, and then with the implementation of *NCLB*, a direct focus was placed on the standardized assessment scores and educators across the country have reacted to ensure student scores were high enough to keep them out of sanctions. And while *NCLB* was able to provide a focus on educational disparities and pushed teachers towards using a standards-based curriculum and analyzing assessment data (Center on Educational Policy 2008), there has also been a

lot of backlash and concerns brought forth. Significant was the complaint that the accountability measures tied to the legislation based on the assessment results led to a narrowing of the curriculum, particularly for schools serving low-income, minority students (Berliner 2009). The *NCLB* era is now being followed up by President Obama's educational reforms that focus on a set of nationally adopted standards (CCSS) and greater accountability for teachers and schools through new teacher evaluation methods, increased school choice options, and the development and implementation of new, nation-wide, standardized assessments (U.S. Department of Education 2010).

Since President Obama's policies have largely occurred after this survey was completed, the perceptions held by the study participants is largely in reaction to the use of standardized assessments from the time of *A Nation at Risk* through the introduction and implementation of *NCLB*. The fact that there are punitive actions to the results of the standardized assessments it is a challenge to disassociate that with the positive impacts. Another practical issue tied to the use of standardized assessment results is the timeliness with which teachers receive the results. In districts where results are not received until shortly before the schools break for the summer, there are not as many benefits – especially for the individual teacher. Despite the issues and concerns with standardized assessment results there are benefits when used properly. Teachers can use the results to identify broad areas where students need additional support and the teaching and curriculum can be focused on. The results can show trends that teachers can address if there is an issue, or build on if the trend is positive. The fact that the pre-service teachers in this study identified the interpretation and use of standardized assessment results as relatively low compared to other assessment categories speaks towards the general



negative feelings towards them. They are used in all schools, they are a source of information that is reliable, and they typically cover a range of topics and content aligned to state standards teachers should be addressing.

The results of this pre-service teacher survey, where a '3' indicates agreement and a '2' indicates disagreement, found an overall mean of 2.44. There was a range of 1.25 between the high and low items, which is large considering this is a 4-point scale. The highest results were for the item that focuses on how standardized assessments provide information to monitor yearly academic progress (3.18). This indicates that the pre-service teachers tend to agree that the standardized assessment results do provide some benefits and can help reflect the overall academic progress made by students. Given that the respondents noted that these assessment results allow teachers to monitor yearly academic progress, it is interesting to note that they did not feel as strongly that standardized assessments and their results serve useful purposes in elementary schools (2.81). Somewhat lower ratings, though seemingly not total disagreement, were given to the idea that these assessment results can help to identify students' academic strengths and weaknesses (2.59).

Even if the enhanced ability to monitor yearly academic progress and potentially identifying students' academic strengths and weaknesses were the only benefits to come from the results that would seemingly provide evidence of serving useful purposes in schools. This likely is where the perceptions of the standardized assessments and bias against them come out. Despite the idea that these results can and do support the teachers' overall assessment of students, the fact that these results are used by some (or many) as a wholesale picture of academic achievement (of individual students,

classrooms, schools, and districts) has many leery of relying on the results. The two lowest rated items seem to reflect this concern. The idea that standardized achievement assessment results can accurately reflect students' educational achievement (2.06) or provide evidence of teachers' effectiveness (1.93), did not find support from this population. This could be from that concern that the results are going to be used for punitive purposes rather than serving as a tool to help teachers either move individual students forward or provide them an opportunity to reflect on their own teaching and know specific areas they need to improve on.

Principals' responses to this survey indicate that they believe standardized achievement results can play a useful role in elementary schools and classrooms. This was indicated in the first item in the question (3.34 rating), where over 98% of the respondents agreed they can serve a useful purpose. This response supports the general notion that standardized assessments can and do play a useful role in education when used properly. Looking closer at the specific topics, principals indicated that they also agreed that they provide useful information to monitor students' yearly academic progress and help identify students' academic strengths and weaknesses (3.21 and 3). A majority of the principals agreed with each item in this question, with the least amount of agreement associated with the idea that these assessments accurately reflect students' educational achievement and that they can serve as an important source of evidence of teachers' effectiveness in the classroom (72% agree or strongly agree).

Principals expressed a greater degree of agreement with the statements made about standardized assessments and their usefulness in elementary schools than pre-service teachers. Principals had a higher rating for every topic covered in this question

and an overall average of 3.04 in comparison to pre-service teachers' average rating of 2.44. The most disagreement came, not surprisingly, with the item covering the idea that standardized assessment results provide an important source of evidence of teachers' effectiveness. This is an argument that is prevalent today as teacher evaluation and pay is increasingly connected to the student results of standardized assessments.

Standardized assessments are a topic that has, and will continue to be, the center of a lot of debate. Effective arguments can be made that support the use of these results of these assessments, yet as misuse continues to run rampant many arguments can also be made against them. In the larger scheme, when teachers, or in this case pre-service teachers, fear that these results are going to be used to judge either individual student or teacher effectiveness, there will be a bias from that population against them. The decisions made at the national, state, and sometimes district level, on how to use these results is something often beyond the control of teachers or teacher training institutions. Research shows that due to *NCLB* and other instances where high stakes are aligned to testing, that teachers changed their teaching practices. They focused more on rote skills and facts and less time being creative and allowing students to use and develop higher order thinking skills (Center of Education Policy, 3).

However, it is important that teachers are should not be afraid to use these results to help inform their instruction. These are assessments that provide a meaningful supplement to the overall picture of student achievement, and when used properly can make a positive impact upon teaching and learning. Teacher training programs provide an opportunity for future teachers to learn the true purposes, benefits, and potential concerns that are tied to standardized assessments before they begin their career.

## Implications

Assessment is a vital part of the educational process. However, the integration of the use of assessment processes into teaching and learning is a practice that requires additional focus as the teachers in our public school system are likely not properly prepared in this area. A significant body of research shows that pre-service and practicing teachers are not always comfortable or capable of properly using assessments and assessment results in their instruction, (Mertler 1999, 1998; Plake 1993, as cited by Mertler 2003). Clearly, teacher education programs should be looking to provide pre-service teachers with as many of the skill sets and as much of a base of knowledge as possible.

The need for teachers to understand assessments seems ever more critical today. Although we are currently getting ready to exit out of the *NCLB* era, testing and accountability routines still have a strong hold in the schools today.

President Obama and his administration are in many ways building upon the foundations of the policies of *NCLB* and moving federal policies deeper into the public school systems. With the implementation of the Race to the Top (RTTT) program that was enacted in 2009, the federal government used a ‘carrot’ of allowing states in financial straits the opportunity to tap into a \$4.3 billion bucket of money to enhance their public educational systems (Ravitch 2010). To access this funding stream, states had to apply for federal grants by agreeing to potentially increase the number of charter schools, begin to evaluate teachers in relation to student test scores, turn-around low-performing schools, and adopt the CCSS in mathematics and English language arts – in essence emphasizing test-based accountability, merit pay, and school choice (2010).

The inclusion of the CCSS is an important distinction from *NCLB*. This is a way to end the practice of states simply lowering student expectations in order to ensure more students score higher on the exams, a practice that states did engage in. The CCSS now provide states that have adopted them with a uniform set of instructional objectives that students are expected to master (Manna and Ryan 2010). The implementation of the CCSS are aimed at ensuring more high order thinking skills will be assessed, and that there will be reliance upon performance-based assessments and provide a richer experience with authentic assessments. The PARCC and SBAC assessment consortia will be developing standardized assessments to measure student achievement in mathematics and English language arts (Afflerbach 2012). Part of the implementation of the CCSS is the assessments that are being developed by PARCC and SBAC to measure student growth. This is significant as the ways these assessments are being developed are going to be different than the typical standardized assessment in place now. The change is that there will now be a greater focus on using performance assessments to measure what students know and can do (2012). This is a major shift in how students will be assessed. First, is the change in the make-up of the standardized assessments, and then the impact of how that will translate into the assessments occurring within classrooms.

Another aspect of the Obama administration's initiatives is the push to have teacher evaluation tied to students' performance on assessments. If there is a concern that teachers 'teach to the test' now, then one will have to wonder as this initiative becomes more commonplace. When a teacher's evaluation is tied directly to one particular test, it will only be natural for the teacher to try and figure out the best way to get the student to achieve. One would hope that teachers would determine the best

method would be to align their instruction closely to the standards being assessed and use a variety of assessment methods to monitor student progress and inform learning.

However, this is not going to be a reality for many teachers. There will be efforts made to ‘game’ the system, attempts to teach to the test, or a narrowing of the curriculum (Ravitch 2011). It is possible that districts can develop an evaluation method that will lessen the likelihood of this happening, but it would take a tremendous amount of coordination, cooperation, and planning – something that educational bureaucracies do not necessarily excel at.

The Chicago Public Schools and the state of Illinois are preparing to evaluate teachers in part on student growth. CPS will use student achievement from a standardized assessment and from district-developed (curriculum and assessment offices and teachers) performance assessments that will be used to show student growth. Our offices are currently developing pre- and post-performance assessments based on the CCSS. While the idea sounds right conceptually, there are concerns in how the process will be implemented. As the past research and the results of this study show, teachers do not have a lot of experience working with performance assessments or rubrics, and issues have been raised about grading for over a century. One concern is that the teachers will be required to use rubrics to score or grade student work. Considering that all teachers of a common course and grade level will be using the same student assessment and scoring tool and making judgments that will potentially impact their careers, it is critical that all teachers can do this in an informed manner. Training will be provided and ideally instructional leaders in the individual schools will provide the necessary support, but in reality there will be a lot of teachers scoring these high-stakes assessments in an

uninformed manner. Additionally, if this process is to work as conceptualized then it will also be incumbent upon all teachers and school principals to ensure that the process is completed in a fair and equitable manner. Issues may arise if the assessments are not scored properly and if the school administrators do not provide the proper oversight (whether that is due to their own lack of assessment knowledge, blindly trusting their teachers, or not taking the time to properly review the scoring process).

It will be very important that school districts that do take this type of teacher evaluation process on carefully think through the best methods of gathering the student data that will be used to evaluate the students. If they rely on standardized assessments such as those relied on in the implementation of NCLB then there should be concern that students will have a much narrower educational experience and will spend a majority of their learning time preparing to take these tests. The unfortunate part of that type of situation is that these assessment results do not really help the students learn, but are really used to help adults (e.g., politicians, educators, and publishers) make arguments about why others should listen to them. And for the districts that are going to follow the guidance associated with the CCSS and identify or develop assessments that are performance-based, they need to take care with that process as well and ensure that the assessments are reliable and the results used in a valid fashion.

Adding to the critical need for teachers to be able to grasp and succeed with the new focus on the CCSS and performance assessments is that another component of the RTTT is that schools that do not meet certain criteria will be restructured (closed and reopened either as a charter school or with a new organizational structure). One of the major complaints of *NCLB* was the punitive nature that many associated with it, and now

that aspect may even intensify. States and districts that are ‘fortunate’ enough to win one of the RTTT grants will have to have a plan in place to restructure schools that are not succeeding. Once again, there will need to be an immediate response from schools to try and figure out the best way to avoid being restructured. The accountability focus that has been growing over the past few decades is going to continue to move forward under the current school reform movement. And while the outcome of wanting all schools to be successful is ideal, to use punitive measures such as closing a school without allowing the needed time for teachers and administrators to adjust, will only create more issues. And further, this could potentially push more highly qualified and skilled teachers away from the profession (Ravitch 2011).

As districts move to implement the RTTT initiative and adopt its accountability components, curriculum will need to better align to the CCSS, performance assessments will become more prevalent, and teachers will need to be prepared to implement this type of assessment in their teaching. This is not a bad thing; in fact, I appreciate the fact that performance assessments are going to have a greater role in teaching. However, many teachers do not have the knowledge or skill sets to use performance assessments, let alone develop them or select ones that are appropriate for their classrooms. And, beyond that, to design and use performance assessments properly is a complex and time-consuming task.

Performance assessments are not used to understand students’ literal comprehension of facts, but capture what students can do by incorporating skills, strategies, and domain learning. For example, a performance task in history could move the focus beyond measuring a student’s knowledge of dates, names, and facts and instead



focus on why history is written, who writes it, how it is written, and how it is read (Afflerbach 2012). To do this, teachers must be able to unpack the standards that are to be learned, identify the tasks to ensure that this learning occurs, develop, implement, and score assessments to check for learning, and then use these results to inform further instruction. The CCSS are fundamentally different from other state standards that have been used to guide teaching over the past decades. The CCSS move from being a collection of pieces of knowledge and facts that are to be covered to a deeper focus on the skills that students must develop to be able to acquire that knowledge. In English Language Arts – History/Social Studies the expectations have been made that students will be able to analyze primary and secondary sources, cite textual evidence to support arguments, consider the influence of author’s perspective, collaborate different sources, and develop written historical arguments (Breakstone, Smith, and Wineburg 2012).

Essentially, teachers are now being asked (or told) to begin using standards that are more complex and challenging to understand, to use assessments methods which many do not have experience with, and to use data from those assessments to guide instruction based on the new standards. Again, I applaud this move and believe that as teachers learn how to use the tools and processes that form the foundation of this work, education will move forward. However, the concern is that the data from past research and this study shows that teachers are not prepared to make these adjustments. Teacher education programs have not done enough to integrate the knowledge or skill sets needed to understand or use assessments within the teaching and learning process and need to adjust to provide this support.

Historically, teachers in the public school system have not been adequately prepared to use assessments in their practice. Issues have included subjective grading of students (from the early 20<sup>th</sup> century through today), misuses of standardized assessments such as labeling and tracking students, allowing accountability measures to not improve our educational system but rather narrow the curricular experience, and seemingly constant public complaints about the education provided. While many of these issues were brought about by political or personal ideologies or simple bureaucracy, there have been issues that need to be addressed. One such example is the ability of teachers to use the available assessment practices to improve teaching. There is ample evidence that shows teachers are not prepared to assess students. Part of that issue can be tied to the ways in which pre-service teachers are prepared to use assessments in teacher education programs. And while much of the focus of the measurement community has focused on the development and use of standardized assessments, there has been recognition that classroom assessment is an issue.

The *Standards for Teacher Competence in Educational Assessment of Students* was developed to guide the training of pre- and in-service teachers. These standards reiterated the important role of student assessments in the educational process. The development of seven guiding standards focused the need for teacher knowledge to include choosing and developing assessment methods; administering, scoring, and interpreting the results; using the results to make instructional decisions; developing valid pupil grading procedures, communicating results to students and parents; and, being aware of inappropriate uses of assessment methods (1990). Teacher preparation programs have not typically prepared aspiring teachers to be prepared to assess students.

In programs where students are required to take a specific assessment course or where a sub-section of a course is focused on the topic, the main topic covered is typically statistical manipulations and standardized assessments. These topics have their place in education and educators should be aware of them. However, teachers and principals have reported that too much focus on these topics is not relevant to them, and that they would prefer to learn more practical information and ways of applying assessment procedures to instruction (Gullickson and Hopkins 1987, Gullickson 1986, Marso and Pigge 1998, Brookhart 2001).

Further, research has shown that pre-service teachers do not have much knowledge on assessments. Surveys, tests, and reviews of products have informed researchers that issues existed with abilities to grade, develop assessments and assessment items, use performance assessments, determine or understand reliability, or using results to make instructional decisions (Gullickson and Hopkins 1987, Brookhart 2001, Campbell, Murphy, and Holt 2002, Stiggins 2001; as cited by Mertler 2003; Campbell and Mertler 2005, Volante and Fazio 2007).

This study found that pre-service teachers had little faith in their abilities to use assessments. The only area that they indicated that they had enough knowledge was with grading. And the principals that participated in the study showed deep concern for the levels of assessment knowledge that new teachers bring with them to the classroom. The lowest rating they provided was based on the ability of teachers to use assessment results to help plan instruction. This alone, should signal that our teachers need more preparation. But even topics such as writing sound test items, evaluating tests and test items, using performance assessments, using data to interpret results, interpreting results

for parents and students, and so on, are all areas that the study found as areas of concern. What purpose do assessments serve if the attained results are not used to inform instructional decisions? And if teachers are not able to properly develop or evaluate selected assessments, then what is being used?

Teacher education programs need to address the assessment process throughout the curriculum. It is important that the pre-service teachers are aware of the different types of assessments and the roles they each play, to have practice at developing assessments and analyzing assessment results, to investigate how assessment results can guide instructional decisions, as well as be prepared to communicate the purpose of the assessments and the meaning of the results with parents and students. And this focus on assessment should not occur separate from the study on pedagogy, curriculum, or cognitive development; but rather in tandem (DeLuca and Klinger 2010). If assessment results will actually be used to make instructional decisions, then it is crucial that pre-service teachers are able to engage in the processes together. And today that need to blend the different components into one has grown exponentially. Literally, the existence of a school or a teacher's job is dependent upon their ability to move students' academic achievement forward. And with the new educational reforms moving quickly forward, teachers will need to be able to understand how develop, use, and interpret performance assessments to allow students to successfully perform. And these are skill sets that teachers should have and utilize. However, the stakes tied to the implementation of these has risen and need to be addressed now. School districts, administrators, and teachers are scrambling to address this situation, but if new teachers are not prepared prior to entering the profession, this will become a task that will be hard to address. Teacher education

programs have the opportunity to prepare young teachers with a solid background of diverse assessment methods and an understanding of how those will apply to their teaching once they enter the classroom.

### **Further Research**

The study did provide a look at the relationship between how a single elementary teacher education program prepares young teachers about and how to use assessments compared with the views of elementary school principals. But there are ways in which additional investigation would provide a more thorough analysis and better insight into what the new and aspiring teachers actually know.

One of the aspects I would be interested in understanding better is what is the actual understanding of the different assessment practices that the pre-service teachers hold? It is one thing to provide self-perceptions on knowledge levels that one holds, but how does that correlate with actual knowledge? Providing the teachers with actual assessments of their understanding would allow for a direct comparison of actual versus perceived assessment knowledge. Interviews with the pre-service teachers to follow-up on the responses they provided on topics such as what is a performance assessment? What is validity? What are strengths and weaknesses associated with standardized assessment results? By comparing the responses given in the survey with focused discussions would allow the results to provide more meaning to how prepared the pre-service teachers are to use the assessment process. Further, by adding the qualitative aspect to the study, it would allow an opportunity for the pre-service teachers to share the methods that were most effective for them to learn about the assessment process, and what methods did not work. Also, especially considering the 'new' era of public

education we are entering, it would be important to understand how comfortable teachers are in using standards (CCSS) guide the development and selection of assessments that will provide evidence on what students know and can do.

Expanding the population of participants in the study would provide a more meaningful analysis. Focusing on one teacher education program provided one perspective on how well pre-service teachers are prepared to assess students, but to add additional programs would allow comparisons that could include types of institutions (i.e., a large, public university compared to a small, liberal arts college), institutions across states (including different regions of the country), as well as undergraduate compared to graduate programs.

In regards to principals, an inclusion of subjects from additional states (and regions of the country), a larger sample of schools from urban and suburban settings, as well as including principals from charter and private school settings would allow for comparisons between the expectations in different settings. And a concern that exists is that not all principals are competent with the assessment process. The study would be stronger if there were steps taken to assess the content knowledge of the principals and to complete an analysis on what assessment topics are covered in the programs they move through on their way to certification.

The pre-service teachers identified their perceptions of exposure to assessment topics, but it would be useful to also get the perspectives of the respective instructors the students took their courses with. This would allow an analysis of what the instructor feels that they cover compared to what the pre-service teacher felt was covered. By looking at the syllabi, assignments, and readings completed in the courses, one could

have a very solid understanding of the actual levels of exposure that the pre-service teachers had towards the different assessment topics. Additionally, what type of experience or background do the teacher educators have with assessment? Do teacher educators with a stronger assessment background do a better job of incorporating the topic into the curriculum?

Including practicing teachers in the study would provide additional benefits. Ideally, I would be able to follow-up with the pre-service teacher participants as they moved through their career. This would allow an opportunity to explore how their perspectives might change as their situation moves from hypothetical to practical. It would allow one to see if the perceptions of exposure levels or knowledge would change as they begin to apply what they have learned to actual practice. Additionally, how the setting one enters impacts those perceptions. And then it would provide an opportunity to see how much support teachers receive in regards to assessment use once they begin teaching. An analysis of the content of the professional development teachers are exposed to, and the implementation of surveys or interviews that address the perceptions of amounts of exposure in professional development sessions could be compared to the respective teacher education programs, and allow for an opportunity to see which supports are more effective and useful for the teachers. And finally, a way to truly understand what assessment practices teachers were able to learn would be to do an analysis of the assessment procedures that they actually apply once they begin teaching.

Teachers' abilities to effectively apply assessment processes into their teaching are important. The informed decisions that this allows can provide teachers with the knowledge that they will need to be able to successfully teach the students in their

classroom. Teachers that do not understand the assessment process will essentially be moving through the teaching and learning process blindly, and potentially providing more harm to the students than good. Teachers enter the profession to work with students and to help them grow socially, emotionally, and academically. But to be able to do that, these young teachers must enter their careers with the proper training and knowledge base that will allow them to succeed. It is incumbent on teacher education programs to prepare these teachers to understand the entire teaching process, and how the different components interact and feed off of each other. A teacher needs to be strong in their understanding of content, pedagogy, students' social and emotional needs, and assessments – not in isolation, but all working together.



## APPENDIX A

## INVITATION FOR PRE-SERVICE TEACHERS TO COMPLETE SURVEY

(SENT ELECTRONICALLY)

Project Title: **Pre-Service Teachers and Assessment**Research Team: **Martin Moe, BA, MA**

We would like to invite you to participate in a research study being conducted by investigators from The University of Iowa. The purpose of the study is to determine what types of assessment knowledge elementary education students are being exposed to during their time in the education program. This study is looking to identify which areas of assessment have or have not been introduced to elementary education students. This information will be useful in providing teacher education programs with the perspective of the students taking their courses of what is or is not being taught in the courses which make up the program. We are inviting you to participate in this research study because you are currently completing your student teaching as an elementary teacher and will possibly be entering a career as a teacher in the near future.

If you agree to participate, we would like you to complete a survey which will ask you to rate a series of items on a scale which will indicate the level of your agreement or disagreement with the statement. This survey can be accessed through a link sent to you. You are free to skip any questions that you prefer not to answer. It will take approximately 20 minutes to complete the study procedures. Once you are finished you may then submit the survey.

To help protect your confidentiality, the survey results can only be accessed via a password protected site. If we write a report or article about this study or share the study data set with others, we will do so in such a way that you cannot be directly identified. Taking part in this research study is completely voluntary. If you do not wish to participate in this study, delete the email containing the link.

We encourage you to ask questions. If you have any questions about the research study itself, please contact: Martin Moe, 312-402-1256 or Dr. Peter Hlebowitsh, 335-5504. If you have questions about the rights of research subjects, please contact the Human Subjects Office, 300 College of Medicine Administration Building, The University of Iowa, Iowa City, IA 52242, (319) 335-6564, or e-mail [irb@uiowa.edu](mailto:irb@uiowa.edu).

Thank you very much for your consideration of this research study.

Sincerely,

Martin Moe

## APPENDIX B

## INVITATION FOR PRINCIPALS TO COMPLETE SURVEY

(SENT ELECTRONICALLY)

To Whom It May Concern:

We invite you to participate in a research study being conducted by investigators from The University of Iowa. The purpose of the study is to determine what types of assessment knowledge elementary education students are being exposed to during their time in education programs and how this correlates with the actual assessment knowledge needed by teachers in the classroom. This study is looking to identify which areas of assessment that have and have not been introduced to elementary education students. This information will be useful in providing teacher education programs with the perspective of administrators as to the assessment knowledge pre-service teachers need to be exposed to prior to entering the classroom.

If you agree to participate, we would like you to feel out a survey that is accessed via the link on this email. You are free to skip any questions that you prefer not to answer. It will take approximately ten to fifteen minutes to complete.

Taking part in this research study is completely voluntary. If you do not wish to participate in this study, please return the survey without answering any of the questions or send an e-mail to martin-moe@uiowa.edu stating such. If we do not hear from you in two weeks, we will send another copy of the survey to you.

If you have any questions about the study, please contact me at my e-mail address or at **312-402-1256**.

Thank you very much for your consideration of this research study.

Sincerely,

Martin S. Moe

## APPENDIX C

## ADMINISTRATOR SURVEY

## SECTION I: Importance of assessment knowledge

**How important is it for new teachers to understand the following:**

	Extremely Important – 5	Very Important – 4	Important – 3	Somewhat Important - 2	Not Important - 1
1.1) To understand how to use different assessment methods/ approaches for a variety of purposes.					
1.2) To understand which assessment methods are appropriate for assessing higher order thinking skills.					
1.3) To understand how to assess for non-cognitive outcomes, such as student attitudes and student values.					
1.4) To understand how to construct test items that assess intended achievement targets.					
1.5) To understand how to evaluate publisher-created tests available with textbook.					
1.6) To understand how to use data to judge whether individual test items are performing as expected.					
1.7) To understand the difference between the uses of formative and summative assessments.					
1.8) To understand the strengths and limitations of the results derived from performance assessments.					
1.9) To understand the strengths and limitations of the results derived from teacher-made tests.					
1.10) To understand the strengths and limitations of the results derived from standardized tests.					

1.11) To understand how to interpret the results of standardized tests using percentile ranks.					
1.12) To understand how to interpret the results of standardized tests using grade equivalents.					
1.13) To understand how to interpret the results of standardized tests using standard scores.					
1.14) To understand how to interpret the scores from standardized tests to parents in a meaningful manner.					
1.15) To understand what the term validity means in the assessment process.					
1.16) To understand what the term reliability means in the assessment process.					
1.17) To understand what the term standard error of measurement means in the assessment process.					
1.18) To understand the main differences between norm-referenced and criterion-referenced tests.					
1.19) To understand how to use standardized test scores to monitor student progress.					
1.20) To understand how to create and use constructed response assessments to monitor student progress.					
1.21) To understand how to use performance assessments to gauge student progress.					
1.22) To understand how to utilize portfolios to gauge student progress.					
1.23) To understand how to utilize observational techniques to gauge student progress.					

1.24) To understand how to use rubrics to gauge student progress.					
1.25) To understand how to use different assessment results to plan classroom instruction.					

## APPENDIX D

## ADMINISTRATOR SURVEY

SECTION II: Perceptions of standardized achievement assessments in elementary schools

**Please indicate your level of agreement with each of the following statements.**

	Strongly Agree - 4	Agree - 3	Disagree - 2	Strongly Disagree - 1
2.1) Standardized achievement assessments can serve useful purposes in elementary schools.				
2.2) Scores from standardized achievement assessment can accurately reflect students' educational achievement in elementary schools.				
2.3) Standardized achievement assessments provide an important source of evidence of teachers' effectiveness in the elementary classroom.				
2.4) Standardized achievement assessment can provide information that makes it possible to monitor year-to-year academic progress with elementary schools students.				
2.5) Standardized achievement assessment can help to identify many of each student's strengths and weaknesses in academic areas.				

## APPENDIX E

## ADMINISTRATOR SURVEY

SECTION III: Perceptions of new teachers' expertise in assessment when they enter the teaching profession.

**Please indicate the degree of your agreement with each statement below.**

	Strongly Agree - 4	Agree - 3	Disagree - 2	Strongly Disagree - 1
3.1) They have sufficient knowledge of classroom assessment methods and procedures.				
3.2) They need additional training in learning how to write sound test items.				
3.3) They need additional training in using data to interpret test results.				
3.4) They need additional help in learning how to use performance assessment methods.				
3.5) They have enough knowledge to grade student work accurately and fairly.				
3.6) They need help interpreting test results for students and parents.				
3.7) They need additional help in learning how to use assessment results to help plan instruction.				

## APPENDIX F

## PRE-SERVICE TEACHER SURVEY

## SECTION I: Instructional exposure to assessment knowledge

*Please indicate your opinion of how much you have been exposed to the following assessment topics during your pre-service education courses (5 being the MOST exposure, 1 the LEAST).*

	Covered In Depth (5)	-- (4)	Covered (3)	-- (2)	Never Covered (1)
1.1) Learning how to use different assessment methods/ approaches for a variety of purposes.					
1.2) Learning how to determine which methods are appropriate for assessing higher-order thinking skills.					
1.3) Learning how to assess for non- cognitive outcomes, such as student attitudes and values.					
1.4) Learning how to construct test items that assess intended achievement targets.					
1.5) Learning how to evaluate the quality of publisher-created tests available with textbooks.					
1.6) Learning how to use data to judge whether individual test items are performing as expected.					
1.7) Learning the difference between the uses of formative and summative assessments.					
1.8) Learning about the strengths and limitations of the results derived from performance assessments.					
1.9) Learning about the strengths and limitations of the results derived from teacher-made tests.					
1.10) Learning about the strengths and limitations of the results derived from standardized tests.					
1.11) Learning how to interpret the results of standardized tests using percentile ranks.					
1.12) Learning how to interpret the results of standardized tests using grade equivalents.					
1.13) Learning how to interpret the results of standardized tests using standard scores.					
1.14) Learning how to interpret the scores from standardized tests to parents in a meaningful					



manner.					
1.15) Learning what the term validity means in the assessment process.					
1.16) Learning what the term reliability means in the assessment process.					
1.17) Learning what the term standard error of measurement means in the assessment context.					
1.18) Learning the main difference between norm-referenced and criterion-referenced tests.					
1.19) Learning how to use standardized test scores to monitor student progress.					
1.20) Learning how to create and use constructed response assessments to monitor student progress.					
1.21) Learning how to use performance assessments to gauge student progress.					
1.22) Learning how to use portfolios to gauge student progress.					
1.23) Learning how to use observational techniques to gauge student progress.					
1.24) Learning how to use rubrics to gauge student progress.					
1.25) Learning how to use different assessment results to plan classroom instruction.					

## APPENDIX G

## PRE-SERVICE TEACHER SURVEY

## SECTION II: Perceptions of standardized achievement assessments in elementary schools

**Please indicate your degree of agreement with each statement.**

	Strongly Agree (4)	Agree (3)	Disagree (2)	Strongly Disagree (1)
2.1) Standardized achievement assessments can serve useful purposes in elementary schools.				
2.2) Scores from standardized achievement assessments can accurately reflect students' educational achievement in elementary schools.				
2.3) Standardized achievement assessments provide an important source of evidence of teachers' effectiveness in the elementary classroom.				
2.4) Standardized achievement assessments can provide information that makes it possible to monitor year-to-year academic progress with elementary school students.				
2.5) Standardized achievement assessments can help to identify many of each student's strengths and weaknesses in academic areas.				

## APPENDIX H

## PRE-SERVICE TEACHER SURVEY

## SECTION III: Perceptions of assessment knowledge

**Please indicate the degree of your agreement with each statement below.**

	Strongly Agree	Agree	Disagree	Strongly Disagree
3.1) I have enough knowledge of classroom assessment methods and procedures.				
3.2) I need additional help in learning how to write sound test items.				
3.3) I need additional help in using data to interpret test results.				
3.4) I need additional help in learning how to use performance assessment methods.				
3.5) I have enough knowledge to grade student work accurately and fairly.				
3.6) I need help interpreting test results for students and parents.				
3.7) I need additional help in learning how to use assessment results to help plan instruction.				

## APPENDIX I

## QUESTION I – CATEGORY #1

Table I1  
Item Response Comparison

Items ranked from highest to lowest (and average rating) Overall Mean Rating: 3.35	Percent of response for each rating (rounded) / N=55				
	5 Covered in-depth	4 -	3 Covered	2 -	1 Not Covered
(3.91) Use observational techniques	35%	27%	33%	6%	0%
(3.84) Use rubrics	29%	38%	22%	9%	2%
(3.74) Use different assessment results to plan instruction	35%	17%	39%	4%	4%
(3.63) Use different assessment methods/approaches	19%	37%	33%	11%	0%
(3.45) Use portfolios	15%	36%	33%	13%	4%
(3.15) Determine methods appropriate to assess higher order thinking skills	9%	29%	33%	26%	4%
(3.13) Use performance assessments	7%	26%	44%	20%	4%
(3.02) Use of standardized test scores	7%	29%	31%	24%	9%
(2.89) Assess for non-cognitive outcomes (student attitudes)	4%	24%	38%	27%	7%
(2.78) Create and use constructed response assessments	6%	19%	39%	22%	15%

## APPENDIX J

## QUESTION I – CATEGORY #2

Table J1  
Understanding the strengths and limitations of different assessment types, Item Response Comparison

Items ranked from highest to lowest average (and average rating) Overall Mean Rating: 3.7	Percent of response for each rating (rounded) / N=55				
	5 Covered in-depth	4 -	3 Covered	2 -	1 Not Covered
<b>1.7 (4.07)</b> Difference in uses of formative and summative assessments	44%	26%	24%	4%	2%
<b>1.10 (3.98)</b> Strengths and limitations of results from standardized tests	33%	40%	22%	4%	2%
<b>1.8 (3.64)</b> Strengths and limitations of results from performance assessments	21%	40%	25%	13%	2%
<b>1.9 (3.11)</b> Strengths and limitations of results from teacher-made tests	13%	22%	33%	29%	4%

## APPENDIX K

## QUESTION I – CATEGORY #3

Table K1

Constructing or evaluating test items, Item Response Comparison

	Percent of response for each rating (rounded) / N=55				
	5 Covered in-depth	4 -	3 Covered	2 -	1 Not Covered
<b>Items ranked from highest to lowest average (and average rating) Overall Mean Rating: 2.8</b>					
<b>1.4 (2.98)</b> Construct test items that assess intended achievement targets	6%	26%	37%	24%	7%
<b>1.6 (2.8)</b> Use data to judge whether individual test items are performing as expected	6%	15%	44%	24%	11%
<b>1.5 (2.61)</b> Evaluate the quality of publisher- created tests	6%	15%	30%	35%	15%

## APPENDIX L

## QUESTION I – CATEGORY #4

Table L1

Interpretation of different score results, Item Response Comparison

	Percent of response for each rating (rounded) / N=62				
	5 Extremely Important	4 -	3 Important	2 -	1 Not Important
<b>Item rank</b> <b>Mean Rating: 4.36</b>					
<b>1.25 (4.73)</b> Use different assessment results	77%	18%	5%	-	-
<b>1.1 (4.6)</b> Use different assessment methods/approaches	66%	27%	7%	-	-
<b>1.21 (4.49)</b> Use performance assessments	57%	34%	8%	-	-
<b>1.24 (4.48)</b> Use rubrics	58%	32%	10%	-	-
<b>1.23 (4.32)</b> Use observational techniques	42%	48%	10%	-	-
<b>1.2 (4.29)</b> Determine methods appropriate to assess high order thinking skills	39%	52%	10%	-	-
<b>1.19 (4.11)</b> Use of standardized test scores	42%	34%	19%	3%	2%
<b>1.20 (4.08)</b> Create and use constructed response assessments	39%	34%	24%	3%	-
<b>1.22 (4.05)</b> Use portfolios	39%	31%	27%	3%	-
<b>1.3 (3.6)</b> Assess for non-cognitive outcomes	11%	45%	36%	8%	-

## APPENDIX M

## QUESTION I – CATEGORY #5

Table M1

Exposure to key assessment terms, Item Response Comparison

Items ranked from highest to lowest average (and average rating) Overall Mean Rating: 2.8	Percent of response for each rating (rounded) / N=55				
	5 Covered in-depth	4 -	3 Covered	2 -	1 Not Covered
<b>1.11 (3)</b> Use of percentile ranks	11%	15%	47%	18%	9%
<b>1.12 (2.76)</b> Use of grade equivalents	9%	13%	36%	29%	13%
<b>1.13 (2.76)</b> Use of standard scores	9%	15%	33%	31%	13%
<b>1.14 (2.69)</b> Interpret results for parents	4%	24%	26%	30%	17%



## APPENDIX N

## QUESTION II – CATEGORY #1

Table N1  
Create and use different assessment methods

	Percent of response for each rating (rounded) / N=55				
	5 Covered in-depth	4 -	3 Covered	2 -	1 Not Covered
<b>Items ranked from highest to lowest average (and average rating) Overall Mean Rating: 3.34</b>					
<b>1.18 (3.82)</b> Difference between norm- and criterion-referenced tests	35%	29%	24%	4%	2%
<b>1.16 (3.33)</b> The meaning of the term reliability	16%	29%	31%	18%	6%
<b>1.15 (3.31)</b> The meaning of the term validity	16%	29%	29%	20%	6%
<b>1.17 (2.89)</b> The meaning of the term standard error of measurement	7%	22%	33%	29%	9%

## APPENDIX O

## QUESTION II – CATEGORY #2

Table O1

Understanding the strengths and limitations of different assessment types

	<b>Percent of response for each rating (rounded) / N=62</b>				
<b>Items ranked from highest to lowest average (and average rating)</b> <b>Overall Mean Rating: 4.35</b>	<b>5 Extremely Important</b>	<b>4 -</b>	<b>3 Important</b>	<b>2 -</b>	<b>1 Not Important</b>
<b>1.7 (4.63)</b> Difference in uses of formative and summative assessments	69%	24%	7%	-	-
<b>1.10 (4.31)</b> Strengths and limitations of results from standardized tests	51%	33%	13%	3%	-
<b>1.8 (4.26)</b> Strengths and limitations of results from performance assessments	36%	34%	22%	7%	-
<b>1.9 (4.18)</b> Strengths and limitations of results from teacher-made tests	44%	34%	18%	2%	2%

## APPENDIX P

## QUESTION II – CATEGORY #3

Table P1

Principals – Learning how to construct or evaluate test items

Items ranked from highest to lowest average (and average rating) Overall Mean Rating: 3.85	Percent of response for each rating (rounded) / N=62				
	5 Extremely Important	4 -	3 Important	2 -	1 Not Important
1.4 (4.11) Construct test items that assess intended achievement targets	50%	23%	18%	8%	2%
<b>1.6 (3.87)</b> Use data to judge whether individual test items are performing as expected	31%	39%	19%	10%	2%
<b>1.5 (3.57)</b> Evaluate the quality of publisher-created tests available with textbooks	21%	34%	30%	10%	5%

## APPENDIX Q

## QUESTION II – CATEGORY #4

Table Q1

The interpretation of different types of standardized score results, item response comparison

<b>Items ranked from highest to lowest average (and average rating) Overall Mean Rating: 4</b>	<b>5 Extremely Important</b>	<b>4 -</b>	<b>3 Important</b>	<b>2 -</b>	<b>1 Not Important</b>
<b>1.14 (4.39)</b> Interpret results for parents	57%	28%	12%	3%	-
<b>1.11 (4.03)</b> Percentile Ranks	39%	34%	19%	8%	-
<b>1.12 (3.92)</b> Grade equivalents	38%	26%	26%	10%	-
<b>1.13 (3.66)</b> Standard scores	27%	31%	27%	10%	5%

## APPENDIX R

## QUESTION II – CATEGORY #5

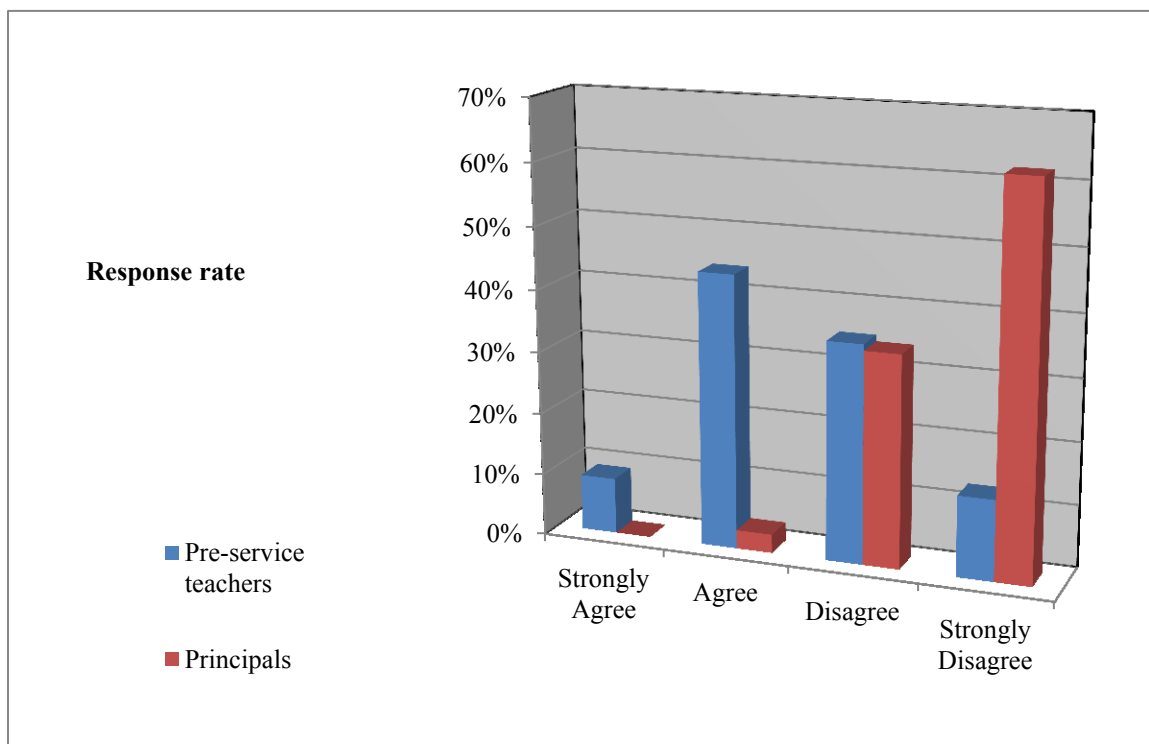
Table R1  
 Knowledge of key assessment terms, item response comparison

<b>Items rank</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>
<b>Average Rating: 3.88</b>	<b>Extremely Important</b>	<b>-</b>	<b>Important</b>	<b>-</b>	<b>Not Important</b>
<b>1.18 (4.28)</b> The difference between norm- and criterion-referenced tests	57%	18%	20%	5%	-
<b>1.16 (3.98)</b> The meaning of reliability	32%	42%	20%	5%	2%
<b>1.15 (3.92)</b> The meaning of validity	31%	39%	21%	7%	2%
<b>1.17 (3.35)</b> The meaning of standard error of measurement	18%	27%	32%	18%	5%

## APPENDIX S

## QUESTION V – ITEM 3.7

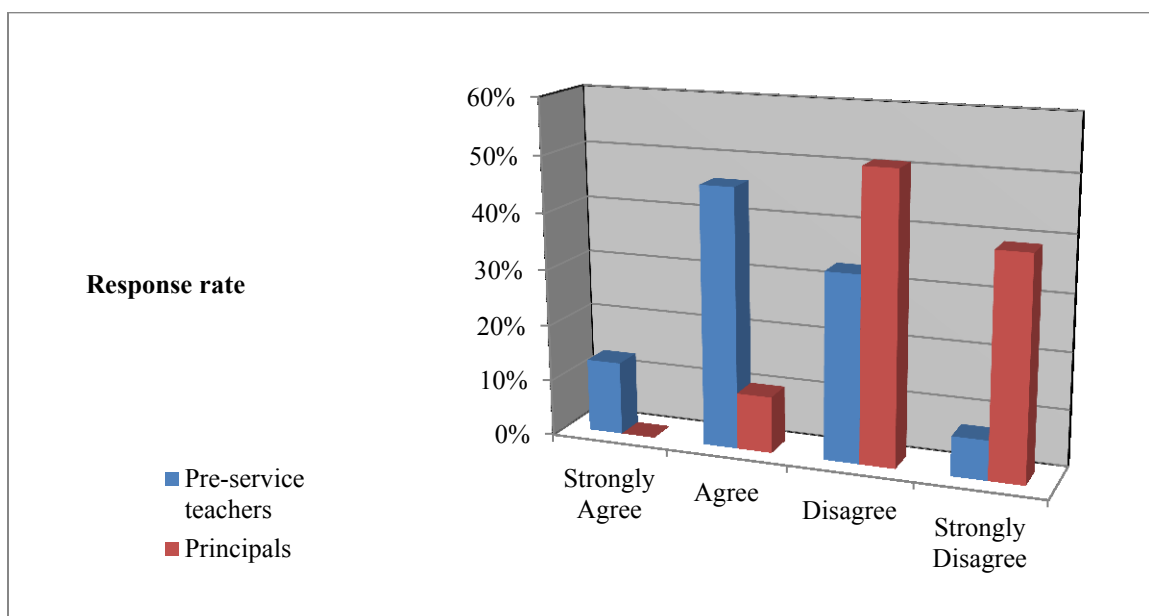
Figure S1  
Sufficient knowledge on how to use assessment results to help plan instruction, Item  
Response Comparison



## APPENDIX T

## QUESTION T – ITEM 3.6

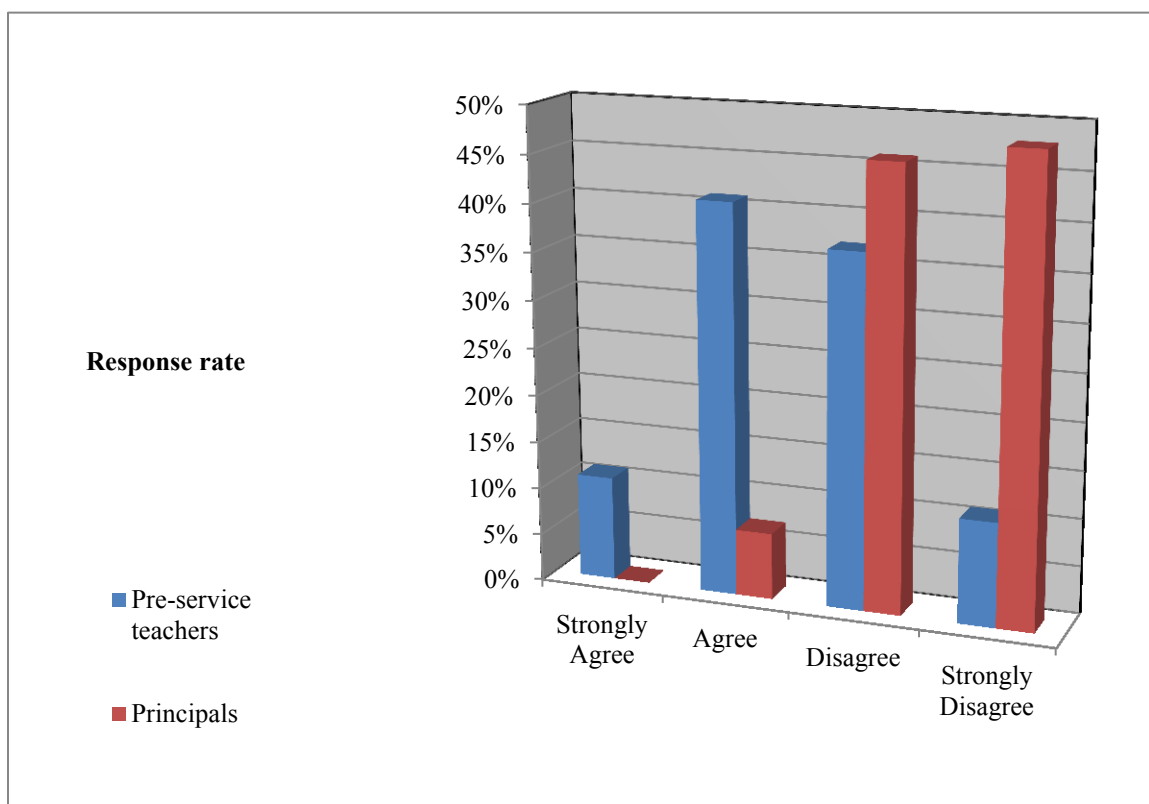
Figure T1  
Sufficient knowledge to use data to interpret test results for students and parents, Item  
Response Comparison



## APPENDIX U

## QUESTION V – ITEM 3.3

Figure U1  
Sufficient knowledge to use data to interpret test results, Item Response Comparison

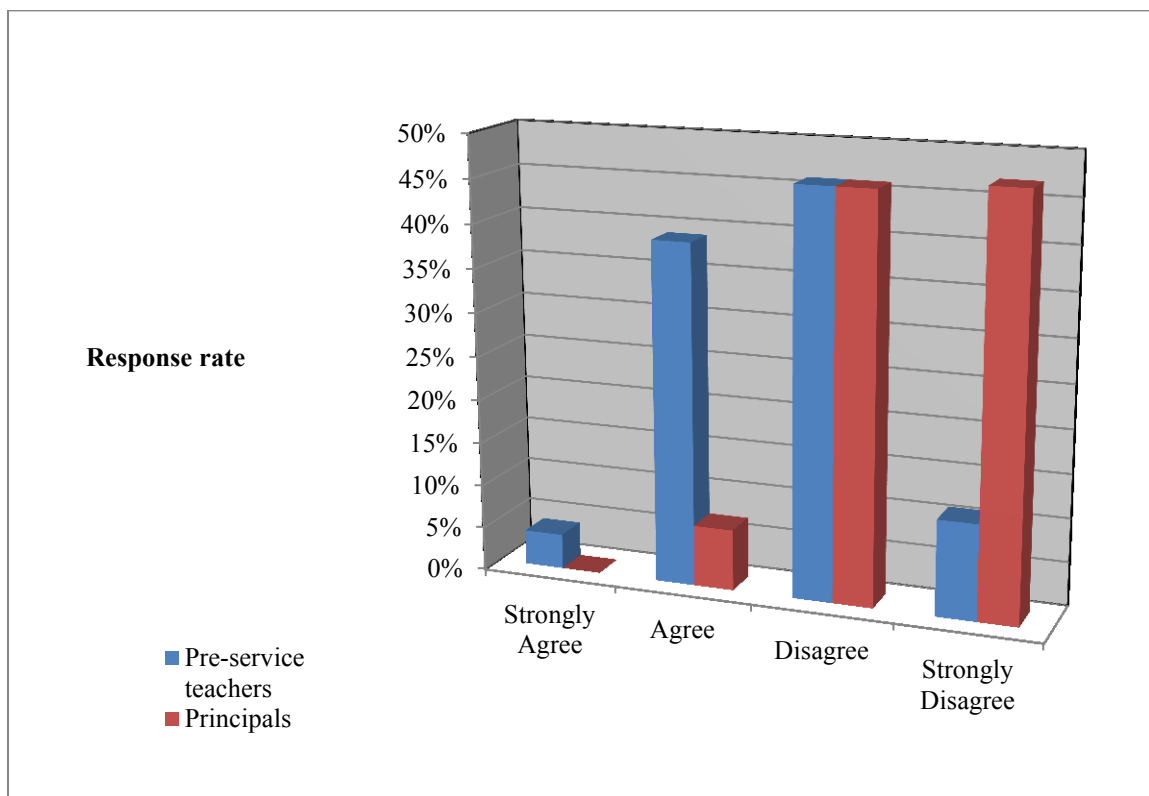




## APPENDIX V

## QUESTION V – ITEM 3.4

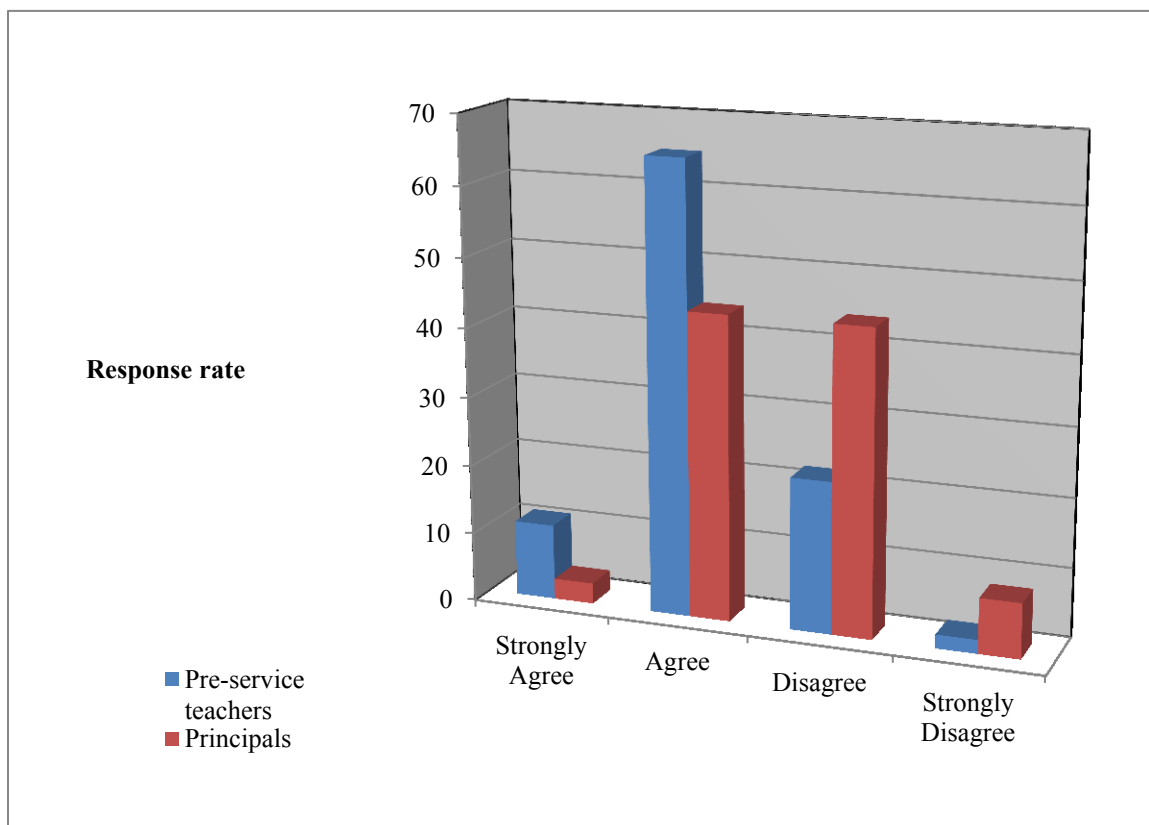
Figure V1  
Sufficient knowledge to use performance assessment methods, Item Response  
Comparison



## APPENDIX W

## QUESTION V – ITEM 3.1

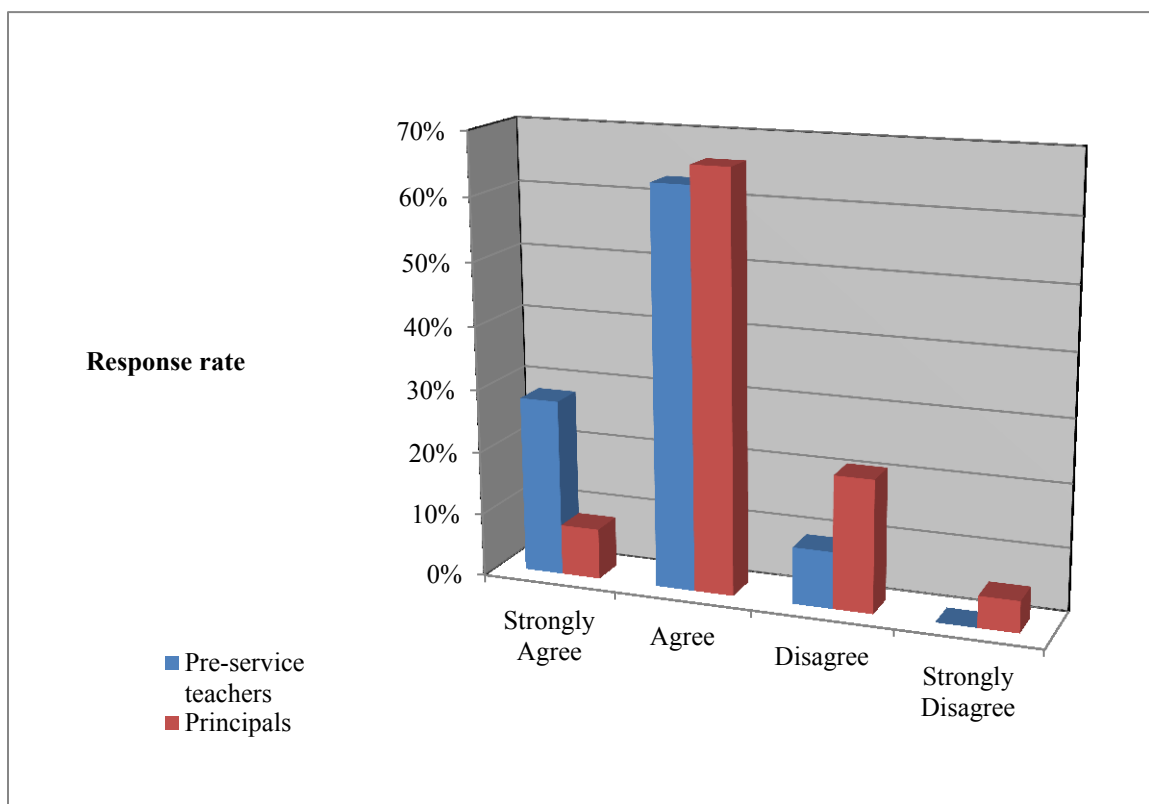
Figure W1  
Sufficient knowledge of classroom assessments and procedures, Item Response  
Comparison



## APPENDIX X

## QUESTION V – ITEM 3.5

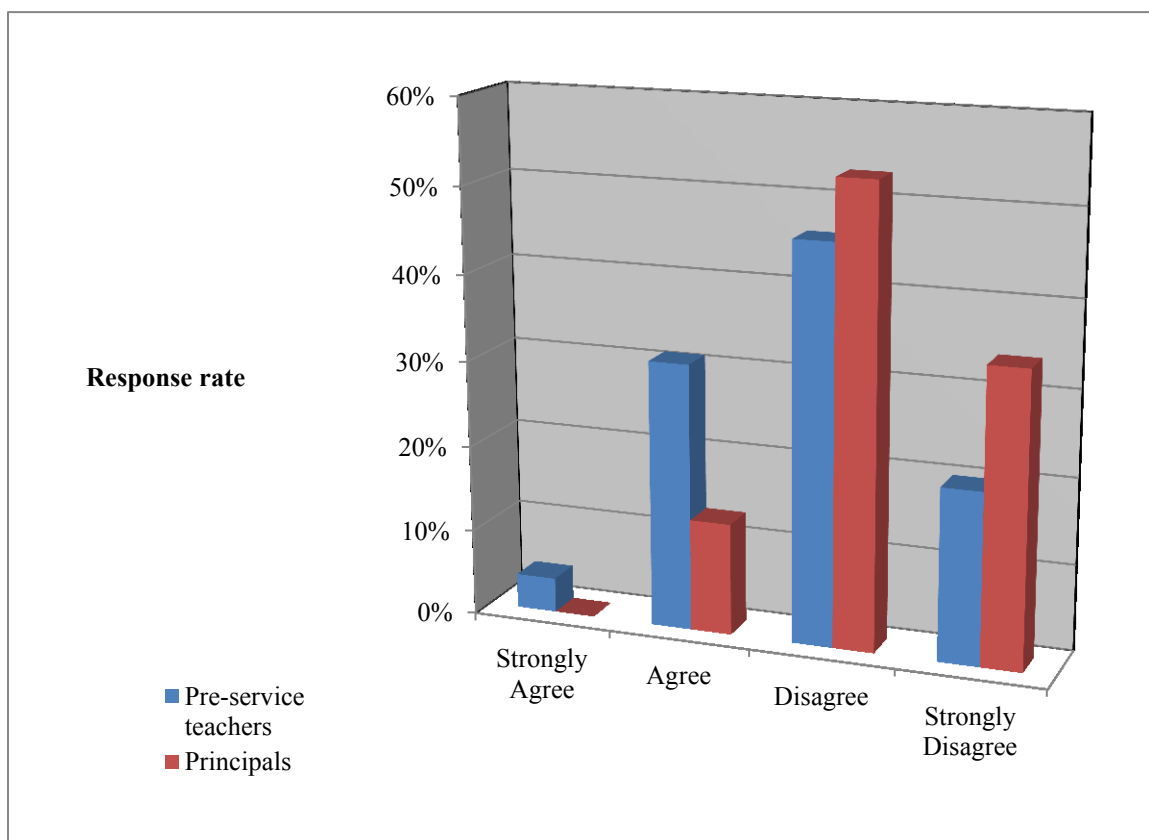
Figure X1  
Sufficient knowledge to grade student work, Item Response Comparison



## APPENDIX Y

## QUESTION V – ITEM 3.2

Figure Y1  
Sufficient knowledge to write sound test items, Item Response Comparison



## BIBLIOGRAPHY

- Afflerbach, Peter. 2012. Understanding Reading Assessment, K – 12 (*second edition*), International Reading Association: 96 – 117.
- ASCD. 2012. Common Core Standards Adoption by State, [www.ascd.org/common-core/common-core-standards](http://www.ascd.org/common-core/common-core-standards). (Accessed 3/13/2012).
- Berliner, David C. June 1, 2009. “*Rational Responses to High-stakes Testing and the Special Case of Narrowing the Curriculum*”, Paper presented at the International Conference on Redesigning Pedagogy, National Institute of Education, Nanyang Technological University, Singapore, 21.
- Black, Dylan; Lee, Clare; Harrison, Christine; and Black, Paul. March 2004. Teachers developing assessment for learning’s impact on student achievement, *Assessment in Education*, Vol. II, No. 1, 1.
- Breakstone, Joel; Smith, Mark; and Wineburg, Sam. 2012. “Beyond the Bubble: New History/ Social Studies Assessments for the Common Core.” Stanford University, 1 – 11, May 1, 2012.
- Brookhart, Susan. 2004. Classroom Assessment: Tensions and Intersections in Theory and Practice, *TCRecord: The Voice of Scholarship in Education*, Volume 106 Number 3, 1-23.
- Center on Education Policy. December 2008. Lessons from the Classroom Level: Federal and State Accountability in Illinois. Washington, D.C.: Center on Education Policy. [www.cep.dc.org](http://www.cep.dc.org). (accessed 6/14/2010).
- Center on Education Policy. December 2009. Improving Low-Performing Schools: Lessons from Five Years of Studying School Restructuring Under No Child Left Behind. Washington, D.C.: Center on Education Policy. [www.cep.dc.org](http://www.cep.dc.org) (accessed 6/14/2010).
- Center for K-12 Assessment & Performance Management at ETS. 2010. “Coming Together to Raise Achievement: New Assessments for the Common Core State Standards,”  
<http://www.ode.state.or.us/teachlearn/subjects/mathematics/standards/comingtogether toraiseachievement.pdf> (Accessed 3/13/2012).
- Chen, C. 2011. Numbers and Types of Public Elementary and Secondary Schools From the Common Core of Data: School Year 2009–10 (NCES 2011-345). U.S. Department of Education. Washington, DC: National Center for Education Statistics. <http://nces.ed.gov/pubsearch> (Accessed 3/11/12).

- DeLuca, Christopher and Klinger, Don A. 2010. Assessment literacy development: identifying gaps in teacher candidates' learning, *Assessment in Education: Principles, Policy & Practice*, 17: 4, 419-438.
- Embertson, Susan E. 2003. *The Second Century of Ability Testing: Some Predictions and Speculations*, Educational Testing Service, Policy Information Center, Princeton, NJ 08541-0001.
- Frisbie, David A. and Friedman, Stephen J. Fall 1987. "Test Standards-Some Implications for the Measurement Curriculum," *Educational Measurement: Issues and Practices*, 17-23.
- Giordano, Gerard. 2005. "How Testing Came to Dominate American Schools: The History of Educational Assessment; Peter Lang Publishing, NY.
- Goe, Laura. Spring 2007. Fighting Inequalities in Education, *News on Research, Products, and Solutions for Learning and Education: Innovations*, Issue 3, ETS.
- Gullickson, Arlen R. and Hopkins, Kenneth D. Fall 1987. The Context of Educational Measurement Instruction for Pre-service Teachers: Professor Perspectives, *Educational Measurement: Issues and Practice*, 12-16.
- Gunzenhauser, Michael G. Winter 2003. High-Stakes Testing and Default Philosophy of Education, *Theory into Practice*, Volume 42, Number 1, College of Education, The Ohio State University.
- Heritage, Margaret. 2008. "Learning Progressions: Supporting Instruction and Formative Assessment", National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California, Los Angeles.
- Hess, F.M. 2003. "Refining or Retreating? High-Stakes Accountability in the States, in No Child Left Behind? The Politics and Practice of School Accountability," Peterson, P., and West M., (eds.), Washington DC: The Brookings Institution.
- Hlebowitsh, Peter S. 2005. *Designing the School Curriculum*, Allyn & Bacon, Inc. 71.
- Hollenweger, Judith. 2011. "Teachers' ability to assess students for teaching and supporting learning." *Prospects*, 2011, UNESCO IBE.
- Jones, M. Gail, Jones, Brett D., and Hargrove, Tracy Y. 2003. *The Unintended Consequences of High-Stakes Testing*, Rowman and Littlefield Publishers, Inc. Lanham, Maryland, 41.

- Kleinsasser, Audrey M. March – April, 1995. “Assessment Culture and National Testing,” *The Clearing House*, Vol. 68, No. 4, pp. 205 – 210, Taylor & Francis, Ltd.
- Linn, Robert L. October 1986. Educational Testing and Assessment: Research Needs and Policy Issues, *American Psychologist*, University of Illinois at Urbana-Champaign, Oct. 1986, Vol. 41, No. 10, 1153, [HTTP://pareONLINE.NET/GETVN.ASP?V=7&N=11](http://pareONLINE.NET/GETVN.ASP?V=7&N=11), (accessed 11/2/2005).
- Linn, R. L., & Gronlund, N. E. 2000. *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall., 31.
- Manna, Paul, and Ryan, Laura L. 2011. Competitive Grants and Educational Federalism: President Obama’s Race to the Top Program in Theory and Practice, *Publius: The Journal of Federalism*, Oxford University Press on behalf of CSF Associates Publius, Inc., 2011, volume 41 number 3, pp. 522-546, <http://publius.oxfordjournals.org> (accessed 3/12/2012).
- Meisels, Samuel J., Atkins-Burnett, Sally, Xue, Yange, Nicholson, Julie, Bickel, Donna DiPrima, and Son, Seung-Hee. February 28, 2003. “Creating a System of Accountability: The Impact of Instructional Assessment on Elementary Children’s Achievement Test Scores,” *Educational Policy Analysis Archives*, Volume 11 Number 9, February 28, 2003.
- Mertler, Craig A. October 15-18, 2003. “Preservice Versus Inservice Teachers’ Assessment Literacy: Does Classroom Experience Make a Difference?” Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus, Ohio, October 15-18, 2003, 2-29.
- Mertler, Craig A. and Campbell, Cynthia. 2005. “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, April 11-15, 2005, 1-27.
- Mertler, Craig A. July 2009. “Teachers’ assessment knowledge and their perceptions of the impact of classroom assessment professional development,” *University of West Georgia, Improving Schools- SAGE Publications, Volume 12, Number 2 July 2009, 101 -113*.
- Nitko, Anthony J. 2001. “Educational Assessment of Students (3<sup>rd</sup> Edition), Prentice Hall, Inc., NJ, pg. 3.
- Noonan, B., & Renihan, P. 2006. Demystifying assessment leadership. *Canadian Journal of Educational Administration and Policy*, 56, 1-20.

- Peterson, Paul E., and West, Martin R. 2003. *No Child Left Behind: The Politics and Practice of School Accountability*, Brookings Institutional Press, Washington, D.C., 3
- Popham, W. James. March 2002. *Implementing ESEA's Testing Provisions: Guidance from an Independent Commission's Requirements*, (University of California, Los Angeles), 3.
- Popham, W. James. 2004. "Curriculum, Instruction, and Assessment: Amiable Allies or Phony Friends?" *Teachers College Record* Volume 106 Number 3, 2004, p. 417 – 428, <http://www.tcrecord.org> (accessed 11/1/2005).
- Popham, W. James. 2009. "Assessment Literacy for Teachers: Faddish or Fundamental?" *Theory Into Practice*, The College of Education and Human Ecology, The Ohio State University, 48: 4-11.
- Ravitch, Diane. 2010. "Dictating to the Schools: A look at the effects of the Bush and Obama administrations on schools." *Virginia Journal of Education*, November 2010, [www.eddigest.com](http://www.eddigest.com) (accessed 5/13/2012).
- Sable, J., Plotts, C., and Mitchell, L. 2010. *Characteristics of the 100 Largest Public Elementary and Secondary School Districts in the United States: 2008–09* (NCES 2011-301). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Schafer, William D. 1991. Essential Assessment Skills in Professional Education of Teachers/ WDS, University of Maryland, 3 – 12, *Educational Measurement: Issues and Practices*, Vol. 10, Number 1, Spring 1991) (Relevant Classroom Assessment Training For Teachers, RJS, Northwest Regional Educational Laboratory, 7 – 12, *Educational Measurement: Issues and Practices*, Vol. 10, Number 1, Spring 1991).
- Schafer, William. 1993. "Assessment Literacy For Teachers", Electronic TIP Volume XXXII No 2, Spring 1993, 119.
- Shavelson, Richard J., Young, Donald B., Ayalya, Carlos C., Brandon, Paul R., Rurtak, Erin Marie, Ruiz-Primo, Maria Araceli, Tomita, Miki K., and Yin, Yue. 2008 'On the Impact of Curriculum-Embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers', *Applied Measurement in Education*, 21: 4, 295-314.
- Shepard, Lorrie A. February 2000. "The Role of Classroom Assessment in Teaching and Learning", CSE Technical Report 517, February 2000, CRESST/ University of Colorado at Boulder.



- Standards for Teacher Competence in Educational Assessment of Students.* 1990. Developed by American Federation of Teachers, National Council on Measurement in Education, National Education Association, Bureau of Educational Research and Statistics, www.unl.edu/buros/bimm/html/article3.html, (accessed 8/01/2005).
- Stiggins, Richard. 1999. "Learning Teams For Assessment Literacy: A Concept Paper", 2-3, written by the Assessment Training Institute.
- Stiggins, Richard J. Fall 2001. *Educational Measurement: Issues and Practice*, Fall 2001, "The Unfulfilled Promise of Classroom Assessment", 5-14, 6.
- Stiggins, Richard J. June 2002. "Assessment Crisis: The Absence of Assessment FOR Learning," *Phi Delta Kappan*, V. 83, No. 10, Pages 758-765.
- Stiggins, Richard J. February 1998. "Auditing the Quality of Classroom Assessment Training in Teacher Education Programs", paper presented at the Annual Meeting of the American Association of Colleges of Teacher Education, New Orleans, LA.
- Stiggins, Richard and DuFour, Rick. May 2009. "Maximizing the Power of Formative Assessments," *Phi Delta Kappan*, Vol. 90, No. 09, May 2009, pp. 640-644.
- Urban, Wayne J., and Wagoner, Jr., Jennings L. 2009. *American Education: A History* (4<sup>th</sup> Edition), Routledge, 235.
- U.S. Department of Education. 2008. 'A Nation Accountable: Twenty-five Years After A Nation at Risk', Washington, D.C.
- U.S. Department of Education. 2010. *ESEA Blueprint for Reform*, Office of Planning, Evaluation and Policy Development, Washington, D.C.
- Valli, Linda, Croninger, Robert G., Chambliss, Marilyn J., Graeber, Anna O., and Buese, Daria. 2008. "Test Driven: High-Stakes Accountability in Elementary Schools," Teachers College Press, New York.
- Whittington, Dale. February 1998. "Standards for Teacher Competence in Educational Assessment of Students: Why They Are Needed, What They Say, and How They Could Influence Preservice Teacher Education", Paper presented at the Annual Meeting of the American Association of Colleges of Teacher Education, New Orleans, LA, John Carroll University, 3
- William, Dylan. 2006. "Formative Assessment: Getting the Focus Right", *Educational Assessment*, II (3 & 4), 283-289, Educational Testing Service, Princeton, NJ
- Wise, Steven L. 1993. "Teachers' Testing Knowledge, Skills, and Practices," Ronald N. Marso and Fred L. Pigge (Bowling Green State University) - 129 – 185, *Teacher*

*Training in Measurement and Assessment Skills*, (University of Nebraska-Lincoln): Volume Editor, Buros-Nebraska Symposium on Measurement & Testing, Copyright- Lincoln, NE