

An Investigation of Charter Schools' School Leader and Teacher Level of Assessment
Literacy

by

Mareen Pfeiffer-Hoens

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctorate of Assessment, Learning and School Improvement

Middle Tennessee State University
August 2017

Dissertation Committee:
Dr. Donald Snead, Chair
Dr. Kathleen Burris
Dr. Patricia Patterson

I dedicate this research to Christian Pfeiffer, my grandfather, and Holger Pfeiffer, my father; both were unable to attend or complete higher education but always supported and encouraged my efforts and dreams.

ACKNOWLEDGMENTS

I would like to thank my family, Joe, Finn, and Laina for loving me, supporting me, and not giving up on me while I traveled down this long doctoral road. During these four years of continuous graduate studies and working full-time, I know all of you sacrificed much due to my absence in mind, body, and spirit in daily activities and family events. Thank you for that sacrifice and supporting me during this journey.

I would also like to thank my dear friend, Kristen Thompson, who helped me keep my sanity and stay in shape by being my devoted running buddy and who planned her wedding around my crazy class schedule.

I also want to give special thanks to my committee members for their investment in and influence on me. Dr. Kathy Burris, thank you for being a part of this journey. Who would have thought that a short conversation of advice ten years ago would have led to this milestone in my life. Your experience and advice helped me choose the path of elementary education and I am grateful to have had such an inspiring mentor. Dr. Patricia Patterson, thank you for jumping in on this journey on such short notice. I appreciate your feedback and time that you have given to help me become Dr. Mareen Hoens. Dr. Donald Snead, thank you for being a part of this incredible journey. What once seemed impossible became possible and I could not have done it without your support. I thank you for your guidance, your patience, your availability, your flexibility, and sharing your knowledge. This journey has been a process but I am glad to have experienced it with such a supportive chair and committee. I feel fortunate to have learned from all three of you—Dr. Snead, Dr. Burris, and Dr. Patterson—during this process.

ABSTRACT

Assessment of student performance is one of the most critical responsibilities of school leaders and teachers. Teachers and school leaders must acquire an understanding of assessment literacy for utilizing data to make sound data-driven decisions. The purpose of this descriptive study was to investigate the levels of assessment literacy among elementary teachers and school leaders in charter schools. The target population was approximately 200 educators from one elementary charter school network with schools primarily in an urban district in Tennessee and in an urban district in California. The instrument used for this research study was the Classroom Assessment Literacy Inventory (CALI) (Mertler, 2003). The inventory consists of two sections. The first section of the survey consists of demographic questions regarding years of experience, level of education, intensity of assessment training, and type of teacher education program. The second section uses the CALI to measure the level of assessment literacy of teachers and school leaders. A total of five school leaders and 58 teachers completed the survey.

The data from the survey showed that teachers and school leaders averaged 20.23 out of 35 questions correct (57.8%). The average is slightly lower than the average obtained in the study of in-service teachers by Plake et al. (1993) and in the study of in-service teachers by Mertler (2003). Findings indicated that the level of assessment literacy has not changed significantly in over twenty years. The results are discussed in terms of use to establish a baseline that can be used in further study of assessment literacy of both classroom teachers and school leaders, especially through professional development.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER ONE INTRODUCTION	1
Assessment.....	4
Assessment Literacy	5
Statement of Purpose	7
Significance of the Study	8
Theoretical Framework.....	11
Description of the Study	14
Definition of Terms	14
Limitations	17
Summary.....	17
CHAPTER TWO LITERATURE REVIEW	19
Introduction.....	19
Charter School—What Is It?.....	24
Assessment Literacy—What Is It?	35
Summary.....	49
CHAPTER THREE METHODOLOGY	51
Problem.....	51
Population.....	52
Instrument	53
Data Collection	55
Procedure	56

Data Analysis.....	57
Summary.....	58
CHAPTER FOUR RESULTS	59
Introduction.....	59
Part I Survey Results—Demographic Information.....	60
Part II Survey Results—CALI.....	62
Summary.....	75
CHAPTER FIVE SUMMARY OF THE STUDY	76
Introduction.....	76
Findings	77
Discussion and Implications	82
Limitations	86
Conclusions.....	87
Recommendations.....	88
Implications for Future Research.....	89
Summary.....	91
REFERENCES	92
APPENDICES	111
APPENDIX A IRB FORM.....	112
APPENDIX B PARTICIPATION INVITATION	114
APPENDIX C CLASSROOM ASSESSMENT LITERACY INVENTORY	115
APPENDIX D MEANS COMPARISON OF CALI SCORES	126
APPENDIX E SUMMARY OF CALI DATA OF COMPARISON GROUPS.....	127

LIST OF TABLES

Table 1 <i>Participants' Faculty Position</i>	53
Table 2 <i>Alignment of Standards with CALI Items</i>	55
Table 3 <i>Summary of Teachers' and School Leaders' Demographic Data</i>	62
Table 4 <i>Means and Standard Deviations for Teachers and School Leaders by Standard and Total on CALI</i>	64
Table 5 <i>Means and Standard Deviations for Teachers and School Leaders by Years of Experience</i>	66
Table 6 <i>Means and Standard Deviations for Teachers and School Leaders by Level of Education</i>	68
Table 7 <i>Means and Standard Deviations for Teachers and School Leaders by Intensity of Assessment Training</i>	70
Table 8 <i>Mean and Standard Deviation for Type of Teacher Education Program</i>	74
Table 9 <i>Means Comparison of CALI Scores</i>	126
Table 10 <i>Summary of CALI Data of Comparison Groups</i>	127

LIST OF FIGURES

Figure 1. Total mean score comparison, traditional versus alternative teacher education program	71
Figure 2. Standard mean score comparison, traditional versus alternative teacher education program	73

CHAPTER ONE

INTRODUCTION

The worth of schools, school leaders, and teachers is judged by assessments. Assessment knowledge is a critical component needed to ensure effective school practices (Popham, 2014). Not only do school leaders and teachers need to take ownership of their environment and maintain an “assessment rich vocabulary” but they also must understand data from looking at each individual student, classroom, school, and state level that determines how to use these assessments to determine their professional value. Two significant problems were noted regarding educators’ understanding of assessment: the lack of pre-service teacher training and the lack of appropriate professional development for teachers (Stiggins, 1998; Black & William, 1998).

Assessment literacy is the fundamental understanding of assessment concepts and procedures intended to influence educational decisions. It is the process of teaching and learning that enables educators to evaluate and enhance student learning (Popham, 2014; Stiggins, 1991, 2014; Lezotte & Snyder, 2011). Lezotte and Snyder (2011) noted that assessment literacy is foundational to data-driven decision making which is integral toward guiding effective school improvement practices. Brookhart (1999) emphasized the importance of teachers using assessments that are valid, reliable, meaningful, and accurate to guide instruction. Zwick et al. (2008) found that educators must acquire an understanding of assessment literacy for utilizing data to diagnose needs of individual students. Popham (2008) and Stiggins (1999, 2014) stated that ineffective assessment practices and lack of assessment literacy can hinder student learning, decrease student achievement, and eventually push students away from school. Mertler (2006) suggested

that lack of exposure to assessment fundamentals helps to explain why teachers do not recognize the importance of assessment for improved instruction, student motivation, and level of student achievement.

Since the passage of the No Child Left Behind Act (NCLB) of 2001, there has been an emerging shift from the era of assessing merely to rank students through norm-referenced results of test scores to educators interpreting criterion-referenced results from tests, quizzes, and many other academic projects (Marzano, 2000). Unfortunately, this emerging shift from assessing to ranking to guiding instruction and student learning has been difficult for many educators. While NCLB pushed this course of accountability of mandated state testing and educators' accountability for students' test scores, the needed assessment training for pre-service and in-service teachers has not been provided (Black & William, 1998). Black and William (1998) note, "There is a wealth of research evidence that the everyday practice of assessment in the classroom is beset with problems and shortcomings" (1998, p. 5). The ever-expanding and changing assessment responsibilities have left school leaders and teachers with only a minimum knowledge of assessment, inadequate application of assessments, and little understanding of how to apply assessments in a standards-based system. This disconnect between national mandates and assessment practices provides necessary evidence for the promotion of assessment literacy.

Whether education and educational policies push a culture of testing or a culture of learning, it is paramount for educators to demonstrate a high level of assessment literacy. As the importance of assessment literacy increases, school leaders and teachers are expected to be skilled assessment practitioners, to design and interpret more student-

involved classroom assessments, and to contribute to political agendas associated with high-stakes testing and achievement data use (Fullan, 2001; Popham, 2008; Stiggins, 2002b).

Yin et al. (2008) found that assessment literacy and assessment reform require significant preparatory measures. Researchers note the need for further exploration of the relationship between traditional teacher education programs and alternative teacher education programs regarding school leaders' and teachers' levels of assessment literacy, especially in charter schools (Mertler, 2005; Veltri, 2008; Siegel & Wissehr, 2011). While state licensing standards require competence in assessment practices, colleges of education and state education agencies continue to require minimal, if any, specific coursework in assessment (Campbell et al., 2002; Kirkpatrick, Lincoln, & Morrow, 2006). According to Boudett, City, and Murnane (2006), the lack of assessment knowledge to lead the process of data-driven inquiry in school leaders was one of the most frequently cited inhibitors to its implementation.

Over the last decade, charter school networks have provided a fertile ground for a growing market of alternative teacher education programs to reduce teacher shortcomings and improve teacher and school leader quality (Cave & Brown, 2010; Carl, 2014; Yancey, 2006). Thus, there is a pressing need for researchers to gather information from practicing charter school leaders and charter teachers about their teacher education background, current levels of assessment literacy, and the resulting relationship among the two variables. It is anticipated that the results of this quantitative study will contribute to an understanding of assessment literacy in school leaders and teachers in charter schools.

Assessment

To improve teachers' competence in educational assessment, the American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), and National Education Association (NEA) developed the Standards for Teacher Competence in Educational Assessment in 1990. There are seven standards addressing seven broad skill areas in educational assessment (Sanders et al., 1990):

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3. Teachers should be skilled in administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods.
4. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
5. Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments.
6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Each standard was considered an expectation for assessment knowledge or skill that a teacher should possess. As a set, the Standards for Teacher Competence were

intended to be a guide for teacher educators in their work with teacher education programs, a self-assessment guide for teachers, and a guide for workshop and educational measurement instructors to conceptualize student assessment more broadly than had been done in the past (Sanders et al., 1990).

While it has been over 25 years since the publication of the Standards for Teacher Competence in Educational Assessment of Students, not much has changed with regard to teacher preparation programs and assessment literacy. The focus on the concept of assessment literacy has drawn attention to the importance of school leaders and teachers incorporating various assessment practices. Despite the increase in expectations of assessment practices, the background knowledge many teachers and school leaders have for assessment-related tasks (such as constructing and using assessments to evaluate learning, understanding results to interpret them for students and parents, and planning instructional programs that meet the needs identified by those results) is limited (Campbell et al., 2002; Popham, 2008; Stiggins, 2002b). In fact, Popham (2009) argues that the lack of assessment literacy knowledge could “cripple the quality of education” (p. 1). Teachers and school leaders still consider themselves assessment-illiterate and unprepared in the use of assessment practices (Campbell & Mertler, 2005; Stiggins, 2014; Popham, 2011; Zwick et al., 2008).

Assessment Literacy

What Is Assessment Literacy?

While various definitions exist throughout assessment literacy research, the fundamental concepts of assessment literacy are defined as the knowledge of 1) the means for assessing what students know and can do, 2) the interpretation of the results

from these assessments, and 3) the application of assessment results to improve student learning and program effectiveness (Popham, 2014). This working definition of assessment literacy will be referred to throughout this study.

Assessment Literacy Inventory

The 1990 Standards for Teacher Competence in the Educational Assessment of Students provided the guiding framework for six of the assessment literacy instruments: The Assessment Literacy Inventory (ALI); the Assessment Practices Inventory (API); the Assessment in Vocational Classroom Questionnaire, Part II; the Classroom Assessment Literacy Inventory (CALI); the revised ALI; and the Teacher Assessment Literacy Questionnaire (TALQ) (DeLuca & Klinger, 2010).

Of these instruments, the Teacher Assessment Literacy Questionnaire (TALQ) developed by Plake and Impara in 1992 served as a blueprint for other instruments measuring educators' levels of assessment literacy. The TALQ is a 35-item, content-based instrument developed to measure in-service teachers' competency in the 7 Standards articulated in the 1990 Standards for Teacher Competence in the Educational Assessment of Students. The survey was one component of a grant from the W.K. Kellogg Foundation to the NCME to develop prototype materials to help teachers become more proficient in assessing student performance. Impara, Plake, and Fager (1993) were the principal investigators for the grant. Since then the TALQ has been revised numerous times to continue to report on research on assessment literacy (Plake, Impara, & Fager 1993).

The Assessment Practice Inventory (API) and the Assessment in Vocational Classroom Questionnaire, Part II, were developed using Likert-type items. The API

(Zhang and Burry-stock, 1997) was developed to measure in-service teachers' perceptions of their assessment skills, while the Assessment in Vocational Classroom Questionnaire, Part II was developed to measure in-service teachers' perceived level of competence in assessment activities (DeLuca and Klinger, 2010).

The instrument used for this research study was the Classroom Assessment Literacy Inventory (CALI). In 2002, Mertler revised the TALQ, primarily changing names of fictitious teachers, changing word choice to improve clarity, and renaming the questionnaire the Classroom Assessment Literacy Inventory (Mertler, 2003, p. 14).

Statement of Purpose

Assessment is a critical component in the process of teaching and learning as it enables educators to evaluate and utilize the information to improve learning and instruction (Harris, Irving, & Peterson, 2008). Brookhart (1999) stresses the need for school leaders and teachers using assessments to make data-driven decisions. Mertler (2006) suggests that lack of exposure to assessment fundamentals helps to explain why teachers do not readily recognize the importance of assessment to improve instruction, student motivation, and level of student achievement. Despite its seemingly obvious relation to enhancement of instruction and raising student achievement, a lack of training in assessment practices has been known for over a decade now (Zwick et al., 2008; Campbell et al., 2002; Stiggins, 2014; Popham, 2011).

The purpose of this study was to determine the levels of assessment literacy of school leaders and teachers employed in a public charter network as measured by the CALI (Mertler, 2003). Demographic data was collected to investigate differences in level of assessment literacy among teachers and school leaders related to their years of

experience (zero to five years and six plus years), level of education, intensity of assessment training, and type of teacher education program. Determining school leaders' and teachers' level of assessment literacy may clarify some important aspects of the year-long professional development implementation regarding assessment literacy. To the interest of the researcher, the following research questions were raised to pursue understanding of school leaders' and teachers' levels of assessment literacy in the public charter school network. Three research questions were asked to address this issue.

1. What is the level of assessment literacy of public charter school teachers and school leaders as measured by the Classroom Assessment Literacy Inventory?
2. Are there differences in level of assessment literacy among teachers and school leaders related to their years of experience (zero to five years and six plus years), level of education, and intensity of assessment training?
3. How does the assessment literacy of teachers and school leaders educated in a traditional education program compare to that of teachers educated in an alternative education program?

Descriptive data was collected and analyzed through the Classroom Assessment Literacy Inventory (CALI) to determine what levels of assessment literacy school leaders and teachers in a specific charter network possess.

Significance of the Study

This study contributes to the broader research in scholarly, policy, and practice circles regarding assessment literacy and teacher education. Although assessment literacy has been explored from various standpoints, looking primarily at pre-service and

in-service teachers in public schools, studies have not examined the level of assessment literacy in school leaders and teachers operating within a charter network. Charter schools have become a distinct and growing part of the public-school system. The purpose of charter schools is to improve the nation's public school system and close the achievement gap. Due to its unique set-up, a charter school has the freedom to be more innovative and flexible in its organization, structure, and instruction. Alternative teacher education programs such as Teach for America (TFA) and The New Teacher Program (TNTP), are preparing teachers to work in hard-to-staff urban school districts, often predominantly charter schools. Limited studies were found regarding the levels of assessment literacy among educators attending alternative education programs.

It is critical that supporting school districts and charter networks can identify the strengths and weaknesses of their school leaders' and teachers' assessment literacy to inform professional development programs aimed at increasing that same assessment literacy. Consequently, the research associated with assisting teachers to build and implement effective assessments will be necessary for educators to develop a system that supports both the formative (during learning) and summative (after learning) functions of assessments.

Educators who are literate in assessment have clear knowledge of the state standards and what students are expected to know at the different grades (Popham, 2011). They can develop and select assessments to fit a context that reflects the specific achievement goals and objectives. Despite its seemingly obvious relation to enhancement of instruction and raising student achievement, a lack of training in assessment practices has been recognized for more than a decade now (Zwick et al.,

2008; Campbell et al., 2002; Stiggins, 2014; Popham, 2011). There are three common reasons that the concept of assessment literacy has received and still receives increased attention in educational reform.

First, the advent of standards-based reform has made student expectations for learning more explicit and has increased the need for measures to determine whether students have attained those learning expectations. It is widely known that standards-based accountability policies that include high-stakes testing are currently the dominant school reform approach in the United States. These policies were designed to raise students' educational outcomes and reduce race and class achievement gaps by linking students' test scores to rewards and sanctions for both schools and students (Diamond & Cooper, 2007). Guskey (2007) maintains that, "Policy makers and legislators at the state and national levels see assessments as essential for change. They believe that good data on student performance drawn from large scale assessments will help focus educators' attention and guarantee success" (p. 15).

Second, the use of different forms of assessments, such as formative (assessment for learning) and summative (assessments of learning) assessments has been more widely accepted. Both a greater emphasis on learning expectations and the more formal use of alternative assessments has increased the burden on teachers and principals to understand how student learning can be adequately assessed, and what meaning should be given to the information produced (Black & William, 1998; Popham, 2011; Stiggins, 2014).

Third, previous research on assessment practices shows that school leader and teacher education programs have only minimal coursework focused on assessment literacy (Mertler & Campbell, 2005; Stiggins & Duke, 2008). A review of the top ten

graduate schools in Educational Administration and Supervision as listed by U.S. News and World Report in 2013 shows that 6 out of 10 programs require a course in assessment or data-driven decision making (Cochran-Smith et al., 2013).

While assessment literacy is only one prerequisite for effective teaching (Stronge, 2007), it is a foundational skill that all school leaders and teachers must have to evaluate what students know and can do, to interpret the results from these assessments, and to apply these results to improve student learning and program effectiveness (Stiggins, 1999; Mertler & Campbell, 2002). Research shows that educators who are literate in assessment and effectively use assessment for and of learning as well increase student learning, student motivation, and student achievement (Black & William, 1998).

With this increased understanding, the researcher may be able to provide better instructional support and professional development to practicing school leaders and teachers within and outside of the charter network.

Theoretical Framework

The theoretical framework directing this study was based on Shepard's (2000) historical framework which highlights and summarizes the key tenets of social efficiency curricula, behaviorist learning theories, and "scientific measurement."

Influenced by the behaviorism theories of Hull, Skinner, and Gagne, Thorndike's theory of learning created an early conceptualization of learning and testing. Shepard (2000) summarizes the key assumptions:

1. Learning occurs by accumulating atomized bits of knowledge.
2. Learning is tightly sequenced and hierarchical.
3. Transfer is limited, so each objective must be explicitly taught.

4. Tests should be used frequently to ensure mastery before proceeding to the next objective.
5. Tests are isomorphic with learning (tests = learning).
6. Motivation is external and based on positive reinforcement of many small steps.

(The Role of Assessment in a Learning Culture. *Journal of Education*, 189(1/2), 95-106).

Shepard (2000) notes the historical influences of Thorndike's contributions to the field of learning theories and educational measurement. Thorndike's influence on the development of the "objective" test (Shepard, 2000) is undeniable since it has been the dominant method of achievement testing in the United States from the beginning of the twentieth century to the present day. While many others followed in the wave of improving instruction through educational measurement, Thorndike set the tone for the next 100 years. Researchers in education such as Bobbitt (1915) and Ballou (1916) (as cited by Shepard, 2000) supported Thorndike's view of educational measurement as a means to study education, but also to intervene to improve education. Ballou (1916), for example, suggested a similar conceptualization of learning and testing:

1. The quality of educational results must be measured by the best available standard tests.
2. Results must be analyzed and suggestions for improvement made where results are unsatisfactory.

3. After reasonable time has elapsed, “similar standard tests must be repeated to determine what effect, if any, the suggestions have had on instruction” (p. 355 as cited by Shepard, 2016).

Shepard (2000) provides an interlocking set of themes that have dominated the learning culture throughout the twentieth century. At the time, learning theories were dominated by the Hereditarian Theory of Intelligence, which explained Intelligence Quotient (IQ) as innate, unitary, and fixed, and the Associationist (Thorndike) and Behaviorist (Hall, Skinner, Gagne) theories. The Associationist and Behaviorist learning theories required learning to be tightly sequenced and hierarchical, where each objective needed to be explicitly taught and assessed, creating a test-teach-test environment. At the time, tests were seen as isomorphic with learning. Scientific measurement pushed the use of IQ testing to rank students by ability and to then measure achievement based on objective tests. The curriculum development was dominated by the need for social efficiency. While differentiated, curriculums existed; they were not based on students' individual needs but rather based on predicted social roles. Many of our current views of assessment and educational belief systems are derived from and often still entwined with these views of the past. Surprisingly, it is the teachers who are holding on to the assessment beliefs of the past. Bliem and Davinroy (as cited by Shepard, 2000) note that teachers feel more comfortable if assessment practice is separate from instruction and objective. Shepard (2000) and Darling-Hammond (1995) suggest that to fulfill our educational purpose of ensuring that all students can learn, teachers and school leaders will need help in learning to use assessment to improve teaching and learning.

Description of the Study

This study describes school leaders' and teachers' levels of assessment literacy. The participating school leaders and teachers had been employed by the public charter school organization for at least one year. The charter school network is a non-profit organization serving primarily low-income students in neighborhoods where access to excellent schools is limited. Assessment literacy developed as a key component of the Standards of Teacher Competence in the Educational Assessment of Students (Sanders et al., 1990). The participants were given a six month window to take the questionnaire online. Once the questionnaires were completed, results were analyzed by the researcher.

In this study, the Classroom Assessment Literacy Inventory (CALI) was used as the instrument to measure school leaders' and teachers' levels of assessment literacy. Part One of the inventory consists of items related to the participants' background and demographics, while Part Two of the CALI assesses the level of assessment literacy. On the CALI, each correct answer is worth one point and an overall score out of a possible 35 points is awarded to measure the 35-item questionnaire.

Definition of Terms

The following terms are defined in the context in which they are utilized in this research:

Accountability assessment. The measurement devices, almost always standardized, used by governmental entities such as states, provinces, or school districts to ascertain the effectiveness of educational endeavors (Popham, 2009).

Alternative teacher education program. The process by which a person is awarded a teaching license even though that person has not completed a traditional

teacher certification program. Alternative certified teachers possess a bachelor's degree in a different field and are completing (or have completed) an alternative education program while teaching full-time (Justice et al, 2003).

Assessment. The process of judging—a formal attempt to determine a student's status with respect to an educational variable of interest (Popham, 2014).

Assessment literacy. The fundamental understanding of assessment concepts and procedures intended to influence educational decisions (Popham, 2009; Stiggins 2014; Mertler, 2004).

Charter school. A non-religious public school operating under a contract or “charter” that governs its operation. All details of school operation—its name, organization, management, and curriculum—are set by the charter, which outlines how the school will measure student performance. Charter schools are publicly funded; they must have open enrollment policies, may not charge tuition, and must still participate in state testing and federal accountability programs (O'Brian & Dervarics, 2010).

Classroom assessment. The formal and informal procedures that teachers employ to make accurate inferences about what their students know and can do (Popham, 2009).

Criterion-referenced assessment. A test or other type of assessment designed to provide a measure of performance interpretable in terms of a clearly defined and delimited domain of learning tasks (Linn & Gronlund, 2000).

Data-driven decision making (DDDM). A process of systematically collecting various types of assessment data and analyzing the data to guide a range of decisions to help improve student achievement (Ikemoto & Marsh, 2007; Ingram, Louis, & Schroeder, 2004).

Formative assessments. Traditionally referred to as pretests or pre-assessments and posttests or post-assessments generated by classroom teachers. These assessments are primarily used to inform instruction (Ainsworth & Viegut, 2006).

Meta-analysis. A quantitative approach to reviewing research literature in a specific area (Hattie, 2012)

Norm-referenced assessment. A test or other type of assessment designed to provide a measure of performance interpretable in terms of an individual's relative standing in some known group (Linn & Gronlund, 2000).

Psychometric analyses. The analysis of psychological tests and measurements to ensure that scores are as reliable and valid as possible (Popham, 2010).

Reliability. The degree to which the result of a test is dependable and has consistent results. Reliability is an indication of the consistency of a student's scores with the same tests, across time, or different tests measuring the same thing (Popham, 2010).

Standardized testing. A standardized test is any test that is administered, scored, and interpreted in a standard, predetermined manner (Popham, 2010).

Summative assessments. Assessments administered at the end of a unit, quarter, course, semester, trimester, or academic school year. These assessments report the results of student learning typically to support the assignment of letter grades or levels of proficiency. Annual standardized tests such as CST and district benchmarks administered after every quarter are examples of summative assessments (Ainsworth & Viegut, 2006).

Traditional teacher education program. Teachers earn certification through completing a bachelor's or master's degree in education, taking Praxis tests, and fulfilling additional state requirements (Justice et al, 2003).

Validity. A measure of whether a test measures what it is intended to measure (Popham, 2010).

Limitations

One significant limitation of the current study was the accuracy of the participants filling out the questionnaire and the demographics associated with the questionnaire. Even though the questionnaire was completely anonymous, the participant may have answered a question as though a superior might discover the answers. In addition, the demographics of the questionnaire may not be accurately completed. Any of these limitations could affect the validity of the study. Furthermore, participants filled out the survey electronically and in a private location. No one was available to answer participants' questions; therefore, misinterpretation of a question could affect the way the question was answered, thus affecting the validity of this study.

Summary

In this chapter, the researcher introduced the importance of an assessment-literate faculty. In this era of high stakes testing, data-driven decision making, and ongoing educational reforms, allow school leaders and teachers to demonstrate a high level of assessment literacy in order to make adequate instructional decisions.

In chapter two, the researcher reviews relevant literature on the topics of assessment, assessment literacy, data-driven decision making (DDDM), charter schools, and teacher education programs. Chapter three describes the research procedures and

materials, and provides a list of the research questions. Chapter four includes the results necessary to answer the research questions. A summary of the investigation and a discussion of the findings and conclusions of the study appear in chapter five.

CHAPTER TWO

LITERATURE REVIEW

Introduction

The review of literature included studies and readings in the following areas: accountability, assessment, assessment history, assessment literacy, charter schools, data-driven decision making, and instructional and school improvement practices. The purpose of this research review was to provide the reader with background information pertaining to assessment literacy in order to create a context and rationale for the current study.

This review of the literature consists of three sections. It begins with a brief historical synopsis of educational assessment development in the United States and its implications regarding assessment literacy. This is followed by a general discussion of definitional issues of assessment literacy, data-driven decision making, charter schools, and alternative teacher education. The third section highlights the current literature about school leaders' and teachers' levels of assessment literacy.

Historical Synopsis of Educational Assessment Development

Educational testing practices have been a part of the United States education system for as long as educational reforms have tried to improve public schools. As early as the 1830s, social reformers pushed for a better-developed, tax-funded, secular public school system. Horace Mann, the Father of the Common School Movement, advocated that all citizens should have equal access to a tuition-free, tax-supported public school system (Reese, 2013). Unlike today, students in the 1800s did not compete in timed,

standardized (timed, written) examinations; instead, schools' reputations rested on impressions and were assessed through exhibitions and popular community events.

By the end of the nineteenth century, an educational testing revolution occurred. Timed paper-pencil exams were now given in class where everyone tested simultaneously, followed by standard procedures comparing students based on numerical results (Clarke, Madaus, & Horn, 2000; Madaus, 1993; Madaus & Russell, 2010; Resnick & Hall, 1998). The collective belief that educational testing for public accountability would lead to school improvement and to high overall student achievement with no significant gaps was strengthened during this time (Lezotte & Snyder, 2011; Stiggins, 1991; Stiggins, 2014; Popham, 2014; Reese, 2013).

Since the beginning of the twentieth century, standardized commercial tests regarded as administratively convenient and inexpensive tools have been used to measure the achievement of students. Additionally, school reformers hoped to solve an array of educational problems using educational testing (NCME, 1997). In fact, current educational reforms still rely heavily on testing to serve a multitude of purposes: 1) to show increased rigor of school curricula; 2) to determine whether students advance or graduate; 3) to judge the effectiveness of schools and teachers; and 4) to compare districts, states, and nations (Clarke, Madaus, & Horn, 2000).

The early decades of the twentieth century were characterized by the introduction of the intelligence (IQ) test and the creation standardized commercial tests using multiple-choice format. These two test-related developments provided the impetus for the growth in standardized testing that began during this period. The development of the first intelligence test also transformed "special education" as we know it today. The IQ

test permitted educators to shift blame for poor attainment away from teaching and toward students' lack of ability to learn. Today, our schools still rely largely on aptitude or IQ testing to determine who has access to enriched programs, a curriculum that is denied to other students who are considered less capable. Thus, some students never get the chance to study a demanding curriculum with high expectations. College acceptance also depends heavily on aptitude-like tests that have little to do with the curriculum (Ali & Ali, 2010; Clarke et al., 2000; Madaus & Russell, 2010). Standardized commercial tests, regarded as administratively convenient and inexpensive tools, have been used to measure the achievement of students.

This early history of assessment parallels the history of industrialization in the United States. As an emerging industrial society in the twentieth century, the U.S. needed to educate its ethnically diverse population of students with maximum efficiency to meet the growing demands in the workforce. To accomplish this, schools were organized following the assembly-line method: a linear progression of grades where students of the same age worked their way through a formal and standardized curriculum with a one-year-per-grade fixed timeframe. The amount learned varied from student to student and schools sorted those students who would go work in factories from those students who would go to college (Ali & Ali, 2010; Clarke et al., 2000; Gerberich, 1963; Madaus & Russell, 2010; Nelson & Dawson, 2014).

The period from 1915 to 1945 is referred to as the “standardized test boom period” because tests were developed for all content areas of the school program. The belief in standardization manifested in the 1930s with the implementation of the Scholastic Aptitude Test (SAT), a standardized test widely used for college admission.

The SAT quickly became a national measure of school accountability and was viewed as the “key to productive educational change” (Stiggins, 1999, p. 192). During World War II, a considerable amount of research on the nature of human abilities was conducted by the military and standardized testing became widely accepted by society. The non-profit organization Educational Testing Service (ETS) was formed in 1947 to take over the testing activities of the American Council on Education (ACE), the Carnegie Foundation for the Advancement of Teaching (CFAT), and the College Entrance Examination Board (CEEB) (Reese, 2013b).

Much of the current assessment craze can be traced back to several significant educational reforms in the post-World War II era. In 1957, the launch of the Sputnik satellite by the former Soviet Union changed the United States’ educational landscape. The former Soviet Union’s victory resulted in the call for educational reform in the United States. To show its continued commitment to educational improvement and change, the federal government created the Elementary and Secondary Education Act (ESEA) of 1965, which distributed additional funding into primary and secondary education with the purpose of improving student achievement. Since its groundbreaking enactment, the federal government has continuously reauthorized the act. Each act was followed by another education reform or movement: the “basic skills” movement of the 1970s, the release of the National Commission on Excellence in Education’s *A Nation at Risk* in 1983, and finally, the *Goals 2000: Educate America Act* in the 1990s. In each of these “reform waves,” testing was seen as an important policy tool (DeLuca & Bellara, 2013).

The reauthorization in 2001 of the Elementary and Secondary Education Act (ESEA), also known as the No Child Left Behind act (NCLB), was intended to solve our nation's complex student achievement shortfalls by calling for increased standards-based reform and standards of accountability for states, school districts, and schools. NCLB was the beginning of a new system of accountability. One significant aspect of this reform was the establishment of its primary goal: to have all students proficient in English language arts and mathematics by 2014. Per the data from the 2007 National Assessment of Educational Progress (NAEP), minimal gains were recorded over a seven-year span and little has been done to close performance gaps between inner-city and suburban students (Duncan & Santy, 2015).

In 2009, in an effort to create a better law that focused on the clear goal of fully preparing all students for success in college and career, President Barack Obama launched the Race to the Top as part of the American Recovery and Reinvestment Act (ARRA). This effort recognized that urgent improvements in education were needed to prepare all students for a globally competitive economy and to drive change for low-income students, students of color, and other groups of students for whom educational progress had come haltingly, if at all. In 2012, the Obama administration began offering flexibility to states regarding specific requirements of the NCLB in exchange for rigorous and comprehensive state-developed plans designed to close achievement gaps, increase equity, improve the quality of instruction, and increase outcomes for all students.

Following the shift in educational standards and assessment, President Obama reauthorized the ESEA in December of 2015, 50 years after its initial enactment; it is now known as the Every Student Succeeds Act (ESSA). One central component of the current

efforts to reform public education is charter schools. While mixed feelings and little research exist regarding the charter school movement, it is set to become a major piece of educational reform, especially since the U.S. Department of Education launched a \$65 million grant competition for creating and expanding high quality public charter schools (Duncan & Santy, 2015).

Darling-Hammond (1995) summarizes the educational development and states: Today's schools were designed when the goal of education was not to educate all students well but to process a great many efficiently, selecting and supporting only a few for "thinking work." Strategies for sorting and tracking students were developed to ration the scarce resources of expert teachers and rich curricula, and to standardize teaching tasks and procedures within groups (p. 153).

Charter School—What Is It?

According to the latest study by the Center for Research on Education Outcomes (CREDO), charter schools are a pronounced and growing part of the public-school system in the United States (CREDO, 2009, Maul & McClelland, 2013). The study reports an estimated 6,400 charters across the country with more than 2.5 million students enrolled. The charter school concept developed from the economic free market theories of Milton Friedman. Friedman envisioned that the charter school system would create a market for school choice, which up until this point was lacking in the public-school sector. Charter schools are a combination of elements from both public and private schools; they do not charge tuition and receive all their funding from state and local governments, school districts, and private charitable donations. Charter schools can

receive waivers from state and local regulations, and are free to create their own curriculum, hire staff, and set budgets (Grady, 2012).

While charter schools vary drastically in student profiles, governance, and academic quality due to the varying state charter laws and policies, the charter movement agrees on its purpose to improve student achievement while eliminating the achievement gap. Zimmer et al. (2009) define charter schools as public schools approved by state entities or local boards that allow greater autonomy over curriculum, instruction, and operations. There are commonly two types of charter schools: start-up and conversion. A start-up charter school is usually governed by a charter network with a specific focus, recruits its students, and is a school of choice. Conversion schools are existing schools that due to low student performance are taken over by a government agency to raise student achievement and most often use the same school facility with already enrolled students (Zimmer et al., 2009).

Effectiveness of Charter Schools

Since charter schools are a part of education reform across the United States, it becomes more important to have current and comprehensible analysis about how well these schools educate students. Research on charter schools is not only limited, with mixed results related to effectiveness, but is also biased in terms of its accuracy (Lubienski & Weitzel, 2010). Zimmer et al. (2009) also raised questions about the validity of charter-school impact studies, suggesting that the conflicting conclusions of different studies are due to variations in charter laws and policies. One study that has experienced such controversy is Bifulco and Ladd's study of North Carolina's charter schools.

Bifulco and Ladd (2006) examined North Carolina charter schools' impact on student achievement. The analysis was based on data from the North Carolina Education Research Data Center and contains individual-level information on test scores and background characteristics for all students in grades 3 through 8 in the state's public schools, both charter and traditional. The data was collected over a five year span, from 1996 – 2000. The researchers evaluated the achievement gains of students attending charter schools versus the gains they would have made in traditional public schools.

Bifulco and Ladd suggest that when studying an entire system of charter schools, it is not possible to conduct a true experiment. The researchers used an extensive, individual-level panel data set, a method that in effect compares the test-score gains of individual students in charter schools with the test-score gains made by the same students when they were in traditional public schools. The analysis of charter school effectiveness is based on the experiences of only those students for whom the researchers observed annual gains in test scores at least once in a charter school and at least once in a traditional public school.

The findings showed that students made considerably smaller achievement gains in charter schools than they would have in traditional public schools. The researchers also concluded that the large negative estimates of the effects of attending a charter school are neither substantially biased nor substantially offset by positive impacts of charter schools on traditional public schools. Finally, Bifulco and Ladd (2003) found that about 30 percent of the negative effect of charter schools is attributable to high rates of student turnover.

In a corresponding article, Roland (2006) refers to Bifulco and Ladd's study as "flawed and fails to count what matters most to parents and students" (p. 1). Roland states that Bifulco and Ladd's study excluded longer-term charter-school attendees. She also criticizes the fact that the data used to assess performance came from state end-of-grade tests measuring knowledge of state curriculum, which does not align with innovative charter schools' alternative curricula. Lastly, she notes that Bifulco and Ladd's data differ from recent Department of Public Instruction statistics. Roland states that in 2004-05, 63 percent of regular North Carolina charter schools made adequate yearly progress under federal accountability guidelines, compared with just 58 percent of traditional public schools.

Hoxby and Rockoff's (2005) study focused on schools operated by the Chicago Charter School Foundation, a charitable organization that has been operating since 1997. It oversees five primary schools, one high school, and one K-12 school. The charter schools used in the study are all located in neighborhoods where the population is disproportionately minority, poor, and in need of bilingual education. The researchers focused on students who participated in the charter school lotteries held in spring 2000, 2001, and 2002. The Consortium on Chicago School Research agreed to match as many of these students as was possible to the Chicago Public Schools' student database using their names, dates of birth, and the school and grade they reported attending when they applied. The data provided information on achievement, as measured by the Iowa Tests of Basic Skills (ITBS), before students applied and with post-application achievement data for students who remained in Chicago's regular public schools. The data provided allowed the researchers to conduct a randomized experiment. The study compared the

achievement of lotteried-in and lotteried-out applicants through the spring of 2004, or up to four years following their initial application. The results reported were adjusted to reflect the fact that not all lotteried-in students enrolled in charter schools. They represent the effect of attending a charter school, not simply of drawing a lottery number low enough to gain admission.

The findings of the study show that students in charter schools outperformed a comparable group of lotteried-out students who remained in regular Chicago public schools by 5 to 6 percentile points in math and about 5 percentile points in reading or, when expressed as gains, 2.5 to 3 points for each year spent in the charter schools. The results are based on students who enter charter schools in kindergarten through grade 5.

Hoxby and Rockoff (2005) reference Bifulco and Ladd's study criticizing their approach of using late-grade entry student data. According to Hoxby and Rockoff . 2005, students entering schools in fifth grade or later are rare and provide an unrepresentative sample of students. They argue that it is unreliable to compare the annual gains made by students in charter schools with the gains made by the same students while attending a traditional public school, especially when drawing only on the experiences of students who were tested for at least two years in the regular public schools before attending a charter school. Because they rely on state tests that are administered for the first time in the 3rd grade, almost all the students included entered charter schools in fifth grade or later.

Chingos and West (2015) conducted a study using statewide student-level longitudinal data extract obtained from the Arizona Department of Education (AZDOE) by the Goldwater Institute. The extract contained test-score, school enrollment,

demographic, and program participation information for students in Arizona public schools from the 2005-06 through 2011-12 school years. The researchers note the difficulty of obtaining the data from the institute, which may explain the paucity of recent charter research using student-level data in Arizona.

Chingos and West used observational methods that compare students in charter schools and traditional public schools with similar characteristics and academic achievement prior to entering those schools. The analysis was focused on middle schools because the researchers could measure student achievement prior to entering middle school and it allowed the tracking of student performance on state tests for multiple years. The study measured students' baseline math and reading test scores based on their standardized test performance in the year and grade prior to their first-time enrollment in a middle school. These baseline scores served as control variables throughout the analysis. The analysis measured charter effects as the cumulative value added to the student's achievement since entering the school, not the relative test-score gains from one year to the next.

The findings indicated that on average, charter schools at every grade level have been modestly less effective than traditional public schools in raising student achievement in some subjects. The data show that each year spent in a charter middle school reduced student achievement in math by roughly two percent of a standard deviation. The researchers observed no difference between charter and TPS students in reading or writing performance, but there was a clear negative effect of charter school attendance on science scores. The researchers note that these effects are not observed consistently across the sector. Some groups of charters produce similar test-score results

as regular district schools, some produce more negative results than the average charter, and some produce more positive results. Chingos and West noted that these differences could be due to geographical location, with urban charter schools producing more positive impacts (relative to the TPS their students would otherwise attend) than charter schools in non-urban settings.

In 2009, the Center for Research on Education Outcomes (CREDO) presented a longitudinal student-level analysis of charter school impacts on more than 70 percent of the students in charter schools in the United States. Since charter schools select their focus, environment, and operations, a wide diversity exists across the sector. The study provided an overview on charter schools' impact on student academic growth. To create a national pooled analysis of the impact of charter schooling on student learning gains, a virtual twin was created based on students who matched the charter students' demographics, English language proficiency, and participation in special education or subsidized lunch programs. The resulting matched longitudinal comparison was used to test whether students who attend charter schools do better than if they had instead attended traditional public schools in their community. The outcome of interest was academic learning gains in reading and math, measured in standard deviation units. Summarized, the analysis showed that charter school students on average saw a decrease in their academic growth in reading of .01 standard deviations compared to their traditional school peers. In math, their learning lagged by .03 standard deviations on average. With regard to different subgroups, African American and Hispanic student gains were significantly lower than those of their traditional school twins. However, charter schools are found to have better academic growth results for students in poverty.

English Language Learners showed significantly better learning gains in charter schools and so did students in Special Education programs (CREDO, 2009). The CREDO report (2009) concludes that despite promising results in several states and within certain subgroups, the overall findings indicate a subset of poorly performing charter schools.

In 2013, CREDO built on its initial study of 2009; this time the study analyzed charter school performance in 27 states for reading and math. The analysis of the pooled 27 states showed that charter schools advanced their students' learning gains in reading more than traditional public schools. Improvement was also seen in charter students' academic growth in math, which was now comparable to the learning gains in traditional public schools. Related results for different student groups indicated that African-american students, students in poverty, and English language learners benefited from attending charter schools. While in both subjects the trend since 2009 is on an upward trajectory, with the relative performance of the charter sector improving each year, the charter school quality was and is uneven across the states and across schools, making it hard to make comparisons and draw conclusions among charter schools (Maul & McClelland, 2013, Raymond et al., 2013).

Uncommon Schools, a charter management organization founded in 2005, not only earned the 2013 Broad Prize for Public Charter Schools, but also outperformed other leading charter schools, Achievement First and KIPP, for having the best overall student academic performance between 2009 and 2012 (Meyer, 2014).

Alternative Education Programs

Alternative education programs such as TeachForAmerica (TFA) or The New Teacher Project (TNTP) have been, and continue to be, a controversial issue in

educational discussions. Many alternatively certified teachers tend to work in urban charter schools. TFA teachers, like many other alternative route teachers, often do not have a degree in education. Following an intensive five-week teacher-training institute, most TFA teachers learn to teach on the job. The TFA program brought teachers with excellent academic credentials to the education system. Veltri (2008) comments on how alternative licensed teachers specifically face many challenges when they begin teaching, including recent graduation from college, attending an intensive five-week training program, and teaching in urban schools. Just as some charter schools have provided alternatives within the public-school system, charter-oriented teacher preparation programs provide improved alternatives to the ways in which most teachers are educated. Traditional teacher education programs associated with larger universities have the benefit of allowing preservice teachers opportunities to deepen their content knowledge and expertise to become leaders in the field of education. Preservice teachers engage in critical and self-reflective analysis of the contexts, theories, and purposes of teaching. Charter-oriented teacher preparation programs focus on the need to quickly prepare teachers for challenging classrooms. The coaching is very prescriptive. Pre-service teachers spend more hours practicing specific moves and techniques than understanding cognitive theories of learning and child development. Many people assume that teachers at charter schools are as qualified as those in public schools, but it's noteworthy that public school teachers usually hold a bachelor's degree and are state-certified or are working towards certification. Certification means that a teacher has gone through the training required by the state, which includes student teaching and course work.

Teachers who work at a charter school may fall under more flexible certification requirements than other public school teachers.

Levine (2006) reports that the nation's teacher education programs leave teachers and principals unprepared. The report, titled *Educating School Teachers*, provides an examination of the successes and failures of university-based teacher education programs. The report identifies several model teacher education programs and critiques their outdated, historically flawed vision of teacher education. Levine states that traditional teacher education programs only introduce students to prolonged, real-life classroom settings late in their pre-service program, if at all. He goes on to argue that students in those programs lack the opportunity for immediate application and testing of learned skills, as well as familiarity with some of the types of situations they will encounter once they have their own classrooms.

Stitzlin and West (2013) analyzed charter-inspired teacher preparation programs. Their analysis highlights the lack of teacher flexibility, limited theoretical knowledge of good teaching, a loss of the overarching purpose of educating for democracy that is integral to traditional teacher colleges, and altering views of the teaching profession itself. Stitzlin and West (2013) acknowledge that charter-inspired teacher preparation programs may potentially do better at training pre-service teachers for the difficult realities of teaching in high-needs classrooms by situating their primary learning experiences in these settings from the start. One noteworthy difference of charter-inspired teacher preparation programs versus traditional teacher programs is that charter-inspired teacher preparation programs strive to produce teachers who demonstrate significant results in improving student test scores during the process of completing their

student teaching. Traditional teacher programs expect preservice teachers to spend a certain number of satisfactory hours supervised in classrooms without having to prove growth in student knowledge or skills. Banks et al. (2005) argue that many pre-service teachers of traditional programs enter the field unprepared, having had no personal experience with struggling schools and underperforming populations. Because of this, they tend to envision the types of schools that they attended as children when visualizing educational methods taught in their college classes, and they lack the immersion experiences necessary to broaden their understanding of school contexts.

Gawlik et al. (2015) conducted a study to determine whether regular public and charter school teachers in the Detroit metropolitan region differ in indicators of teacher quality. Michigan is an important state in which to study these issues given its early adoption of charter schools dating back to 1993 and their strong presence in the Detroit metropolitan region. The population of traditional public and charter school teachers from the tri-county Detroit metropolitan region was used for the study. It consisted of 26,135 teachers distributed across 794 elementary and middle schools, including 23,171 teachers in 708 traditional public schools and 2,964 teachers in 86 charter schools. All data were for the 2005-2006 school year, the most recent year for which a complete data set could be assembled. High schools were omitted due to data and school coding limitations. Descriptive statistics and tests of differences of means were used to compare teacher quality indicators in charter schools and traditional public schools while the coefficient of variation was used to assess within group variation. Pearson correlation was used to determine to what extent teacher quality indicators were associated with

teacher effectiveness; and to what extent teacher sorting took place across charter and traditional public schools.

The findings concluded that charter schools differ significantly from traditional public schools. Overall, the charter school teacher workforce was more likely to be noncertified, inexperienced, and to hold a substitute permit. Although charter school teachers were more likely to be graduates of a competitive college and to hold a major or minor in their teaching subject matter area, they left teaching at a higher rate than traditional public school teachers. Only in few cases were teacher quality indicators consistently associated with teacher effectiveness, except for the percentage of certified teachers. These ambiguous results may be, at least in part, an artifact of the use of a single year of data and associated lack of school-level, value-added estimates of teacher effectiveness (Gawlik et al., 2015).

Assessment Literacy—What Is It?

What Is Assessment Literacy?

In chapter 1, the researcher presented a working definition of assessment literacy. The following are various definitions that exist throughout assessment literacy research.

Chappius & Stiggins (2011) explained that assessment-literate educators come to any assessment knowing what they are assessing, why they are doing so, how best to assess the achievement of interest, how to generate sound samples of performance, what can go wrong, and how to prevent these problems before they occur.

Brookhart (2011) defined assessment literacy as the knowledge and skills educators need to: 1) identify, select, or create assessments optimally designed for various purposes and 2) analyze, evaluate, and use the quantitative and qualitative

evidence generated by assessments to make appropriate decisions to advance student learning.

Fullan (2002) defined assessment literacy as the educator's capacity to examine student performance evidence and discern quality work through the analysis of achievement scores and disaggregation of data. Additionally, he summarized the need for teachers to be knowledgeable in the formation of action plans aimed to increase student achievement, and related assessment literacy to educators' contributions to political agendas associated with high-stakes testing and achievement data use. Webb (2002) stated that assessment literacy is knowledge about how to assess what students know and can do, interpret the results of these assessments, and apply these results to improve student learning and program effectiveness. Lastly, Stiggins (2002) noted that assessment literacy is the understanding of the basic principles of sound assessment. Gotch and French (2014) placed assessment content and methods at the center of elements influencing student learning. They suggested that educational leaders must have the tools to improve classroom instruction, including a focus on what to attend to in improving teaching, observing classrooms, using achievement data, and considering samples of student work.

Current thinking regards assessment as more than just a method to measure student achievement, emphasizing the use of assessment practices and results as learning tools that promote instructional and effective school practices as well as enhancing student learning. Brookhart (1999a) indicated that good classroom assessment allows teachers to draw accurate inferences about individual student achievement, communicate that information to students and parents, and focus further instruction. Chappius and

Stiggins (2011) noted that assessment-literate educators (regardless of whether they are teachers, administrators, or superintendents) come to any assessment knowing what they are assessing, why they are doing so, how best to assess the achievement of interest, how to generate sound samples of performance, how to determine what can go wrong, and how to prevent these problems before they occur. DeLuca and Klinger (2010) asserted that the ineffective use of classroom assessment leads to reduced reliability and validity, consequently leading to inappropriate educational decisions.

The primary force for examining teachers' assessment literacy was the development and distribution of the 1990 Standards for Teacher Competence in Educational Assessment of Students (Sanders et al., 1990), which defined the competencies necessary for teachers to conduct sound assessments.

Initially, these 7 Standards for Teacher Competence in Educational Assessment of Students (see page 4) were developed to address the problem of inadequate assessment training for teachers. The working definition of assessment as defined by the standards 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" (Sanders et al., 1990). The 7 Standards provide criteria for teacher competence with respect to the various components of this definition of assessment.

While the Standards have served as the basis for studies of school leader and teacher knowledge regarding assessment for the last 20 years, Brookhart (2011) argued that the standards have become out dated. She suggested that to accommodate school leaders' and teachers' knowledge and skills, the Standards need to reflect the current

conceptions of formative assessments and promote the knowledge and skills needed to successfully work in the current accountability and “standards-based reform” context. Understanding formative assessment practice has developed over time, and received a major upswing when Black and Wiliam (1998) conducted a meta-analysis of more than four thousand studies on assessment undertaken during the last 40 years and concluded that formative assessment improves learning. According to Black and Wiliam, formative assessment can effectively improve student learning when implemented effectively. While formative assessment became popular in theory, it was rarely used in classroom routine (Popham, 2008).

Hattie’s (2009) meta-analysis confirms Black and Wiliam’s findings, stating that the use of formative assessments to provide descriptive feedback improves student learning. According to Stiggins (2008), Hattie (2009), and Wiliam (2010), student involvement in assessment enhances the value of classroom assessment and, ultimately, student achievement. Wiliam (2010) noted that effective student involvement thrives when teachers employ effective strategies such as clarifying and sharing learning intentions and criteria for success.

In general, the literature review indicated that preservice teachers, in-service teachers, and school leaders represent low, insufficient levels of assessment literacy (Plake & Impara, 1993; Mertler, 2003; Gotch & French, 2013). Studies yielded common areas of weakness including development of assessment methods appropriate for instructional decisions, development of valid student grading procedures that use student assessments, and communication of assessment results to students, parents, and other audiences (Plake & Impara, 1993; Quilter & Gallini, 2000; Mertler 2003, 2009).

Effectiveness of the Assessment Literacy Inventory

Studies on assessment literacy and studies that used the assessment literacy inventory were used to support the positive impact that assessment-literate school leaders and teachers can have on student learning, and confirm that a lack of assessment literacy can hinder instructional practices that foster student learning. Among the studies presented, all support the need for better assessment training as well as improved pre-service training.

Plake (1993) reported the findings of a national survey of teacher levels of assessment knowledge based on the Standards of Teacher Competency of Educational Assessment. The study sample of teachers was selected by testing coordinators within each of the selected six districts per state. The instrument was mailed to testing coordinators and a total of 555 teachers (48 percent) returned their survey responses. The data represents all but 5 states; a total of 98 districts in 45 states made up the final sample.

This study focused on levels of assessment literacy among in-service teachers. The Teacher Assessment Literacy Questionnaire (TALQ), which was the first version of the Classroom Assessment Literacy Inventory (CALI) used for the current study, is made up of two parts. Part One consisted of a 35-item multiple choice instrument, five items designed to assess each of the seven competency areas of the Standards for Teacher Competence in the Educational Assessment of Students. The seven broad skills areas in educational assessment were: 1) choosing assessment methods appropriate for instructional decisions; 2) developing assessment methods appropriate for instructional decisions; 3) administering, scoring, and interpreting the results 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 that use pupil assessments; 6) communicating assessment results to students, parents, other lay audiences, and other educators; and 7) recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (pp. 2-5). Part Two analyzes teacher perceptions and background.

The overall test results indicated that teachers answered slightly over 23 of the total 35 items correctly (66%). Out of the seven competency areas, teachers scored highest in “administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods” (mean = 3.96; maximum possible score = 5). The lowest score was recorded in the competency area of “communicating achievement” with a mean of 2.70 out of the possible five items. Ninety percent or more of the teachers answered 10 of the 35 items correctly. These items focused on areas such as making assessment selection decisions, teacher knowledge of acceptable test taking behavior for timed tests, and explanation of a child’s grade to parents. On the other hand, less than 30% of teachers answered items correctly in areas of “using assessment in grading” and “recognizing unethical practices standards.”

Part Two of the national survey analyzed teacher perceptions and background. Plake (1993) reported that most teachers had between six and 12 years of teaching experience. Only few teachers had less than five years of experience or more than 15. Over 85% of teachers responded that they felt positive and comfortable about the use of teacher-made tests for instructional decisions, but only 34% felt positive and comfortable

using standardized tests in this way. Per the study, over 70% of teachers reported prior exposure to tests and measurement content, but many teachers stated that the content was irrelevant for their current use of assessment practices. Over 85% of teachers preferred to receive additional training through in-service mode of delivery, requesting that it be relevant to the current needs of assessment practices. Based on the results of this national survey, the competency area of “communicating and interpreting test results” was given priority and training models were designed to foster teacher competency in this standard.

Campbell, Murphy, & Holt (2002) conducted a similar study using the renamed Assessment Literacy Inventory (ALI). The inventory was administered to 220 undergraduate pre-service teachers following completion of course work in tests and measurement. The pre-service teachers averaged 21 questions correctly out of the 35 multiple-choice items, two fewer correct answers than the in-service teachers. The data from the preservice teachers showed a higher level of reliability (.74) than data from the in-service teachers in the Plake study (Campbell, Murphy, & Holt, 2002). The in-service teachers scored higher on 6 out of the 7 Standards. Standard 1 was the only Standard that received a lower score. The pre-service teachers scored highest on Standard 1, Choosing Appropriate Assessment Methods. The in-service and pre-service teachers scored lowest on Standard 6, Communicating Assessment Results (Mertler, 2003).

Another study conducted in the fall of 2002 by Mertler (2003) surveyed both pre-service and in-service teachers with respect to their assessment literacy. Mertler compared the two groups statistically. The study surveyed 67 pre-service teachers and 197 in-service teachers. The two groups were surveyed using an instrument called the Classroom Assessment Literacy Inventory (CALI). The CALI was adapted from the

Teacher Assessment Literacy Questionnaire used by Plake in 1993. The data set included 67 pre-service teachers (N=67) and demonstrated a reliability of .74. On average, pre-service teachers answered slightly less than 19 out of 35 items correctly. Again, the highest score was on Standard 1, Choosing Appropriate Assessment Methods, while the lowest score was on Standard 5, Developing Valid Grading Procedure.

The data from the in-service teachers (N=197) demonstrated a reliability of .57. The pre-service teacher results showed higher reliability than the in-service teacher results. The in-service teachers averaged less than 22 out of 35 items correctly. The in-service teachers scored highest on Standard 3, Administering, Scoring, and Interpreting the Results of Assessments, and scored lowest on Standard 5, Developing Valid Grading Procedures.

Scores for the two groups of teachers on each standard as well as the total scores were compared by conducting independent sample t-tests. The analyses showed that significant differences existed between the two groups on 5 of 7 standards and for the total score. In each case where there was a significant difference, the in-service teacher scored significantly higher than the pre-service teachers. The survey supported the data from the earlier studies in that in-service teachers showed a higher level of assessment literacy than pre-service teachers.

Both researchers, Campbell et al. (2002) and Mertler (2003), recommended revisions of the assessment literacy instrument to accurately measure teachers' levels of assessment literacy. The revised assessment literacy instrument was piloted during the fall of 2003. One hundred fifty-two undergraduate pre-service teachers from two large Midwestern institutions participated in the study.

Following the item analysis, Campbell et al. and Mertler made appropriate revisions to items on the ALI. The second phase of data collection occurred in the spring of 2004, with 250 undergraduate pre-service teachers following their completion of tests and measurement coursework. Analysis of the item results from this phase disclosed similar psychometric properties. The reliability was equal to KR20 (rKR20) .76; the mean item difficulty was equal to .681, and the mean item discrimination was equal to .313.

The data revealed a mean score of 23.83 out of a possible 35 points; approximately 68% of items were answered correctly. The pre-service scores were far lower than expected given their recent completion of coursework in classroom assessment. The researchers concluded that possible reasons may be related to pre-service teachers' limited classroom experience and the close alignment of the ALI to real-world applications of assessment concepts and competencies outlined in the Standards of Teacher Competency of Educational Assessment.

Mertler and Campbell (2005) administered the ALI again, this time to a small sample of in-service teachers ($n = 7$) following an intensive two-week professional development training course on classroom assessment. The average scores of in-service teachers were much higher (mean = 28.29), at approximately 81% of items answered correctly. While the sample was too small to stress a link between teaching experience and assessment competency, the study highlights an important aspect that cannot be overlooked.

Volante and Fazio (2007) examined the assessment literacy of preservice teacher candidates within a large Canadian urban setting. Candidates from each year of the

program completed a survey pertaining to self-described level of assessment literacy, main purposes of assessment, utilization of different assessment methods, need for further training, and suggested methods for promoting assessment literacy in university and practice teaching settings. The sample consisted of 69 teacher candidates.

Respondents ranged from 19 to 51 years of age, and had acquired classroom experience ranging from 0 to 10 years during practice teaching placements and previous work as an early childhood educator, teaching assistant, and/or private school teacher. Using a convenience sample, the researchers asked participants to complete a survey during the final stages of their academic year. By this time, pre-service candidates had completed their practice teaching assignments and were preparing for final exams.

The survey consisted of a series of open- and closed-ended questions in four broad areas: self-described level of assessment literacy, main purposes of assessment, utilization of different assessment methods and need for further training, and suggested methods for promoting assessment literacy in university and practice teaching settings. Collectively, these four areas provided a broad understanding of the perspectives of teacher candidates. The responses that aligned with more recent formative and student metacognitive purposes were noted to a lesser extent. These findings are troubling given that assessment for learning has shown the greatest promise for improving student learning and achievement within schools.

The survey identified a need for further training in authentic assessment approaches that typically support newer conceptions of assessment and performance assessment. These more recent forms of authentic assessment are increasingly seen as important for promoting improved student learning and achievement (Hattie, 2011;

Stiggins, 2005). More established traditional methods, observation, communication, and selected response assessments were ranked relatively lower for further training.

It was the noticeable absence of a balance of classroom assessment techniques that best characterized most teacher candidates' responses. This lack of balance occurred even though all teacher candidates in their second, third, or fourth year had completed a course that presumably addressed a range of observation techniques, documentation procedures, authentic assessment, and formative and summative evaluation procedures.

To improve their assessment knowledge, teacher candidates overwhelmingly endorsed the development of a specific course focused on classroom assessment and evaluation. Interestingly, this suggestion came from all students – even second, third, and fourth year students who had completed the previously noted observation and evaluation course.

Beziat and Coleman (2015) conducted a longitudinal study measuring pre-service teachers' knowledge of standards-based classroom assessment strategies during their teacher preparation. The participating pre-service teachers were enrolled in a small university in the Southeast part of the United States. The Teacher Assessment Literacy Questionnaire (TALQ) by Plake & Impara (1992) was used to assess the pre-service teachers' levels of assessment literacy.

The TALQ was administered four times during teacher training. Forty-nine students completed the pre-test of data collection, but only 26 of those students completed the post test. No significant differences were found between the pre-test and post-test results. Additionally, scores were compared based on the different programs of study (early, elementary, and secondary) and whether or not students were already

admitted to the teacher education program. Pre-service teachers in the secondary programs scored higher on both administrations of the test in comparison to those in the early and elementary programs. Pre-service teachers who were admitted to the school of teacher education program outperformed those who were not admitted to the teacher education program.

The TALQ results revealed that pre-service teachers lack assessment literacy, despite completing courses in classroom assessment. The results also indicated that preservice teachers were not improving in assessment literacy. Specifically, their knowledge of classroom assessment practices seemed to remain the same over the course of the semester.

Butty (1998) conducted a study to assess how teachers and university faculty perceive the need for and importance of professional development in performance-based assessment. This study surveyed 19 middle school mathematics teachers and 18 school of education faculty members. The study was conducted in an urban school district (Washington, D.C.) that was part of a larger effort to improve teachers' performance and personal efficacy in performance-based assessment.

The study defined performance-based assessment as a type of assessment where students typically engage in the use of physical materials and the construction of responses to performance tasks. The instrument used for the study was a 22-item questionnaire, the Performance-Based Educational Experiences Survey. Twenty items addressed the extent to which teachers and university faculty believe that performance-based assessment experiences are important to or needed by teachers. The participants were asked to rate their responses to the statements on two 5-point Likert-type scales

ranging from 1 (not important) to 5 (very important) and, correspondingly, from 1 (not needed) to 5 (highly needed). The results indicated that teachers perceived a great need for, and interest in, understanding different types of assessment, aligning classroom instruction with performance-based education, and understanding the fundamental concepts of measurement. A large proportion of the education faculty indicated that professional development in the area of performance-based assessment was needed, and their responses for higher-order thinking skills and scoring rubrics were lower than those of the teachers. The results of the survey suggested that professional development activities planned by university faculty might generally be improved and more useful for teachers if they were based on targeted input from teachers.

Per Johnson et al. (1998), teachers will feel more comfortable and efficacious in developing and carrying out performance-based instruction and assessments when trained in ways that they deem helpful to them.

Data-Driven Decision Making (DDDM)

Data-driven decision making (DDDM) is defined as a process of systematically collecting various types of assessment data and analyzing the data to guide a range of decisions to help improve student achievement (Ikemoto & Marsh, 2007; Ingram, Louis, & Schroeder, 2004). Educational reforms have placed increasing pressure on state departments of education, districts, and schools to improve the way school leaders and teachers use data to make data-driven decisions. Everyone from policymakers to school leaders and teachers has moved towards a culture of being data-driven, and has declared data use in schools to be significant for school improvement and accountability (Ikemoto & Marsh, 2007). As states and school districts make great strides in creating a culture of

data-driven decision making—collecting, analyzing, and interpreting data—individual school leaders and teachers struggle to make sense of all the data and how to effectively use the data to make instructional decisions.

President Obama’s American Recovery and Reinvestment Act of 2009 and the reauthorization of the Elementary and Secondary Education Act in 2011 reiterated the importance of data-driven decision making, with high quality assessments and data analysis in schools as the leading indicator for educational reform. Race to the Top, the federal grant competition, was another major effort by the administration to create data-rich cultures in schools. The competition encouraged states to build comprehensive data systems that measure student success and inform school leaders and teachers about how they can improve their practices. Because of such initiatives, the education community witnessed increased interest in data-driven decision making (Duncan & Santy, 2015).

The big assumption underlying data use in schools is that results from state, district, and local assessments will be used to enhance decisions about how to allocate resources and improve student learning and classroom practice (Ikemoto & Marsh, 2007; Ingram, Louis, & Schroeder, 2004). Hess and Kelly (2007) argued that “data-driven decision making does not simply require good data; it also requires good decisions” (p. 17). Ikemoto and Marsh (2007) stated, “These calls for data-driven decision making often imply that data use is a relatively straightforward process. As such they fail to acknowledge the different ways in which practitioners use and make sense of data to inform decisions and actions” (p. 105). While educators claim to be data-driven, systematically collecting and analyzing various types of data to guide a range of decisions to help improve student achievement, a 2006 RAND report investigated the

types of data available to administrators and teachers as well as the factors influencing data use, and concluded that data-driven decision making does not guarantee effective decision-making (Marsh et al., 2006). And while data use in schools has played a prominent role in district efforts to provide systematic change and improve student achievement, these efforts assume that the classroom teachers are fully equipped with a repertoire of methods and skills in data analysis and know how to use the data to improve student learning.

The importance of teachers knowing how to use assessment data to improve student learning cannot be overestimated. The effective use of high-quality assessment data by classroom teachers can improve instruction (Stiggins, 2002; Popham, 2011). Fullan, Hill, and Crevola (2006) contended that to improve the effectiveness of classroom instruction, teachers need to become proficient in using assessment to monitor and manage student learning. The notion is that by supporting teachers with effective use of data analysis, they can make informed decisions about instruction. Stiggins (2002) suggested that it is essential for schools to build an assessment-literate faculty and focus on effective data analysis use.

Summary

Research has linked assessment literacy with notable benefits in students' learning and improvements in teaching. This chapter provided a historical synopsis of educational assessment development and a description of the charter school movement. Details about the benefits of assessment literacy for the use of data-driven decision making were identified, and their essential nature in today's standards-based education system was explained.

Findings from studies that support the need for assessment-literate educators were discussed, and studies that used the TALQ, CALI, and ALI to determine levels of assessment literacy among school leaders and in-service and pre-service teachers were reviewed.

Chapter three contains the description of the research procedures and materials, and a list of the research questions. Chapter four includes the results necessary to answer the research questions. A summary of the investigation and a discussion of study findings and conclusions appear in chapter five.

CHAPTER THREE

METHODOLOGY

Problem

The purpose of this non-experimental (descriptive) quantitative study was to determine the assessment levels of elementary school leaders and teachers. Each group was classified according to faculty position (school leader and teacher) and type of teacher education program. In a traditional teacher education program, teachers earn certification by completing a bachelor's or master's degree in education, taking Praxis tests, and fulfilling additional state requirements (Justice et al., 2003). An alternative teacher education program is the process by which a person is awarded a teaching license even though that person has not completed a traditional teacher education program. Alternative certified teachers possess a bachelor's degree in a different field and are completing (or have completed) an alternative education program while teaching full-time (Justice et al., 2003).

This non-experimental quantitative study used a descriptive research method design. Descriptive research is primarily concerned with finding out "what is" and allows the researcher to describe the data with respect to social implications. According to Knupfer and Mclellan (1996), descriptive research plays an important role in educational research because it produces statistical information about aspects of education that can be described in greater or less depth depending on the interest of the researcher, policymakers, and educators. The National Assessment of Educational Progress (NAEP), which collects descriptive information about how well the nation's youth are doing, is an example of descriptive research that heavily impacts educational

decision making in the United States. Glass & Hopkins (1984, as cited by Knupfer & Mclellan) note that descriptive research involves gathering data that describes events and then organizes, tabulates, depicts, and describes the data collection. Borg & Gall (1989, as cited by Knupfer & Mclellan) note two critical components of descriptive research: sound methodology and a well-designed data collection instrument.

The identification of school leaders' and teachers' levels of assessment literacy provides guidance to the charter network and individual school campuses to develop adequate school leader and teacher professional development and promote assessment literacy. Brookhart (1999) noted the need for school leaders and teachers to be assessment-literate in order to make sound data-driven decisions. Assessment-literate school leaders can promote and develop effective schools, while supporting teachers with their instructional practices. Assessment-literate teachers can improve their instructional practices and increase students' levels of achievement (Brookhart, 1999a; Mertler, 2004; Stiggins, 1999a).

Population

The target population was approximately 200 educators from elementary charter schools within the same charter network. The elementary schools were located in an urban area in Tennessee and in an urban district in California. All of the approximately 200 educators were invited to participate in taking the questionnaire. Eighty-six of the 200 educators attempted the survey but only a total of 63 participants completed the entire survey. Participants who did not complete the survey were excluded from the data analysis. Table 1 summarizes the participants' faculty positions. A total of 58 teachers and five school leaders completed the questionnaire: five school leaders, 12 kindergarten

teachers, 5 first grade teachers, 6 second grade teachers, 10 third grade teachers, 8 fourth grade teachers, 6 fifth grade teachers, one special area teacher, and ten special education teachers, academic coaches, and interventionists.

Table 1

Participants' Faculty Position

School Leaders & Teacher Participants	Responses
School leaders	5
Kindergarten	12
First Grade	5
Second Grade	6
Third Grade	10
Fourth Grade	8
Fifth Grade	6
Special Area (Drama, PE, Art)	1
Academic Coaches, Special Education Teachers, Interventionists	10
Total	63

Due to the anonymous nature of the questionnaire, the data does not reveal the school of each participant. Although all schools are different, even within the same school district and charter network. All schools were located in large urban areas and served disadvantaged children, including a large percentage of students (90% to 98%) who receive free and reduced cost lunches. Student populations ranged from about 400 to 550.

Instrument

The instrument used for this research study was the Classroom Assessment Literacy Inventory (CALI) (Mertler, 2003). The CALI is a revised version of the Teacher Assessment Literacy Questionnaire (TALQ) initially developed by Impara, Plake, and

Fager in 1993 as part of a grant from the W.K. Kellogg Foundation to the NCME to develop prototype materials to help teachers become more proficient in assessing student performance. The inventory is commercially available through the National Council on Measurement in Education (NCME).

In this study, the inventory consisted of two parts. Part one collected demographic information. Participants were asked for information regarding current position, highest completed educational degree, years in education, number of years since completion of a teacher education program, experience in a stand-alone course in assessment, and type of teacher education program. Part two used Mertler's Classroom Assessment Literacy Inventory (CALI). The CALI was designed to find the level of assessment knowledge of teachers. The CALI is a 35-item measure of teachers' level of assessment literacy. The online survey presented participants with 35 items; each item was an assessment-related scenario, followed by a question with a specific correct answer. All items used a multiple-choice format with four options, one of which was the correct response. Table 2 delineates the items on the CALI as they relate to the standards. Each standard is aligned with five items throughout the inventory.

Table 2

Alignment of Standards with CALI Items

Standards for Teacher Competence	CALI Item Numbers
Standard 1- Choosing Appropriate Assessment Methods	1, 8, 15, 22, 29
Standard 2- Developing Appropriate Assessment Methods	2, 9, 16, 23, 30
Standard 3- Administering, Scoring, and Interpreting the Results of Assessments	3, 10, 17, 24, 31
Standard 4- Using Assessment Results to Make Decisions	4, 11, 18, 25, 32
Standard 5- Developing Valid Pupil Grading Procedures	5, 12, 19, 26, 33
Standard 6- Communicating Assessment Results	6, 13, 20, 27, 34
Standard 7- Recognizing Unethical or Illegal Practices	7, 14, 21, 28, 35

Note: From Secondary Teachers' Assessment Literacy: Does Classroom Experience Make a Difference? by Craig A. Mertler, 2004, p. 55.

Data Collection

Before contacting the charter school network and the charter school principals regarding participation in this study, the researcher submitted the required materials to the Institutional Review Board (IRB) at Middle Tennessee State University (MTSU). Upon receipt of IRB approval (Appendix A), an electronic letter was sent to the chief executive officer (CEO) of the charter network, requesting permission to conduct this study within the public charter network of elementary schools. A copy of the Participation Invitation (Appendix B) and the survey (Appendix C) were provided to the CEO.

Upon receiving permission from the CEO, an initial email was sent to the principals in late May 2016, inviting each school leader to preview the survey to

determine participation of faculty members. The purpose of the study, importance of voluntary participation, and confidentiality assurance were included in this correspondence. Upon principals' consent, a second email was sent to the principals directing them to forward the survey to the target population. This second email contained the letter of participation to school leaders and teachers and included a live link to the survey. Survey responses were then collected for a six-month period for each school.

The quantitative design of this study included an online survey of participants. The inventory was administered through Survey Monkey, an online software program. Survey Monkey was chosen for several reasons: it has multiple layers of security and firewalls, data can be downloaded in multiple forms and directly into SPSS, respondents can be tracked, and service is available to the researcher at minimal cost.

Procedure

The researcher requested consent from the charter network to conduct the study and discussed the nature of the study with administrators. After consent was given, the elementary school leaders were contacted by email to explain the study and request consent to invite teachers and school leaders to participate in the study. After consent was granted, the faculty was contacted by email to explain the study and invite their participation. An electronic link to the Classroom Assessment Literacy Inventory was emailed to all participants. A paragraph at the beginning of the questionnaire stated that taking the survey implied the participant's consent to the study.

Data Analysis

Part one of the survey consisted of demographic questions, specifically asking each respondent to give information regarding years of experience in teaching and/or years in administration, education level, assessment coursework, number of years since the completion of a teacher and/or principal preparation program, and type of teacher preparation program. The years of experience allowed the data to reflect the amount of time spent in the classroom and/or as a school leader. The education level also allowed the data to show how much education has been attained after the initial teacher preparation program. The question regarding assessment coursework determined whether the teacher and school leader took any courses specifically in assessment. The question regarding number of years since the teacher and/or principal education program showed the amount of time lapsed since the formal training setting, and the type of teacher training program provided information regarding traditional and alternative teacher education programs.

Descriptive analyses at the individual item level included frequencies and variability analyses; descriptive analyses were also conducted for the seven composite scores based on the standards. The data was used to determine the level of school leaders' and teachers' assessment literacy. The results are reported in chapter 4 and are used to answer the following research questions.

1. What is the level of assessment literacy of public charter school teachers and school leaders as measured by the Classroom Assessment Literacy Inventory?

2. Are there differences in level of assessment literacy among teachers and school leaders related to their years of experience (zero to five years and six plus years), level of education, and intensity of assessment training?
3. How does the assessment literacy of teachers and school leaders educated in a traditional education program compare to that of teachers educated in an alternative education program?

Summary

The introduction and background for the problem were presented in chapter one. The problem was stated and the need for the study, definitions, procedures, theoretical framework, and limitations were discussed. Chapter two analyzed the relevant literature. Chapter three described the research procedures and materials, and a list of the research questions. Chapter four presents the results necessary for answering the research questions. A summary of the investigation and a discussion of the findings and conclusions of the study appear in chapter five.

CHAPTER FOUR

RESULTS

Introduction

The purpose of the study was to investigate assessment literacy among elementary teachers and school leaders in charter schools. First, a description of the study sample is presented. Second, the quantitative results (demographic results and CALI results) are presented, which include variables of teachers' and principals' assessment literacy. The findings of the study were guided by the following research questions:

1. What is the level of assessment literacy of public charter school teachers and school leaders as measured by the Classroom Assessment Literacy Inventory?
2. Are there differences in level of assessment literacy among teachers and school leaders related to their years of experience (0 to 5 years and more than 5 years), level of education, and intensity of assessment training?
3. How does the assessment literacy of teachers and school leaders educated in a traditional education program compare to that of teachers educated in an alternative education program?

A survey, the CALI, was delivered electronically to approximately 200 educators in the charter school network. Eighty-six educators (43%) attempted the survey, but only 63 participants (31.5%) completed the entire survey. The initial time frame of four months was extended to six months to achieve a higher participant rate. The researcher reached out to school leaders and provided reminder emails to be distributed to faculty. Two out of five principals assured the researcher that time was given to teachers during weekly professional development to complete the survey. Participants who did not

complete the survey were excluded from the data analysis. The sample was composed of teachers (n=58) and school leaders (n=5) who represent five elementary charter schools within two urban districts across two states, Tennessee and California.

Part I Survey Results—Demographic Information

The demographic section of the survey sent to participants solicited the following information:

- a) Current position/grade level;
- b) Highest completed educational degree;
- c) Years in education;
- d) Number of years since completion of a teacher education program;
- e) Experience in a stand-alone course in assessment;
- f) Type of teacher education program.

The current faculty position was provided in Table 1 of chapter 3. A summary of the remaining demographic findings is shown in Table 3.

The results regarding the level of education obtained by teachers and school leaders were split evenly between 29 teachers with a bachelor's degree and 29 teachers with a master's degree. Among the surveyed school leaders, three held bachelor's degrees, one a master's degree, and one a doctoral degree.

The results of participants' years of teaching experience showed a more diverse distribution. Thirty-six teachers and three school leaders had five or fewer years' teaching experience. Only one teacher and two school leaders had six to 10 years of teaching experience. 21 teachers had more than 11 years of teaching experience but no school leader had more than 10 years of teaching experience.

Only 27 of the 58 teachers had completed their teacher education program in the last five years or less. All five school leaders had completed their teacher and/or school leader preparation program in the last 10 years or less. Three school leaders fell in the zero to five-year category and two school leaders fell between six and 10 years of teacher/school leader program completion. Nine teachers had completed their teacher preparation program in the last six to 10 years, while 22 teachers had completed their program in the last 11 or more years.

To the interest of the researcher the demographic findings included whether the participants had taken a stand-alone course in assessment (a specific higher education course in assessment) during their teacher/principal education program. The results showed that 41 teachers—over 70%—had not taken a stand-alone course in assessment, nor had four out of the five school leaders. Seventeen teachers and one school leader marked that they had attended a stand-alone course in assessment, but the researcher is unaware of the specific content and quality of the assessment course.

One of the demographic questions asked survey participants to indicate whether they had attended a traditional teacher education program or an alternative teacher education program. The numbers in Table 3 represented by the teachers and school leaders show an almost even split between those who attended a *traditional teacher education program* and those who attended an *alternative teacher education program* (see Definition of Terms, page 14).

Table 3

Summary of Teachers' and School Leaders' Demographic Data

	Teacher (n=58)	School leaders (n=5)
Highest completed educational degree	29 Bachelor degree 29 Master degree	3 Bachelor degree 1 Master degree 1 Doctoral degree
Years in education	0-5 years = 36 6-10 years = 1 11+ years = 21	0-5 years = 3 6-10years = 2 11+ = 0
Number of years since completion of a teacher education program	0-5 years = 27 6-10 years = 9 11+ years = 22	0-5 years = 3 6-10 years = 2 11+ years = 0
Experience in a stand-alone course in assessment	17 attended stand-alone course 41 did not attend stand-alone course	1 attended stand-alone course 4 did not attend stand-alone course
Type of teacher education program	27 Traditional Teacher Ed. Program 31 Alternative Teacher Ed. Program	3 Traditional Teacher Ed. Program 2 Alternative Teacher Ed. Program

Part II Survey Results–CALI

The participants' responses were evaluated for correctness, with a correct response given a value of 1 and an incorrect answer given a value of 0. Tallying the total number of correct responses for the five questions derived a composite score for each standard. Means approaching 5 indicated a greater knowledge for each specific standard. This procedure is the same as that used by Plake et al. (1993) in the national administration of the Teacher Assessment Literacy Questionnaire (TALQ) and the administration of the Assessment Literacy Inventory (ALI) by Mertler and Campbell (2005). The instrument used for this study, the CALI, was adapted from the TALQ by

Mertler in 2003 and renamed the Classroom Assessment Literacy Inventory. The results that follow are presented by each individual research question.

Research Question 1

What is the level of assessment literacy of public charter school leaders and school teachers as measured by the Classroom Assessment Literacy Inventory?

Research question one was designed to find the level of assessment literacy for the charter network educators as measured by the CALI. Initially the researcher thought to evaluate levels of assessment literacy for school leaders and teachers, but since only five school leaders participated in the survey the results are reported for teachers and school leaders combined (n=63). On average, teachers and school leaders answered 20 (M=20.23) out of 35 items correctly. Out of the seven competency areas, as delineated by the *Standards for Teacher Competence in the Educational Assessment of Students*, the highest overall performance for teachers and school leaders was found for Standard 4, Using Assessment Results to Make Decisions (M=3.65; maximum possible score = 5). The lowest performance was found for Standard 7, *Recognizing Unethical or Illegal Practices* (M=1.44; maximum possible score = 5). The results for the teachers and school leaders for each of the seven standards are presented in Table 4.

Table 4

Means and Standard Deviations for Teachers and School Leaders by Standard and Total on CALI

Standard	Mean	Standard Deviation
Standard 1 <i>Choosing Appropriate Assessment Methods</i>	3.44	0.91
Standard 2 <i>Developing Appropriate Assessment Methods</i>	3.51	0.98
Standard 3 <i>Administering, Scoring, and Interpreting the Results of Assessments</i>	2.68	1.15
Standard 4 <i>Using Assessment Results to Make Decisions</i>	3.65	1.12
Standard 5 <i>Developing Valid Pupil Grading Procedures</i>	2.86	0.99
Standard 6 <i>Communicating Assessment Results</i>	2.69	0.89
Standard 7 <i>Recognizing Unethical or Illegal Practices</i>	1.4	0.98
Total Score	20.23	4.01
Note: n=63		

Teachers and school leaders answered five out of all the 35 questions on the CALI inventory with greater than 90%. Two of these questions came from Standard 1, Choosing Appropriate Assessment Methods, and two from Standard 2, Developing Appropriate Assessment Methods. One question that received greater than 90% accuracy

came from Standard 6, Communicating Assessment Results. Out of the 35 total questions, five questions had 25% or fewer participants choose the correct answer. Two of the questions came from Standard 7, Recognizing Unethical or Illegal Practices; one item each came from Standard 1, Choosing Appropriate Assessment Methods, and Standard 2, Developing Appropriate Assessment Methods.

Research Question 2

Are there differences in level of assessment literacy among teachers and school leaders related to their years of experience (zero to five years and more than five years), level of education, and intensity of assessment training?

Research question two was designed to analyze differences in level of assessment literacy among teachers and school leaders related to their years of experience (zero to five years and six plus years), level of education, and intensity of assessment training. Table 5 compares the mean and standard deviation of teachers and school leaders with five or less years of experience to teachers that have six or more years of experience. Thirty-nine participants had five or less years of experience and 24 participants had six or more years of experience. Teachers and school leaders with five or less years of experience scored an average of 19.44 items correctly, while teachers and school leaders with six or more years of experience scored an average of 20.91 items correctly. By conventional criteria, this difference is considered to be not statistically significant. The two-tailed P value equals 0.6056. Teachers and school leaders with six or more years of experience scored highest on Standard 4 (M=3.92; maximum possible score = 5). The lowest standard for both groups was Standard 7, Recognizing Unethical or Illegal Practices (M=1.33 and M=1.60; maximum possible score = 5). On average, teachers

with six or more years scored higher than teachers and school leaders on five out of the seven standards.

Table 5

Means and Standard Deviations for Teachers and School Leaders by Years of Experience

Assessment Literacy Standard	Mean & Standard Deviation	0-5 years' experience (n=39)	6 or more years' experience (n=24)
Standard 1 <i>Choosing Appropriate Assessment Methods</i>	Mean	3.36	3.48
	SD	0.91	0.93
Standard 2 <i>Developing Appropriate Assessment Methods</i>	Mean	3.44	3.35
	SD	1.04	0.87
Standard 3 <i>Administering, Scoring, and Interpreting the Results of Assessments</i>	Mean	2.46	2.97
	SD	1.07	1.19
Standard 4 <i>Using Assessment Results to Make Decisions</i>	Mean	3.41	3.92
	SD	1.14	0.99
Standard 5 <i>Developing Valid Pupil Grading Procedures</i>	Mean	2.87	2.76
	SD	0.89	1.17
Standard 6 <i>Communicating Assessment Results</i>	Mean	2.56	2.83
	SD	0.91	0.83
Standard 7 <i>Recognizing Unethical or Illegal Practices</i>	Mean	1.33	1.60
	SD	1.00	0.92
Total (n=63)	Mean	19.44	20.91
	SD	3.83	3.98

The results for the comparison of teachers and school leaders by level of education are shown in Table 6. Thirty-two of the survey participants had a bachelor's degree and 32 participants had a masters' degree. Teachers and school leaders with a bachelor's degree scored an average of 19.86 items correctly (M=19.86; maximum score is 35 items correct). Teachers and school leaders with a master's or higher degree scored an average of 19.88 items correctly (M=19.88; maximum score is 35 items correct). By conventional criteria, the difference between the types of degree is considered to be not statistically significant. The two-tailed P value equaled 0.9753. Teachers and school leaders with a bachelor's degree scored higher on Standard 2, Developing Appropriate Assessment Methods; Standard 5, Developing Valid Pupil Grading Procedures; Standard 6, Communicating Assessment Results; and Standard 7, Recognizing Unethical or Illegal Practices. Both groups scored lowest on Standard 7.

Table 6

Means and Standard Deviations for Teachers and School Leaders by Level of Education

Assessment Literacy Standard	Mean & Standard Deviation	Bachelor's (n=32)	Master's + (n=31)
Standard 1 <i>Choosing Appropriate Assessment Methods</i>	Mean	3.33	3.56
	SD	0.79	0.98
Standard 2 <i>Developing Appropriate Assessment Methods</i>	Mean	3.58	3.47
	SD	1.02	0.92
Standard 3 <i>Administering, Scoring, and Interpreting the Results of Assessments</i>	Mean	2.72	2.75
	SD	1.16	1.11
Standard 4 <i>Using Assessment Results to Make Decisions</i>	Mean	3.58	3.75
	SD	1.22	1.02
Standard 5 <i>Developing Valid Pupil Grading Procedures</i>	Mean	2.88	2.75
	SD	0.85	1.16
Standard 6 <i>Communicating Assessment Results</i>	Mean	2.72	2.65
	SD	0.94	0.83
Standard 7 <i>Recognizing Unethical or Illegal Practices</i>	Mean	1.47	1.44
	SD	0.99	0.98
Total (n=63)	Mean	19.86	19.88
	SD	3.88	4.06

The results in Table 7 show the comparison by intensity of assessment training.

Teachers and school leaders were asked about having received a stand-alone course in assessment or not. Eighteen teachers and school leaders had taken a stand-alone course

in assessment while 45 survey participants did not. Teachers and school leaders who had taken a stand-alone course scored an average of 19.71 items correctly. Teachers and school leaders who had not taken a stand-alone course in assessment scored an average of 19.66 items correctly. Teachers and school leaders who had taken a stand-alone course in assessment scored higher on all seven standards, but the difference between both groups is considered not statistically significant due to a two-tailed P value of 0.7049.

Table 7

Means and Standard Deviations for Teachers and School Leaders by Intensity of Assessment Training

Assessment Literacy Standard	Mean & Standard Deviation	Stand-alone course (n=18)	No stand-alone course (n=45)
Standard 1	Mean	3.5	3.46
<i>Choosing Appropriate Assessment Methods</i>	SD	1	0.92
Standard 2	Mean	3.5	3.45
<i>Developing Appropriate Assessment Methods</i>	SD	1	1.04
Standard 3	Mean	2.75	2.52
<i>Administering, Scoring, and Interpreting the Results of Assessments</i>	SD	1.02	1.11
Standard 4	Mean	3.65	3.55
<i>Using Assessment Results to Make Decisions</i>	SD	1.09	1.14
Standard 5	Mean	3.1	2.73
<i>Developing Valid Pupil Grading Procedures</i>	SD	1.165	0.91
Standard 6	Mean	2.8	2.6
<i>Communicating Assessment Results</i>	SD	0.77	0.91
Standard 7	Mean	1.2	0.91
<i>Recognizing Unethical or Illegal Practices</i>	SD	0.95	1.35
Total	Mean	19.71	19.66
	SD	4.02	3.90

Research Question 3

How does the assessment literacy of teachers educated in a traditional education program compare to that of teachers educated in an alternative education program?

Figure 1 shows the total mean score comparison of teachers and school leaders educated in traditional and alternative teacher education programs as measured by the CALI. The teachers and school leaders educated in traditional teacher education programs (M=20.78; maximum possible score = 35) scored slightly higher than teachers and school leaders from alternative teacher education programs (M=20.32; maximum possible score = 35). By conventional criteria, T-test analysis showed the difference in Mean scores to be not statistically significant (p value equals 0.6598). The mean of traditional teacher education program minus alternative teacher education program equals 0.4600.

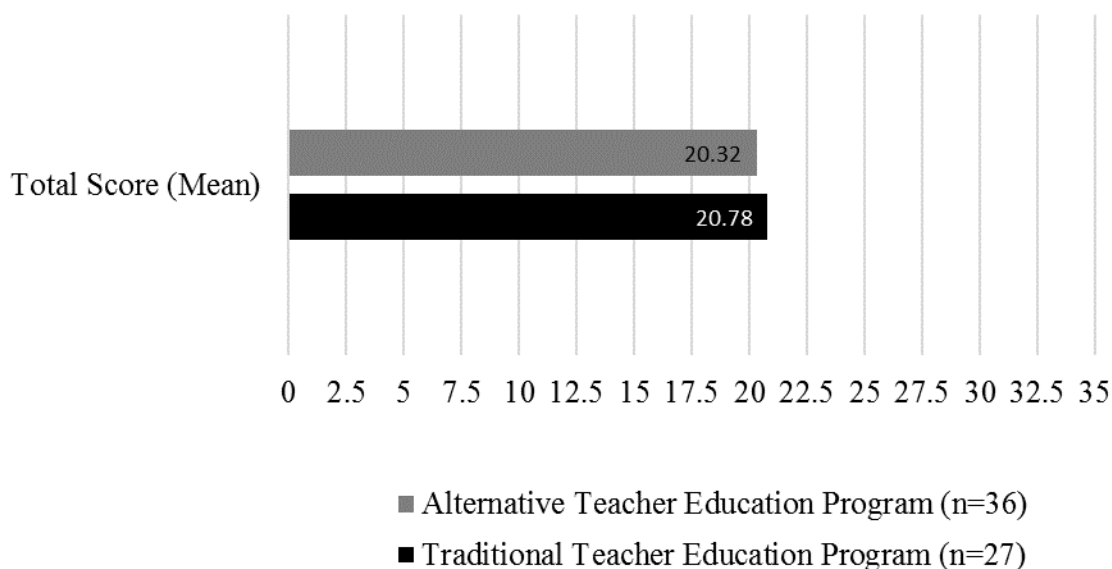


Figure 1. Total mean score comparison, traditional versus alternative teacher education program

Out of the seven competency areas as delineated by the Standards, the highest overall performance for traditionally educated teachers and school leaders was found for

Standard 4, Using assessment Results to Make Decisions. Teachers and school leaders educated in alternative programs scored highest on Standard 2, Developing Appropriate Assessment Methods. The biggest difference between the two groups was found for Standard 2, Developing Appropriate Assessment Methods, and Standard 7, Recognizing Unethical and Illegal Practices. On average, teachers and school leaders from alternative education programs scored 0.34 higher on Standard 2, while traditionally educated teachers and school leaders scored 0.46 points higher on Standard 7. Traditionally educated teachers and school leaders scored higher than teachers and school leaders from alternative education programs on four out of the seven standards (Standards 1, 3, 4, and 7). The standard mean comparison results for the two groups of teachers and school leaders for each of the seven standards are presented in Figure 2. Table 8 lists the mean and standard deviation scores for the traditional teacher education program participants (n=27) and the alternative teacher education program participants (n=36).

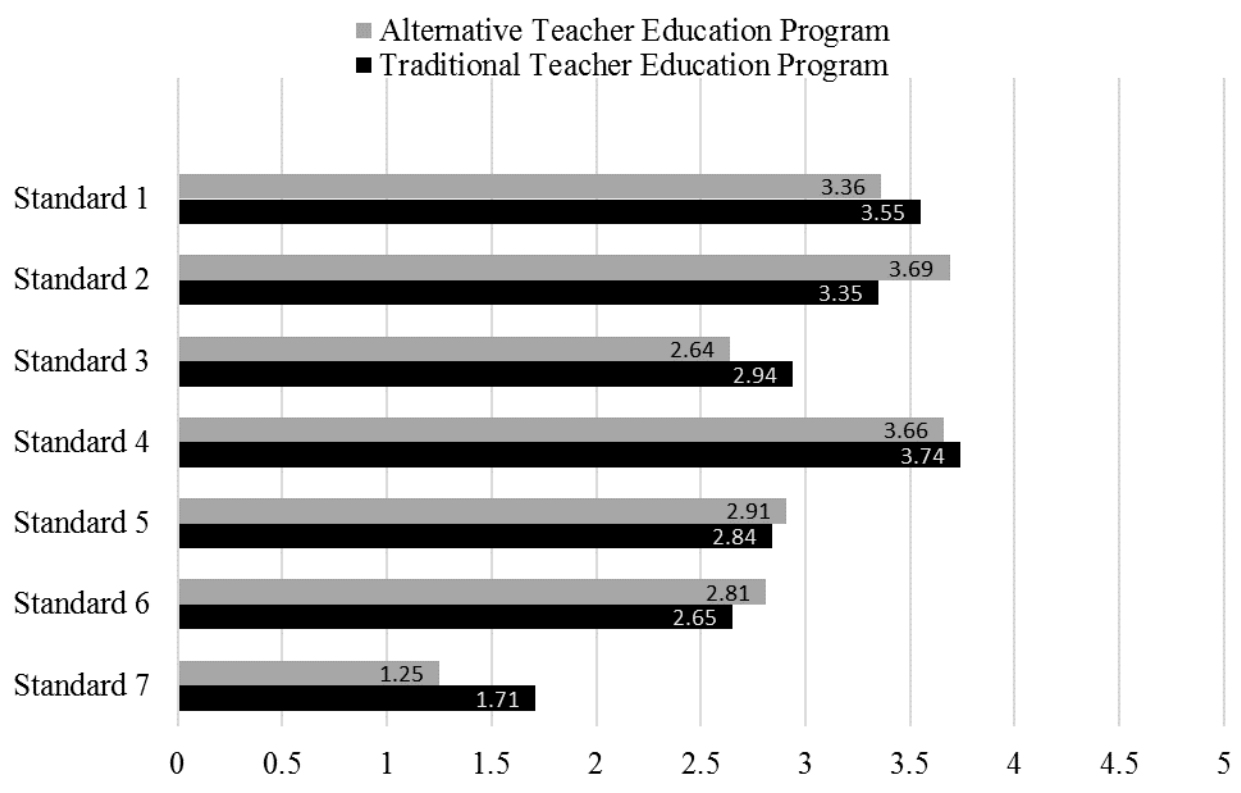


Figure 2. Standard mean score comparison, traditional versus alternative teacher education program

Out of the seven standards, both groups scored higher on Standard 1, Choosing Appropriate Assessment Methods; Standard 2, Developing Appropriate Assessment Methods; and Standard 4, Using Assessment Results to Make Decisions. Standard 7, Recognizing Unethical or Illegal Practices, was the lowest standard for both groups.

Table 8

Mean and Standard Deviation for Type of Teacher Education Program

Assessment Literacy Standard	Mean & Standard Deviation	Traditional (n=27)	Alternative (n=36)
Standard 1 <i>Choosing Appropriate Assessment Methods</i>	Mean	3.55	3.36
	SD	0.92	0.89
Standard 2 <i>Developing Appropriate Assessment Methods</i>	Mean	3.35	3.69
	SD	0.91	1.01
Standard 3 <i>Administering, Scoring, and Interpreting the Results of Assessments</i>	Mean	2.94	2.64
	SD	1.12	1.24
Standard 4 <i>Using Assessment Results to Make Decisions</i>	Mean	3.74	3.66
	SD	1.21	1.04
Standard 5 <i>Developing Valid Pupil Grading Procedures</i>	Mean	2.84	2.91
	SD	1.04	0.97
Standard 6 <i>Communicating Assessment Results</i>	Mean	2.65	2.81
	SD	0.95	0.82
Standard 7 <i>Recognizing Unethical or Illegal Practices</i>	Mean	1.71	1.25
	SD	1.10	0.81
Total Score	Mean	20.78	20.32
	SD	4.26	3.95
Note: n=63			

Summary

Results were generated based upon survey data collected among teachers (n=58) and school leaders (n=5) representing five elementary charter schools in an urban metropolitan area in California and an urban metropolitan area in Tennessee. The research fills a gap regarding the level of assessment literacy for teachers and school leaders employed by a charter network. The survey gathered research data pertaining to assessment literacy for teachers and school leaders educated in both traditional and alternative teacher education programs. The data were gathered using a previously used survey known as the CALI.

CHAPTER FIVE

SUMMARY OF THE STUDY

Introduction

The purpose of this descriptive study was to investigate the levels of assessment literacy among elementary teachers and school leaders in charter schools as measured by the Classroom Assessment Literacy Inventory (CALI).

Assessment of student performance is one of the most critical responsibilities of classroom teachers. Teachers and school leaders must acquire an understanding of assessment literacy in order to utilize data to make sound data-driven decisions. The identification of school leaders' and teachers' levels of assessment literacy provides guidance to the charter network and individual school campuses to develop adequate school leader and teacher professional development promoting assessment literacy. Brookhart (1999) noted the need for school leaders and teachers to be assessment-literate to make sound data-driven decisions. School leaders can promote and develop effective schools, while supporting teachers with their instructional practices. Teachers can improve their instructional practices and increase students' levels of achievement (Brookhart, 1999a; Mertler, 2004; Stiggins, 1999a).

To the interest of the researcher, the following research questions were raised to pursue an understanding of the levels of assessment literacy among teachers and school leaders in charter schools.

1. What is the level of assessment literacy of public charter school teachers and school leaders as measured by the Classroom Assessment Literacy Inventory?

2. Are there differences in level of assessment literacy among teachers and school leaders related to their years of experience (zero to five years and more than five years), level of education, and intensity of assessment training?
3. How does the assessment literacy of teachers and school leaders educated in a traditional education program compare to that of teachers educated in an alternative education program?

This chapter presents a detailed discussion of the findings as they relate to the research questions and the implications these findings have for future studies.

Findings

Demographic Discussion

The participant population was split evenly between 29 teachers with a bachelor's degree and 29 teachers with a master's degree. Among the surveyed school leaders three had bachelor's degrees, one a master's degree, and one a doctoral degree. The teaching experience of the survey participants was not as evenly split. Most participants (n=39) had less than five years' experience, three participants fell between six and 10 years' experience, 12 participants between 10 and 15 years, four participants between 16 and 20 years, and five participants had more than 20 years of experience. While 39 participants had five or less years of teaching experience, most teachers and school leaders had completed their teacher/leadership preparation program in the last five years or less. Only 30 participants indicated that they had completed their preparation program in the last five years or less. This indicates that some teachers and school leaders did not start teaching right away after the completion of their teacher preparation program. The largest discrepancy was found between the number of participants who had taken a stand-

alone course in assessment versus those who had not. Forty-five survey participants had not taken a stand-alone course in assessment, while 18 teachers and school leaders had taken a stand-alone course in assessment. Despite the large discrepancy in participant numbers, the data reveal no significant difference in the overall mean score levels of assessment literacy between participants who had taken a stand-alone course in assessment and those who had not. One possible reason could be the lack of classroom application practice. Teachers and school leaders might gain assessment knowledge in theory but many school leaders and teacher seem to struggle with the application of assessment practice.

Classroom Assessment Literacy Inventory Discussion

Research Question One: What is the level of assessment literacy of public charter school teachers and school leaders as measured by the Classroom Assessment Literacy Inventory?

The data from the survey showed that teachers and school leaders averaged 20.23 out of 35 questions correct (57.8%). The average is slightly lower than the average obtained in the study of in-service teachers (M=23.20/66.3%) by Plake et al. in 1993 which had the highest participant number (n=555). In Mertler's (2003) study, 197 in-service teachers were surveyed. The average score was 21.96 out of 35 questions correct (62.74%). The results show that the overall scores for teachers and school leaders decreased by since the study conducted by Mertler (2003) and Plake et al. (1993). In an era of increased accountability and data-driven decision making the scores do not show any increase in assessment knowledge, which should be cause for concern. A

comparison of the teachers and school leaders' scores for this study, Plake et al. and Mertler's study (2003), are listed in Table 9 (Appendix D).

The data from this study showed that participants had the highest scores for Standard 4; (M=3.65; maximum possible score = 5), Using Assessment Results to Make Decisions. The literature review discussed the importance and emphasis on data-driven decision making. Plake et al. (M=3.40) and Campbell and Mertler (M=3.36) received similar results in their studies. Standard 7, Recognizing Unethical or Illegal Practices, had by far the lowest scores out of all seven standards. Participants scored an average of 1.44 questions correct, which is significantly lower than the average obtained by Plake et al. (M=3.26) and Campbell and Mertler (M=3.10).

Research Question Two: Are there differences in level of assessment literacy among teachers and school leaders related to their years of experience (0 to 5 years and more than 5 years), level of education, and intensity of assessment training?

The overall percentage of correct answers for the teachers and school leaders with five or less years of teaching experience was 55%. The participants averaged 19.44 out of 35 questions (55.5%). The overall percentage of teachers and school leaders with more than five years' experience was slightly higher at 60%. The participants averaged 20.91 out of 35 questions correct (59.7%). The difference in years of experience among the survey participants was 4.2%, which equates to approximately one question difference.

Teachers and school leaders with five years or less experience scored higher on Standard 2, Developing Appropriate Assessment Methods, and Standard 5, Developing Valid Pupil Grading Procedures. The average was 3.44 questions correct (68.8%)

compared to 3.35 questions correct (67.0%) on Standard 2 and 2.87 questions correct (57.4%) compared to 2.76 questions correct (55.2%) on Standard 5.

Secondly, the researcher analyzed the difference in average responses based on participants' level of education. Teachers and school leaders with a bachelor's degree scored an average of 20.31 out of 35 items correctly (58%). Teachers and school leaders with a master's or higher degree scored an average of 20.37 out of 35 items correctly (58.2%). The percent difference between levels of education was less than 1%. The results indicate that a higher level of education did not increase a teacher or school leader's level of assessment literacy. Teachers and school leaders with a bachelor's degree scored higher on Standard 2, Developing Appropriate Assessment Methods; Standard 5, Developing Valid Pupil Grading Procedures; Standard 6, Communicating Assessment Results; and Standard 7, Recognizing Unethical or Illegal Practices. Both groups scored lowest on Standard 7. The most significant difference in the comparison was found for Standard 1, Choosing Appropriate Assessment Methods. Teachers and school leaders with a bachelor's degree received an overall percentage of 66.6% items correct, while teachers and school leaders with master's or higher degree received an overall percentage of 71.2% items correct.

The impact of intensity of assessment training was also of interest to the researcher. Eighteen teachers and school leaders stated that they had taken a stand-alone course in assessment in their teacher education program, while 45 survey participants stated that they did not participate in a stand-alone course in assessment. The data showed that teachers and school leaders who took a stand-alone course scored an average of 20.5 items correctly (58.6%). Teachers and school leaders who did not take a stand-

alone course in assessment scored an average of 19.72 items correctly (56.3%). Teachers and school leaders who took a stand-alone course of assessment scored higher on all 7 standards. The percent difference between the two groups was 2.2%.

The data indicates that teachers and school leaders are obtaining the assessment training while serving in the classroom or in the school setting.

Research question three: How does the assessment literacy of teachers and school leaders educated in a traditional education program compare to that of teachers educated in an alternative education program?

Teachers and school leaders educated in a traditional teacher education program scored an average of 20.78 items correctly (59.37%), while teachers and school leaders educated in an alternative teacher education program scored 20.32 items correctly (58.06%). The data show that teachers and school leaders educated in a traditional teacher education program scored 1.3% percent higher than teachers and school leaders educated in an alternative education program. The standards that received a higher score for teachers and school leaders educated in a traditional program were Standard 1, Choosing Appropriate Assessment Methods; Standard 3, Administering, Scoring, and Interpreting the Results of Assessments; Standard 4, Using Assessment Results to Make Decisions; and Standard 7, Recognizing Unethical or Illegal Practices. Teachers and school leaders educated in an alternative program scored higher on Standard 2, Developing Appropriate Assessment Methods; Standard 5, Developing Valid Pupil Grading Procedures; and Standard 8, Communicating Assessment Results. The data reveals no significant difference in the level of assessment literacy based on type of teacher education program.

Discussion and Implications

The findings in the current study are consistent with the information found in earlier studies by Plake et al. (1993), and Mertler (2003). These researchers measured the level of assessment literacy of in-service teachers in public schools and found comparable levels of assessment literacy. Considering that federal and state accountability mandates have demanded an increase in the emphasis placed on assessment training of teachers and school leaders, the findings show that the accountability movement has not been accompanied by an increase in levels of assessment literacy. Among the studies presented, all support a need for better assessment training as well as improved pre-service training. In general, the literature review indicated that pre-service teachers, in-service teachers, and school leaders have low, insufficient levels of assessment literacy (Plake et al., 1993; Mertler 2003; Gotch & French, 2013). Studies by Plake et al., 1993); Quilter and Gallini (2002); and Mertler (2003, 2009) yielded common areas of weakness including Developing Assessment Methods Appropriate for Instructional Decisions; Developing Valid Student Grading Procedures That Use Student Assessments; and Communicating Assessment Results to Students, Parents, and Other Audiences. The current study, even though conducted at least ten years later, continues to show a significant lack of assessment literacy. This study's data revealed that the weakest areas continue to be Administering, Scoring, and Interpreting the Results of Assessment; Communicating Assessment Results; and Recognizing Unethical or Illegal Practices. On reflection, the researcher wonders if the reason for the continuous low levels of assessment literacy are due to the inventory being outdated and not aligned with current assessment practices. Another possible explanation is that teacher training programs are

not able to keep up with the ever-changing assessment practices and assessment types. The constant introduction of new and revised mandates makes it hard for education programs, teachers, and school leaders to maintain currency.

The literature review also discussed possible reasons that teachers and school leaders lack the needed levels of assessment literacy. Levine (2006) reported that the nation's teacher education programs leave teachers and principals unprepared. The report identified several model teacher education programs and critiqued their outdated, historically flawed vision of teacher education. He states that traditional teacher education programs only introduce students to prolonged, real-life classroom settings late in their pre-service program, if at all. Levine goes on to argue that students in teacher education programs lack the opportunity for immediate application and testing of learned skills, as well as familiarity with some of the types of situations they will encounter once they have their own classrooms. Research has shown that traditional teacher preparation courses in classroom assessment are not well matched with what teachers need to know for classroom practice (Schafer, 1993).

Gawlik et al. (2012) conducted a study to determine whether regular public and charter school teachers in the Detroit metropolitan region differed in indicators of teacher quality. The findings concluded that the charter school teacher workforce was more likely to be noncertified and inexperienced, and to hold a substitute permit. Although charter school teachers were more likely to be graduates of a competitive college and to hold a major or minor in their teaching subject matter area, they left teaching at a higher rate than traditional public school teachers. Teacher quality indicators were consistently associated with teacher effectiveness only in a few cases, except for the percentage of

certified teachers. Stitzlin and West (2014) analyzed charter-inspired teacher preparation programs. Their analysis highlights the lack of teacher flexibility, limited theoretical knowledge of good teaching, a loss of the overarching purpose of educating for democracy that is integral to traditional teacher colleges, and altering views of teaching professions themselves. Stitzlin and West (2014) acknowledge that charter-inspired teacher preparation programs may potentially do better at training preservice teachers for the difficult realities of teaching in high-needs classrooms by situating their primary learning experiences in these settings from the start. One noteworthy difference of charter-inspired teacher preparation programs versus traditional teacher programs is that charter-inspired teacher preparation programs strive to produce teachers who demonstrate significant results in improving student achievement.

The findings for research question three addressed this issue by analyzing differences in levels of assessment literacy based on the teachers' and school leaders' type of teacher education program. The study findings were not what the researcher anticipated. The researcher expected that teachers and school leaders educated in a traditional teacher education program would outperform the teachers and school leaders educated in alternative teacher education programs. However, the difference in levels of assessment literacy between the two types of programs was minimal. Teachers and school leaders educated in a traditional teacher education program scored an average of 20.78 out of 35 items correctly (59.37%), while teachers and school leaders educated in an alternative teacher education program scored 20.32 out of 35 items correctly (58.06%). The data show that teachers and school leaders educated in a traditional teacher education program scored 1.3% percent higher than teachers and school leaders

educated in an alternative education program. While teachers from a traditional teacher education program scored higher on four out of the seven standards (see table 8 and table 13).

One other issue that emerged from the data was the significantly lower scores on Standard 7, Recognizing Unethical or Illegal Practices; by far the lowest scores out of all seven standards. Participants scored an average of 1.44 questions correct, which is significantly lower than the average obtained by Plake et al. ($M=3.26$) and Campbell and Mertler ($M=3.10$). The researcher wondered how this happened in light of recent fraud scandals (in the Atlanta Public Schools scandal in 2015, about 150 educators resigned, retired, or lost their appeals to have their jobs reinstated after being involved in unethical or illegal assessment practices (Colbert, 2016); in the Texas school scandal in 2012, a superintendent faced prison after putting pressure on educators and school boards to refrain from testing low-achieving students (Sanchez, 2013). The researcher assumed that the scores would have increased due to the immense impact the accountability movement has had on districts budgets, school rankings, and school leader and teacher evaluations. However, one factor that might impact the current research study's outcome is the population sample. In the two previous studies the in-service teachers were part of the public-school system and the pre-service teachers had been educated in a traditional teacher program. While public charter schools, in general, are administering the same standardized end-of-year tests as public schools, many public charter school networks, including the one used for this study, have their own internal assessment system to evaluate student growth and to evaluate regions, schools, school leaders, and teachers. The pressure on teachers and school leaders appears to be less intense than it is in many

public schools where school leaders and teachers are threatened by state takeover or shutdowns (Wells, 2002; Cohodes, 2016). Table 10 depicts a summary of the data analyzed for research questions two and three. It highlights that no matter how participants were grouped the lowest standard is Standard 7.

Most concerning and surprising to the researcher were the fact that the level of assessment literacy decreased over the last 20 plus years. The researcher wonders if more assessment training was provided to in-service teachers in 1993 to 2003 given the newly created Standards for Teacher Competence of Educational Assessment. Often with a new innovation comes an increase in training, materials, and supervision.

Another thought that the researcher formed is the influence Black and Wiliam' research (1998, 2009, 2010) had on assessment literacy. Soon after the Standards for Teacher Competence of Educational Assessment were published, Black and Wiliam transformed classroom assessment practices with their profound research on formative assessment. The researcher wondered if professional development was/is focused primarily on formative assessments rather than a deep understanding of assessment literacy and how high level levels of assessment literacy can impact student learning outcomes.

Limitations

This study had several limitations. First, it involved teachers and school leaders voluntarily taking the Classroom Assessment Literacy Inventory (CALI). This limitation did not promote the collection of an equal and balanced representation of teacher and school leader participants for the sample size. In addition, it limited the sample to only those who were motivated to participate and did not allow the researcher to control the

sample size. Of approximately 200 teachers and school leaders who were invited to participate, 87 attempted the inventory but only 63 (31.5%) completed the inventory.

A second limitation of the study was the fact that the inventory was completed in isolation by the participant, which did not allow for monitoring or assistance if needed. This isolation did not give participants an opportunity to ask questions in the event that clarification was needed regarding procedures or question items. This limitation may have affected the accuracy of the data gathered from the CALI. Additionally, because participants were not monitored while completing the inventory, participants could have worked together or researched to answer the CALI questions, thus influencing the answers of other participants.

A third limitation of the study involved the participants' comprehension of the vocabulary and terms used in gathering the data from participants. This includes the demographic data and responses to the CALI.

Conclusions

Within the limitations established, the following conclusions seem justified:

1. Teachers' and school leaders' level of assessment literacy has not changed significantly over the last 20 years.
2. Teachers and school leaders feel unprepared to assess their students adequately regardless of their years of experience, level of education, and time passed since leaving their teacher education program.
3. A stand-alone course in assessment has little impact on teachers' and school leaders' level of assessment literacy. It emphasizes the thought that

assessment training is an “on-the-job” type of training; in other words, assessment skills are best learned through experience and application.

4. Whether teachers and school leaders attended a traditional or alternative teacher education program has no significant impact on their level of assessment literacy.
5. Teachers and school leaders seem to be unaware of the implications of recognizing unethical or illegal practices.

Recommendations

The recommendations are, first and foremost for teachers, school leaders, and everyone in contact with students on a day-to-day basis. Assessment literacy is too important to be neglected. Assessment practices are not just a means to measure students or to reduce them to numbers, it is way to analyze, evaluate, guide, and push students to reach their highest potential. Secondly, to reduce the amount of professional development needed to fill gaps in assessment literacy, teacher preparation programs need to be restructured to provide future school leaders and teachers with high levels of assessment literacy. Therefore, these recommendations are also intended for colleges of education and universities. As a result of this study’s findings the following recommendations are offered:

1. Because the need of an assessment literate faculty is essential, schools and districts have to focus on effective data analysis use.
2. Because federal and state accountability mandates are ever-changing, school leaders and teachers need continual professional development regarding

assessment literacy to acquire knowledge and deepen understanding and practice regarding assessment.

3. Because Standard 7, Recognizing Unethical or Illegal Practices, was the lowest standard reported by teachers and school leaders. The researcher recommends that the charter network and school leadership explore why teachers appear to be unaware of these practices and develop a plan to address the issue more effectively.
4. Because assessment practices are continuously evolving and changing, it would serve the network and school leadership well to be patient and supportive of their teachers as they journey through the process of change in implementing new assessments.
5. Because professional development activities might generally be improved and more useful for teachers and school leaders if they were based on targeted input from teachers and school leaders.
6. Because teachers and school leaders enter the profession ill-prepared embedding assessment courses throughout a pre-service teachers' education training would support the understanding of assessment being an integral interdisciplinary part of day.

Implications for Future Research

The current study was based on two earlier studies, one by Plake et al. in 1993 and one by Mertler in 2003. The data for the levels of assessment literacy as measured by the CALI was similar for all three studies. A question for further research is whether the CALI continues to be an adequate measure of levels of assessment literacy as outlined

by current educational mandates. Brookhart (2011) suggested the need for an updated list of knowledge and skills that teachers and school leaders need to perform the assessment-related aspects of their work in a competent and professional manner. The question to be considered is whether the Standards for Teacher Competence on Educational Assessment of Students as developed in 1991 still cover the skills important for teachers and school leaders to meet the current testing requirements.

While the Standards for Teacher Competence on Educational Assessment of Students outline a solid foundation for assessment practices, the researcher wonders if the alignment of questions to each Standard is outdated and/or might require some critical revising to embed more up-to-date assessment jargon as well practices. The researcher believes a more complex component on formative versus summative practices should be included considering the influence Black and Wiliam have had on classroom assessment practices.

Another area of future research would be to look at the data for teachers and school leaders from a different charter network considering the pressure and scrutiny that charter schools are presently under.

In addition, it would be interesting to see whether differences exist in levels of assessment literacy based on the various types of teacher education programs, particularly looking at the level, intensity, and relevance of assessment courses within each program. Assessment course requirements vary greatly from state to state and from college to college.

It would be of interest to investigate if assessment courses are covered deeper at a Masters' level compared to the Bachelor level? Considering many teachers require their

Master degree after having already taught, the researcher wonders if assessment courses at a Master's level are more meaningful to teachers and school leaders because they are able to apply the learned practices in their classroom/school setting.

Summary

This chapter presented a discussion of the study findings. The researcher concluded that the levels of assessment literacy for teachers and school leaders employed by a charter network are comparable to the level of assessment literacy for public school in-service teachers of previous studies. The researcher also noted that variables such as years of experience (zero to five years and six plus years), level of education, and intensity of assessment training are not relevant when determining levels of assessment literacy. Lastly, the researcher found that the type of teacher education program does not impact a teacher's or school leader's level of assessment literacy.

REFERENCES

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). Standards for educational and psychological testing.
- Ainsworth, L., & Viegut, D. (2006). *Common formative assessments*. Thousand Oaks, CA: Corwin Press.
- Ali, A., & Ali, U. (2010). Educational measurement and testing: Historical perspectives. *Journal of Educational Research*, (1027-9776), 13(2), 216–221.
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education Policy Analysis Archives*, 10(18), 1-74.
- Andrews, K., & Rothman, M. (2002). Cultivating innovation: How a charter/district network is turning professional development into professional practice. *Phi Delta Kappan*, 83(7), 506-512.
- Bauerlein, M. (2015). Standardized culture. *First Things: A Monthly Journal of Religion & Public Life*, 249, 15-17.
- Bennett, J. (2008). Brand-name charters. *Education Next*, 8(3), 28-34.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Beziat, T. L., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, 27(1), 25-30.
- Bifulco, R., & Ladd, H. (2003). Charter school impacts on student performance: Evidence from North Carolina. APPAM paper.

- Bifulco, R., & Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy, 1*(1), 50-90.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* (formerly: *Journal of Personnel Evaluation in Education*), *21*(1), 5.
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81-90.
- Boudett, K. P., City, E. A., & Murnane, R. J. (2006). The “data wise” improvement process. *Harvard Education Letter, 11*(4), 1-3.
- Brookhart, S. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice, 18*(1), 5-13.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues & Practice, 30*(1), 3-12. Retrieved from: <https://ezproxy.mtsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=508199957&site=eds-live&scope=site>
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers*. Columbus: Mid-Western Educational Research Association.
- Carl, N. M. (2014). Reacting to the script: teach for America teachers' experiences with scripted curricula. *Teacher Education Quarterly, 41*(2), 29.

- Cave, A., & Brown, C. W. (2010, June). When learning is at stake: Exploration of the role of teacher training and professional development schools on elementary students' math achievement. *National Forum of Teacher Education Journal*, 20(3), 1-21.
- Center for Research on Educational Outcomes (CREDO). (2009). Multiple choice charter school performance in 16 states. Retrieved from:
https://credo.stanford.edu/reports/MULTIPLE_CHOICE_CREDO.pdf
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-44.
- Chingos, M. M., & West, M. R. (2015). The uneven performance of Arizona's charter schools. *Educational Evaluation and Policy Analysis*, 37(1_suppl), 120S-134S.
- Clarke, M. M., Madaus, G. F., & Horn, C. L. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies*, 32(2), 159-181. <http://doi.org/10.1080/002202700182691>
- Cochran-Smith, M., Piazza, P., & Power, C. (2013, January). The politics of accountability: Assessing teacher education in the United States. In *The Educational Forum* (Vol. 77, No. 1, pp. 6-27). Taylor & Francis Online.
- Colbert, J. (2016). The Atlanta Public Schools Scandal: Educator Fraud, RICO, and COSO. *Journal of Business Case Studies (Online)*, 12(3), 103.
- Cohodes, S. R. (2016). Teaching to the student: Charter school effectiveness in spite of perverse incentives. *Education Finance and Policy*, 11(1), 1-42.
- Creswell, W. J. (2013). *Qualitative inquiry and research design*. Los Angeles, CA: Sage Publications, Inc.

- Darling-Hammond, L. (1995). Restructuring schools for student success. *Daedalus*, 124(4), 153-162. Retrieved from <http://www.jstor.org/stable/20027341>
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64(4), 356-372. <http://doi.org/10.1177/0022487113488144>
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- DeLuca, C., & Lam, C. (2014). Preparing Teachers for Assessment within Diverse Classrooms: An Analysis of Teacher Candidates' Conceptualizations. *Teacher Education Quarterly*, 41(3), 3-24. Retrieved from <http://www.jstor.org/stable/teaceducquar.41.3.3>
- Diamond, J. B., & Cooper, K. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago. *Yearbook of the National Society for the Study of Education*, 106(1), 241-263.
- Diket, R. M., & Brewer, T. M. (2011). NAEP and policy: Chasing the tail of the assessment tiger. *Arts Education Policy Review*, 112(1), 35-47.
- Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly*, 49(5), 838-882.
- Duncan, A., & Santy, R. (2015). US Department of Education. *Race to the Top*. Retrieved from <http://www.ed.gov/about/reports/annual/index.html>.

- Eisner, E. W. (2001). *Educational imagination: On the design and evaluation of school programs* (3rd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
- Erwin, T. D., & DeFilippo, J. G. (2010). The evolution of assessment policy: A view from Virginia. *Change: The Magazine of Higher Learning*, 42(2), 40-44.
- ESEA Reauthorization | ED.gov Blog. (n.d.). Retrieved from:
<http://www.ed.gov/blog/topic/esea-reauthorization/>
- Fullan, M. (2002). The change leader. *Educational Leadership*, 59(8), 16-20.
Retrieved from:
<http://eds.a.ebscohost.com.ezproxy.mtsu.edu/eds/pdfviewer/pdfviewer?vid=4&sid=c8519908-181c-4cdd-a603-5d73bd731d43%40sessionmgr101>
- Fullan, M., Hill, P., & Crévola, C. (Eds.). (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.
- Gawlik, M. A., Kearney, C. P., Addonizio, M. F., & LaPlante-Sosnowsky, F. (2012). Teacher quality in Michigan: A school-level analysis of the Detroit metropolitan region. *Education and Urban Society*, 44(4), 412-447.
- Gerberich, J. R. (1963). Development of educational testing. *Theory Into Practice*, 2, 184-191. <http://doi.org/10.1080/00405846309541861>
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues & Practice*, 33(2), 14-18.
<http://doi.org/10.1111/emip.12030>
- Grady, D. P. (2012). Charter school revocation: A method for efficiency, accountability, and success. *JL & Educ.*, 41, 513-518.

- Guskey, T. R. (2007). Using assessments to improve teaching and learning. In *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 15-29). Bloomington, IN: Solution Tree, 2007.
- Hailaya, W., Alagumalai, S., & Ben, F. (2014). Examining the utility of Assessment Literacy Inventory and its portability to education systems in the Asia Pacific region. *Australian Journal of Education* (Sage Publications Ltd.), 58(3), 297. doi:10.1177/0004944114542984
- Haney, W. M., & Others, A. (1993). *The fractured marketplace for standardized testing: Evaluation in education and human services*. Norwell, MA: Kluwer Academic Publishers.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Harris, L.R., Irving, S.E. and Petersen, E. (2008). *Secondary teachers' conceptions of the purpose of assessment and feedback*. Paper presented to the Annual Conference of the Australian Association for Research in Education (AARE). Brisbane, Australia.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Bookcraft Ltd.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
- Hess, F. M., & Kelly, A. P. (2007). Learning to lead: What gets taught in principal-preparation programs. *Teachers College Record*, 109(1), 244-274.

- Honawar, V. (2007). Gains seen in retooled teacher education. *Education Week*, 27(10), 1-13. Retrieved from: <https://eric.ed.gov/?id=EJ779497>
- Horace Mann (1796-1859) - Education and Training, Career and Contribution. (n.d.). Retrieved April 27, 2015, from: <http://education.stateuniversity.com/pages/2197/Mann-Horace-1796-1859.html#ixzz3YSnjYuw3>
- Horace Mann | biography - American educator. (n.d.). Retrieved April 27, 2015, from: <http://www.britannica.com/EBchecked/topic/362466/Horace-Mann>
- Hoxby, C. M. (2009). A statistical mistake in the CREDO study of charter schools. Retrieved June, 8, 2016.
- Hoxby, C. M., & Rockoff, J. E. (2005). Findings from the city of big shoulders. *Education Next*, 5(4), 52-58.
- Ikemoto, G., & Marsh, J. (2007, August). Cutting through the “data driven” mantra: Different conceptions of data-driven decision making. *Yearbook of the National Society for the Study of Education*, 106(1), 105-131.
- Impara, J. (1993). *Joint Committee on Competency Standards in Student Assessment for Educational Administrators Update: Assessment survey results*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Atlanta, GA.
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Teachers’ assessment background and attitudes toward testing. *Theory Into Practice*, 32, 113-117.

- Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258-1287.
- Johnson, S. T., Thompson, S. D., Wallace, M. B., Hughes, G. B., & Butty, J. A. L. M. (1998). How teachers and university faculty perceive the need for and importance of professional development in performance-based assessment. *Journal of Negro Education, 67*(3), 197-210.
- Justice, M., Greiner, C., & Anderson, S. (2003). Determining the influences of traditional Texas teachers vs. teachers in the emergency teaching certification program. *Education, 124*(2), 376. Retrieved from:
http://www.jstor.org/stable/2668189?seq=1#page_scan_tab_contents
- Kahn, S. (2012). *The one world school house: Education reimagined* (Kindle edition). New York, NY: Grand Central Publishing.
- Kakadelis, L. (2002). Cap on charter schools discourages change. In W.D. Lewis (Ed.), *The politics of parent choice in public education: The choice movement in North Carolina and the United States* (pp. 62 – 63). New York, NY: Springer Publishing.
- Kirkpatrick, L., Lincoln, F., & Morrow, L. R. (2006). Assessment of a collaborative teacher preparation program: Voices of interns. *Delta Kappa Gamma Bulletin, 73*(1), 36-54.
- Knupfer, N. N., & McLellan, H. (1996). Descriptive research methodologies. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 1196–1212). New York, NY: Macmillan.

- Kripalani, L. (2002). Charter schools: Our challenge, our opportunity. *Montessori Life*, 14(3), 43-44.
- Kronholz, J. (2012). Teaching the teachers. *Education Next*, 12(3), 8-15.
- Leistyna, P. (2007). Corporate testing: Standards, profits, and the demise of the public sphere. *Teacher Education Quarterly*, 34(2), 59-84.
- Levine, A. (2005). Educating school leaders. *The Education Schools Project*. Retrieved from www.edschools.org/pdf/Final313.pdf
- Lezotte, L. W., & Snyder, K. T. (2011). *What effective schools do: Re-envisioning the correlates*. Bloomington, IN: Solution Tree Press.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Prentice Hall.
- Lubienski, C. A., & Weitzel, P. C. (2010). *The charter school experiment: Expectations, evidence, and implications*. Cambridge, MA: Harvard Education Press.
- Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, 20(5), 523-535.
<http://doi.org/10.1016/j.tate.2004.04.008>
- Madaus, G. F. (1993). A national testing system: Manna from above? An historical/technological perspective. *Educational Assessment*, 1(1), 9. Retrieved from:
<https://ezproxy.mtsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=7403536&site=eds-live&scope=site>
- Madaus, G., & Russell, M. (2010). Paradoxes of high-stakes testing. *Journal of Education*, 190(1/2), 21-30. Retrieved from:

<https://ezproxy.mtsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=54018443&site=eds-live&scope=site>

Marzano, R. J. (2000). *Introduction to the special section implementing standards in schools updating the standards movement*. Retrieved from:

<http://journals.sagepub.com/doi/abs/10.1177/019263650008462001?journalCode=buld>

Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandria, VA: ASCD.

Marsh, C. J. (2007). A critical analysis of the use of formative assessment in schools. *Educational research for policy and practice*, 6(1), 25-29.

Maul, A., & McClelland, A. (2013). Review of national charter school study 2013. Boulder, CO: National Education Policy Center. Retrieved September, 2, 2016.

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120(2), 285-296.

Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the annual meeting of the MidWestern Educational Research Association, Columbus, OH. Retrieved from: <https://eric.ed.gov/?id=ED482277>

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49-64.

Mertler, C. A. (2005). Erratum: Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(2), 76-92.

- Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instructional decision making*. Thousand Oaks, CA: Sage Publications.
- Mertler, C. A. (2014). *Action research: Improving schools and empowering educators*. Los Angeles, CA: Sage Publications, Inc.
- Mertler, C.A. & Campbell, C.S. (2005) *Measuring teachers' knowledge and application of classroom assessment concepts: development of the Assessment Literacy Inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada. Retrieved from:
<https://eric.ed.gov/?id=ED490355>
- Meyer, P. (2014). Uncommon success: A conversation with Brett Peiser. *Education Next*, 14(3), 46-51.
- Mittenfelner, C. N. (2014). Reacting to the script: Teach for America teachers' experiences with scripted curricula. *Teacher Education Quarterly*, 41(2), 29-50.
- Muhammad, A. (2009). *Transforming school culture*. Bloomington, IN: Solution Tree Press.
- National Center for Education Statistics. (2012). *The nation's report card: Vocabulary results from the 2009 and 2011 NAEP reading assessments. National Assessment of Educational Progress at Grades 4, 8, and 12*. NCES 2013-452. Retrieved from:
<http://eric.ed.gov/?id=ED537659>
- Nelson, R., & Dawson, P. (2014). A contribution to the history of assessment: How a conversation simulator redeems Socratic method. *Assessment & Evaluation in Higher Education*, 39(2), 195-204. <http://doi.org/10.1080/02602938.2013.798394>

- O'Brien, E. M. & Dervarics, C. (2012). *Charter schools: Finding out the facts: At a glance*. Retrieved from <http://www.centerforpubliceducation.org/MainMenu/Organizing-a-school/Charter-schools-Finding-out-the-facts-At-a-glance>
- Plake, B. S., & Impara, J. C. (1993). *Teacher assessment literacy: Development of training modules*. Paper presented at the annual meeting of the national council on measurement in education. Retrieved from: <https://eric.ed.gov/?id=ED358131>
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: ASCD.
- Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*, 62(1), 82.
- Popham, J. W. (2006). Needed: A dose of assessment literacy. *Educational Leadership*, 63(6), 84-85.
- Popham, W. J. (2009). *Instruction that measures up: Successful teaching in the age of accountability*. Alexandria, VA: ASCD.
- Popham, W. J. (2010). *Everything school leaders need to know about assessment*. Thousand Oaks, CA: Corwin Press.

- Popham, W. J. (2011). *Transformative assessment in action: An inside look at applying the process*. Alexandria, VA: ASCD.
- Popham, W. J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). New York, NY: Pearson Education, Inc.
- Popham, W. J. (2014). Looking at assessment through learning-colored lenses. In *Designing assessment for quality learning* (pp. 183-194). New York, NY: Springer Publishing.
- Popham, W. J. (2014). The right test for the wrong reason. *Phi Delta Kappan*, 96(1), 46-52. <http://doi.org/10.1177/0031721714547862>
- Ramirez, A. (1999). Assessment-driven reform: The emperor still has no clothes. *Phi Delta Kappan*, 81(3), 204-208.
- Raptis, H., & Fleming, T. (2006). Large-scale assessment outcomes in British Columbia, 1876-1999. *Canadian Journal of Education*, 29(4), 1191-1222.
doi:10.2307/20054215
- Ravitch, D. (2010). *The life and death of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.
- Raymond, M. E., Woodworth, J. L., Cremata, E., Davis, D., Dickey, K., Lawyer, K., & Negassi, Y. (2013). National charter school study. Center for Research on Education Outcomes. Stanford, CA: Stanford University. Retrieved from: <http://credo.stanford.edu/documents/NCSS,202013>.
- Reese, W. J. (2013b). *Testing wars in the public schools: A forgotten history*. Cambridge, MA: Harvard University Press.

- Reeves, D. B. (2004). *Accountability in action: A blueprint for learning organizations*. Englewood, CO: Advanced Learning Press.
- Reeves, D. B. (Ed.). (2009). *Ahead of the curve: The power of assessment to transform teaching and learning*. Bloomington, IN: Solution Tree Press.
- Resnick, L. B., & Hall, M. W. (1998). Learning organizations for sustainable education reform. *Daedalus*, 127(4), 89-118.
- Roland, D. (2006). School choice and the North Carolina constitution. *Institute for Justice*. Retrieved from: <http://files.eric.ed.gov/fulltext/ED505528.pdf>
- Sanchez, C. (2013). *El Paso schools cheating scandal: Who's accountable?* Washington DC: National Public Radio.
- Sanders, J.R., Hills, J.R., Nitko, A.J., Merwin, J.C., Trice, C., Dianda, M., & Schneider, J. (1990). *Buros center for testing: Standards for teacher competence in educational assessment of students*. Retrieved from <http://buros.org/standards-teacher-competence-educational-assessment-students>
- Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice*, 32(2), 118-126.
- Scott, J., & Marshall, G. (2009). *A dictionary of sociology*. New York, NY: Oxford University Press.
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M.K. & Yin, Y. (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, L. A. (2000). The role of assessment in a learning culture. *Journal of Education*, 189(1/2), 95-106.
- Shepard, L. A. (2010). What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education*, 85(2), 246-257.
- Shepard, L. A. (2016). Testing and assessment for the good of education. *Educational Researcher*, 45(2), 112. doi:10.3102/0013189X16639599
- Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, 22(4), 371-391.
- Stiggins, R. J. (1991). Facing challenges of a new era of educational assessment. *Applied Measurement in Education*, 4(4), 263. Retrieved from:
<https://ezproxy.mtsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=7366090&site=eds-live&scope=site>
- Stiggins, R. J. (1991a). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. J. (1991b). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, 10(1), 7-12.
- Stiggins, R. J. (1998). *Classroom assessment for student success*. Washington, DC: National Education Association.
- Stiggins, R. J. (1999). Are you assessment literate? *The High School Journal*, 6(5), 20-23.
- Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81(3), 191-198. Retrieved from:
<https://ezproxy.mtsu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=rgm&AN=503597909&site=eds-live&scope=site>

- Stiggins, R. J. (2002a). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. J. (2002b). Assessment for learning. *Education Week*, 21(26), 30, 32-33.
- Stiggins, R. J. (2014). *Revolutionize assessment: Empower students, inspire learning*. Thousand Oaks, CA: Corwin.
- Stiggins, R. J., & Bridgeford, J. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22, 271-286.
- Stiggins, R. J., & Conklin, N. F. (1988). *Teacher training in assessment*. Portland, OR: Northwest Regional Education Laboratory.
- Stiggins, R. J., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285-291.
- Stiggins, R. J., & Stiggins, R. J. (2001). *Student-involved classroom assessment*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Stitzlein, S. M., & West, C. K. (2014). New forms of teacher education: Connections to charter schools and their approaches. *Democracy & Education*, 22(2), 2.
- Stronge, J. H. (2007). *Qualities of effective teachers*. Alexandria, VA: ASCD.
- Thurlow, M. L. (2000). Standards-based reform and students with disabilities: Reflections on a decade of change. *Focus on Exceptional Children*, 33(3), 1-16.
- Tucker, M. S. (2014). Fixing our national accountability system. National Center on Education and the Economy. Retrieved from: <https://eric.ed.gov/?id=ED556313>
- Uriegas, B. (2012). *The impact of certification type and student teaching experience on classroom management in a rural south Texas district*. Texas A&M University-

Kingsville. Retrieved from:

<http://search.proquest.com/docview/1282630994?accountid=31255>

- Uriegas, B., Kupczynski, L., & Mundy, M. A. (2014). A comparison of traditional and alternative certification routes on classroom management. *SAGE Open*, 4(4), 2158244014553599.
- Veltri, B. T. (2008). Teaching or service? The site-based realities of Teach for America teachers in poor, urban schools. *Education & Urban Society*, 40(5), 511-542.
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770. Retrieved from:
<http://files.eric.ed.gov/fulltext/EJ780818.pdf>
- Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need—and what we can do about it*. New York, NY: Basic Books.
- Webb, N. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states: A study of the State Collaborative on Assessment & Student Standards (SCASS), Technical Issues in Large-Scale Assessment (TILSA)*. Council of Chief State School Officers.
- Wells, A. S. (Ed.). (2002). *Where charter school policy fails: The problems of accountability and equity* (Vol. 12). New York, NY: Teachers College Press.
- William, D. (2007). Changing classroom practice. *Educational leadership*, 65(4), 36.

- William, D. (2010). The role of formative assessment in effective learning environments. In H. Dumont, D. Instance, and F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 135-155). OECD online.
- Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149-162.
doi:10.1016/j.tate.2016.05.010
- Yancey, P. (2006). From the tightrope: Designing, developing, and delivering an alternative teacher education model. *Multicultural Education, 14*(2), 24.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., ... & Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education, 21*(4), 335-359.
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives, 18*(19), 1-37.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323-342.
- Zimmer, R., Gill, B., Booker, K., Lavertu, S., Sass, T. R., & Witte, J. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration, and competition* (Vol. 869). Santa Monica, CA: Rand Corporation.
- Zuckerbrod, N. (2012, November). City schools gain, yet still lag nation. Retrieved June 11, 2015, from:

[http://www.lasentinel.net/index.php?option=com_content&view=article&id=8393
:city-schools-gain-yet-still-lag-nation&catid=65&Itemid=155](http://www.lasentinel.net/index.php?option=com_content&view=article&id=8393:city-schools-gain-yet-still-lag-nation&catid=65&Itemid=155)

Zwick, R., Sklar, J., Wakefield, G., Hamilton, C., Norman, A., & Folsom, D. (2008, Summer). Instructional tools in educational measurement and statistics (ITEMS) for school personnel: Evaluation of three web-based training modules. *Educational Measurement: Issues and Practice*, 27, 14-27.

APPENDICES

APPENDIX A

IRB FORM

IRB
INSTITUTIONAL REVIEW BOARD
 Office of Research Compliance,
 010A Sam Ingram Building,
 2269 Middle Tennessee Blvd
 Murfreesboro, TN 37129



IRBN007 – EXEMPTION DETERMINATION NOTICE

Wednesday, May 18, 2016

Investigator(s): Mareen Pfeiffer-Hoens, and Dr. Donald Snead
 Investigator(s) Email(s): mp2r@mtmail.mtsu.edu
 Department: College of Education (Ed. Leadership)

Study Title: An Investigation of Charter Schools' Principals, Assistant Principals, and Teachers Level of Assessment Literacy
 Protocol ID: 16-1271

Dear Investigator(s),

The above identified research proposal has been reviewed by the MTSU Institutional Review Board (IRB) through the **EXEMPT** review mechanism under 45 CFR 46.101(b)(2) within the research category (2) *Educational Tests*. A summary of the IRB action and other particulars in regard to this protocol application is tabulated as shown below:

IRB Action	EXEMPT from further IRB review***	
Date of expiration	NOT APPLICABLE	
Sample Size	Up to 100	
Participant Pool	Adults (18 years of age or older)	
Mandatory Requirements	Must collect informed consent forms	
Additional Restrictions	Leadership and staff at Rocketship Northeast and at Rocketship United Academy	
Comments	N/A	
Amendments	Date	Post-Approval Amendments
	N/A	N/A

***This exemption determination only allows above defined protocol from further IRB review such as continuing review. However, the following post-approval requirements still apply:

- Addition/removal of subject population should not be implemented without IRB approval
- Change in investigators must be notified and approved
- Modifications to procedures must be clearly articulated in an addendum request and the proposed changes must not be incorporated without an approval
- Be advised that the proposed change must comply within the requirements for exemption
- Changes to the research location must be approved – appropriate permission letter(s) from external institutions must accompany the addendum request form
- Changes to funding source must be notified via email (irb_submissions@mtsu.edu)
- The exemption does not expire as long as the protocol is in good standing

- Project completion must be reported via email (irb_submissions@mtsu.edu)
- Research-related injuries to the participants and other events must be reported within 48 hours of such events to compliance@mtsu.edu

The current MTSU IRB policies allow the investigators to make the following types of changes to this protocol without the need to report to the Office of Compliance, as long as the proposed changes do not result in the cancellation of the protocols eligibility for exemption:

- Editorial and minor administrative revisions to the consent form or other study documents
- Increasing/decreasing the participant size

The investigator(s) indicated in this notification should read and abide by all applicable post-approval conditions imposed with this approval. [Refer to the post-approval guidelines posted in the MTSU IRB's website.](#) Any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918 within 48 hours of the incident.

All of the research-related records, which include signed consent forms, current & past investigator information, training certificates, survey instruments and other documents related to the study, must be retained by the PI or the faculty advisor (if the PI is a student) at the secure location mentioned in the protocol application. The data storage must be maintained for at least three (3) years after study completion. Subsequently, the researcher may destroy the data in a manner that maintains confidentiality and anonymity. IRB reserves the right to modify, change or cancel the terms of this letter without prior notice. Be advised that IRB also reserves the right to inspect or audit your records if needed.

Sincerely,

Institutional Review Board
Middle Tennessee State University

Quick Links:

[Click here](#) for a detailed list of the post-approval responsibilities.
More information on exempt procedures can be found [here](#).

APPENDIX B

PARTICIPATION INVITATION

Hello Rocketeers,

My name is Mareen P. Hoens, and I am a Doctoral student at Middle Tennessee State University. I am inviting you to participate in a research study. I am working on completing the Doctorate of Education degree in the Assessment, Learning, and School Improvement Program. This study will be the focus of my dissertation that is a portion of the degree requirements.

The purpose of this study is to determine the level of assessment literacy in administrators and teachers in a charter school setting. Assessment literacy refers to educators' understanding of the principles of sound assessment to skillfully develop assessments that transform learning purposes into assessment activities which exactly demonstrate student understanding and achievement. Educators with assessment literacy know what they assess, why they assess, how to assess, what the possible problems with assessment are, and how to prevent them from occurring. Assessment of student performance is one of the most critical responsibilities of classroom teachers because to a great extent it influences everything that teachers do. Building principals and assistant principals need the assessment literacy skills to function as effective instructional leaders. An effective instructional leader needs the skills to read and interpret assessment results to improve instruction, guide decisions about program selection, curriculum arrangement, and professional development for teachers and school resources within his or her building. As a current Rocketship United Academy teacher I want to support Rocketship education in its mission to close the achievement gap of our lifetime.

The Classroom Assessment Literacy Inventory (CALI) contains a consent form, demographic questions, and scenarios followed by questions concerning Assessment Literacy. These questions concentrate on extracting the knowledge, perceptions, and educational practices as they relate to Assessment Literacy, through an online survey. No risks beyond the standard risks associated with responding to an online survey will be experienced by the participants.

To complete the survey online, please go to the URL <https://www.surveymonkey.com/r/HOENScali> and follow the online survey instructions. You will be prompted to complete the Consent Form prior to beginning the survey. Your answers will be completely confidential. The results of the survey will be reported in a summary format, so again no one will link you to your responses.

Thank you in advance for your participation in this important research study. If you have any questions about the administration of the survey, please contact Mareen Pfeiffer-Hoens, Middle Tennessee State University, at mp2r@mtmail.mtsu.edu.

Sincerely,
Mareen Pfeiffer-Hoens, M. Ed.

APPENDIX C

CLASSROOM ASSESSMENT LITERACY INVENTORY

[Adapted from the *Teacher Assessment Literacy Questionnaire* (1993), by Barbara S. Plake & James C. Impara, University of Nebraska-Lincoln, in cooperation with The National Council on Measurement in Education & the W.K. Kellogg Foundation] **Craig A. Mertler, Ph.D. Bowling Green State University**

Description of the inventory:

This inventory consists of two parts. Part I consists of items related to your background as an educator. Part II consists of 35 items related to the seven “Standards for Teacher Competence in the Educational Assessment of Students.” Some of the items are intended to measure general concepts related to testing and assessment, including the use of assessment activities for assigning student grades and communicating the results of assessment to students and parents; other items are related to knowledge of standardized testing and the remaining items are related to classroom assessment.

Directions: Please read each item carefully and select the response you think is the best one by shading the corresponding circle. Even if you are not sure of your choice, but you think you know which is best, mark that response.

1. Which of the following is the **most appropriate** description of the level at which you teach?

- elementary – primary (K – grade 3)
- elementary – intermediate (grades 4 – 6)

2. Which **best** describes the educational level you have attained?

- B.A. or B.S.
- M.A. or M.S.
- Ed. Specialist
- Ed.D./Ph.D.
- other _____

3. **Including the current year**, how many years of experience do you have as a classroom teacher?

- 1 – 5 years
- 6 – 10 years
- 11 – 15 years
- 16 – 20 years
- 21 – 25 years

- 26 – 30 years
- more than 30 years
- other _____

4. **Including the current year**, how many years since you completed your teacher preparation program?

- 1 – 5 years
- 6 – 10 years
- 11 – 15 years
- 16 – 20 years
- 21 – 25 years
- 26 – 30 years
- more than 30 years
- other _____

5. **Including the current year**, how many years since you completed your principal preparation program?

- 1 – 5 years
- 6 – 10 years
- 11 – 15 years
- 16 – 20 years
- 21 – 25 years
- 26 – 30 years
- more than 30 years
- other _____

6. **Including the current year**, how many years of experience do you have as a principal or assistant principal?

- 1 – 5 years
- 6 – 10 years
- 11 – 15 years
- 16 – 20 years
- 21 – 25 years
- 26 – 30 years
- more than 30 years
- other _____

7. To the best of your knowledge, did you take a **standalone** course in classroom assessment as part of your undergraduate teacher preparation?

- yes
- no
- other _____

PART II

1. What is the **most** important consideration in choosing a method for assessing student achievement?

- The ease of scoring the assessment.
- The ease of preparing the assessment.
- The accuracy of assessing whether or not instructional objectives were attained.
- The acceptance by the school administration.

2. When scores from a standardized test are said to be “reliable,” what does it imply?

- Student scores from the test can be used for a large number of educational decisions.
- If a student retook the same test, he or she would get a similar score on each retake.
- The test score is a more valid measure than teacher judgments.
- The test score accurately reflects the content of what was taught.

3. Mrs. Bruce wished to assess her students’ understanding of the method of problem solving she had been teaching. Which assessment strategy below would be **most** valid?

- Select a textbook that has a “teacher’s guide” with a test developed by the authors.
- Develop an assessment consistent with an outline of what she has actually taught in the class.
- Select a standardized test that provides a score on problem solving skills.
- Select an instrument that measures students’ attitudes about problem solving strategies.

4. What is the **most** effective use a teacher can make of an assessment that requires students to show their work (e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion)?

- Assigning grades for a unit of instruction on problem solving.
- Providing instructional feedback to individual students.
- Motivating students to attempt innovative ways to solve problems.
- None of the above.

5. Ms. Green, the principal, was evaluating the teaching performance of Mr. Williams, the fourth grade teacher. One of the things Ms. Green wanted to learn was if the students

were being encouraged to use higher order thinking skills in the class. What documentation would be the **most** valid to help Ms. Green to make this decision?

- Mr. Williams' lesson plans.
- The state curriculum guides for fourth grade.
- Copies of Mr. Williams' unit tests or assessment strategies used to assign grades.
- Worksheets completed by Mr. Williams' students, but not used for grading.

6. A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the **best** evidence for this purpose?

- Have other teachers judge whether the assessment strategy covers what was taught.
- Match an outline of the instructional content to the content of the actual assessment.
- Let students in the class indicate if they thought the assessment was valid.
- Ask parents if the assessment reflects important learning outcomes.

7. Which of the following would most likely **increase** the reliability of Mrs. Lockwood's multiple choice end-of-unit examination in physical science?

- Use a blueprint to develop the test questions.
- Change the test format to true-false questions.
- Add more items like those already on the test.
- Add an essay component.

8. Ms. Gregory wants to assess her students' skills in organizing ideas rather than just repeating facts. Which words should she use in formulating essay exercises to achieve this goal?

- compare, contrast, criticize
- identify, specify, list
- order, match, select
- define, recall, restate

9. Mr. Woodruff wanted his students to appreciate the literary works of Edgar Allen Poe. Which of his test items shown below will **best** measure his instructional goal?

- "Spoke the raven, nevermore" comes from which of Poe's works?
- True or false: Poe was an orphan and never knew his biological parents.
- Edgar Allen Poe wrote: 1. Novels 2. Short stories 3. Poems 4. All of the above.
- Discuss briefly your view of Poe's contribution to American literature.

10. Several students in Ms. Atwell’s class received low scores on her end-of-unit test covering multi-step story problems in mathematics. She wanted to know which students were having similar problems so she could group them for instruction. Which assessment strategy would be **best** for her to use for grouping students?

- Use the test provided in the “teacher’s guide.”
- Have the students take a test that has separate items for each step of the process.
- Look at the student’s records and standardized test scores to see which topics the students had not performed well on previously.
- Give students story problems to complete and have them show their work.

11. Many teachers score classroom tests using a 100-point percent correct scale. In general, what does a student’s score of 90 on such a scale mean?

- The student answered 90% of the items on this test correctly.
- The student knows 90% of the instructional content of the unit covered by this test.
- The student scored higher than 90% of all the students who took the test.
- The student scored 90% higher than the average student in the class.

12. Students in Mr. Jakman’s science class are required to develop a model of the solar system as part of their end-of-unit grade. Which scoring procedure below will **maximize** the objectivity of assessing these student projects?

- When the models are turned in, Mr. Jakman identifies the most attractive models and gives them the highest grades, the next most attractive get a lower grade and so on.
- Mr. Jakman asks other teachers in the building to rate each project on a 5-point scale based on their quality.
- Before the projects are turned in, Mr. Jakman constructs a scoring key based on the critical features of the projects as identified by the highest performing students in the class.
- Before the projects are turned in, Mr. Jakman prepares a model or blueprint of the critical features of the product and assigns scoring weights to these features. The models with the highest scores receive the highest grade.

13. At the close of the first month of school, Mrs. Friend gives her fifth grade students a test she developed in social studies. Her test is modeled after a standardized social studies test. It presents passages and then asks questions related to understanding and problem definition. When the test was scored, she noticed that two of her students—who had been performing well in their class assignments—scored much lower than other students. Which of the following types of additional information which would be most helpful in interpreting the results of this test?

- The gender of the students.
- The age of the students.
- Reliability data for the standardized social studies test she used as the model.
- Reading comprehension scores for the students.

14. Frank, a beginning fifth grader, received a G. E. (grade equivalent score) of 8.0 on the Reading Comprehension subtest of a standardized test. This score should be interpreted to mean that Frank:

- can read and understand 8th grade reading level material.
- scored as well as a typical beginning 8th grader scored on this test.
- is performing in Reading Comprehension at the 8th grade level.
- will probably reach maximum performance in Reading Comprehension at the beginning of the 8th grade.

15. When the directions indicate each section of a standardized test is timed separately, which of the following is **acceptable** test-taking behavior?

- John finishes the vocabulary section early; he then rechecks many of his answers in that section.
- Mary finishes the vocabulary section early; she checks her answers on the previous test section.
- Jane finishes the vocabulary section early; she looks ahead at the next test section but does not mark her answer sheet for any of those items.
- Bob did not finish the vocabulary section; he continues to work on that section when the testing time is up.

16. Ms. Camp is starting a new semester with a factoring unit in her Algebra I class. Before beginning the unit, she gives her students a test on the commutative, associative, and distributive properties of addition and multiplication. Which of the following is the **most** likely reason she gives this test to her students?

- The principal needs to report the results of this assessment to the state testing director.
- Ms. Camp wants to give the students practice in taking tests early in the semester.
- Ms. Camp wants to check for prerequisite knowledge in her students before she begins the unit on factoring.
- Ms. Camp wants to measure growth in student achievement of these concepts, and scores on this test will serve as the students' knowledge baseline.

17. To evaluate the effectiveness of the mathematics program for her gifted first graders, Ms. Allen gave them a standardized mathematics test normed for third graders. To decide how well her students performed, Ms. Allen compared her students' scores to those of the third-grade norm group. Why is this an **incorrect** application of standardized test norms?

- The norms are not reliable for first graders.
- The norms are not valid for first graders.
- Third grade mathematics items are too difficult for first graders.
- The time limits are too short for first graders.

18. When planning classroom instruction for a unit on arithmetic operations with fractions, which of these types of information have **more** potential to be helpful?

Norm-referenced information: describes each student's performance relative to other students in a group (e.g., percentile ranks, stanines), or

Criterion-referenced information: describes each student's performance in terms of status on specific learning outcomes (e.g., number of items correctly answered for each specific objective).

- Norm-referenced information.
- Criterion-referenced information.
- Both types of information are equally useful in helping to plan for instruction.
- Neither, test information is not useful in helping to plan instruction.

19. Students' scores on standardized tests are sometimes inconsistent with their performances on classroom assessments (e.g., teacher tests or other in-class activities). Which of the following is **not** a reasonable explanation for such discrepancies?

- Some students freeze up on standardized tests, but they do fine on classroom assessments.
- Students often take standardized tests less seriously than they take classroom assessments.
- Standardized tests measure only recall of information while classroom assessments measure more complex thinking.
- Standardized tests may have less curriculum validity than classroom assessment.

20. Elementary school teachers in the Baker School system collectively designed and developed new curricula in Reading, Mathematics, and Science that is based on locally developed objectives and objectives in state curriculum guides. The new curricula were not matched directly to the content of the fourth grade standardized test. A newspaper reports the fourth grade students in Baker Public Schools are among the lowest scoring districts in the State Assessment Program. Which of the following would **invalidate** the comparison between Baker Public Schools and other schools in the state?

- The curriculum objectives of the other districts may more closely match those of the State Assessment.
- Other school systems did not design their curriculum to be consistent with the State Assessment test.
- Instruction in Baker schools is poor.
- Other school systems have different promotion policies than Baker.

21. Which of the following choices typically provides the **most** reliable student-performance information that a teacher might consider when assigning a unit grade?

- Scores from a teacher-made test containing two or three essay questions related directly to instructional objectives of the unit.
- Scores from a teacher-made 20 item multiple-choice test designed to measure the specific instructional objectives of the unit.
- Oral responses to questions asked in class of each student over the course of the unit.
- Daily grades designed to indicate the quality of in-class participation during regular instruction.

22. A teacher gave three tests during a grading period and she wants to weight them all equally when assigning grades. The goal of the grading program is to rank order students on achievement. In order to achieve this goal, which of the following should be **closest** to equal?

- Number of items.
- Number of students taking each test.
- Average scores.
- Variation (range) of scores.

23. When a parent asks a teacher to explain the basis for his or her child's grade, the teacher should:

- explain that the grades are assigned fairly, based on the student's performance and other related factors.
- ask the parents what they think should be the basis for the child's grade.
- explain exactly how the grade was determined and show the parent samples of the student's work.
- indicate that the grading scale is imposed by the school board and the teachers have no control over grades.

24. Which of the following grading practices results in a grade that **least** reflects students' achievement?

- Mr. Jones requires students to turn in homework; however, he only grades the odd numbered items.
- Mrs. Brown uses weekly quizzes and three major examinations to assign final grades in her class.
- Ms. Smith permits students to redo their assignments several times if they need more opportunities to meet her standards for grades.
- Miss Engle deducts 5 points from a student's test grade for disruptive behavior.

25. During the most recent grading period, Ms. Johnson graded no homework and gave only one end-of-unit test. Grades were assigned only on the basis of the test. Which of the following is the **major** criticism regarding how she assigned the grades?

- The grades probably reflect a bias against minority students that exists in most tests.
- Decisions like grade assignment should be based on more than one piece of information.
- The test was too narrow in curriculum focus.
- There is no significant criticism of this method providing the test covered the unit's content.

26. In a routine conference with Mary's parents, Mrs. Estes observed that Mary's scores on the state assessment program's quantitative reasoning tests indicate Mary is performing better in mathematics concepts than in mathematics computation. This probably means that:

- Mary's score on the computation test was below average.
- Mary is an excellent student in mathematics concepts.
- the percentile bands for the mathematics concepts and computation tests do not overlap.
- the mathematics concepts test is a more valid measure of Mary's quantitative reasoning ability.

27. Many states are revising their school accountability programs to help explain differences in test scores across school systems. Which of the following is **not** something to be considered in such a program?

- The number of students in each school system.
- The average socio-economic status of the school systems.
- The race/ethnic distribution of students in each school system.
- The drop-out rate in each school systems.

28. The following standardized test data are reported for John:

Subject -- Stanine Score

Vocabulary -- 7

Mathematics Computation -- 7

Social Studies -- 7

Which of the following is a **valid** interpretation of this score report?

- John answered correctly the same number of items on each of the three tests.
- John's test scores are equivalent to a typical seventh grader's test performance.
- John had the same percentile rank on the three tests.
- John scored above average on each of the three tests.

29. Mr. Klein bases his students' grades mostly on graded homework and tests. Mr. Kaplan bases his students' grades mostly on his observation of the students during class. A major difference in these two assessment strategies for assigning grades can **best** be summarized as a difference in:

- formal and informal assessment.
- performance and applied assessment.
- customized and tailored assessment.
- formative and summative assessment.

30. John scored at the 60th percentile on a mathematics concepts test and scored at the 57th percentile on a test of reading comprehension. If the percentile bands for each test are five percentile ranks wide, what should John's teacher do in light of these test results?

- Ignore this difference.
- Provide John with individual help in reading.
- Motivate John to read more extensively outside of school.
- Provide enrichment experiences for John in mathematics, his better performance area.

31. In some states testing companies are required to release items from prior versions of a test to anyone who requests them. Such requirements are known as:

- open-testing mandates.
- gag rules.
- freedom-of-information acts.
- truth-in-testing laws.

32. Mrs. Brown wants to let her students know how they did on their test as quickly as possible. She tells her students that their scored tests will be on a chair outside of her room immediately after school. The students may come by and pick out their graded test from among the other tests for their class. What is wrong with Mrs. Brown's action?

- The students can see the other students' graded tests, making it a violation of the students' right of privacy.
- The students have to wait until after school, so the action is unfair to students who have to leave immediately after school.
- Mrs. Brown will have to rush to get the tests graded by the end of the school day, hence, the action prevents her from using the test to identify students who need special help.
- The students who were absent will have an unfair advantage, because her action allows the possibility for these students to cheat.

33. A state uses its statewide testing program as a basis for distributing resources to school systems. To establish an equitable distribution plan, the criterion set by the State

Board of Education provides additional resources to every school system with student achievement test scores above the state average. Which cliché **best** describes the likely outcome of this regulation?

- Every cloud has its silver lining.
- Into each life some rain must fall.
- The rich get richer and the poor get poorer.
- A bird in the hand is worth two in the bush.

34. In a school where teacher evaluations are based in part on their students' scores on a standardized test, several teachers noted that one of their students did not reach some vocabulary items on a standardized test. Which teacher's action is considered **ethical**?

- Mr. Jackson darkened circles on the answer sheet at random. He assumed Fred, who was not a good student, would just guess at the answers, so this would be a fair way to obtain Fred's score on the test.
- Mr. Hoover filled in the answer sheet the way he thought Joan, who was not feeling well, would have answered based on Joan's typical in-class performance.
- Mr. Stover turned in the answer sheet as it was, even though he thought George, an average student, might have gotten a higher score had he finished the test.
- Mr. Lund read each question and darkened in the bubbles on the answer sheet that represented what he believed Felicia, a slightly below average student, would select as the correct answers.

35. Mrs. Overton was concerned that her students would not do well on the State Assessment Program to be administered in the Spring. She got a copy of the standardized test form that was going to be used. She did each of the following activities to help increase scores. Which activity was **unethical**?

- Instructed students in strategies on taking multiple choice tests, including how to use answer sheets.
- Gave students the items from an alternate form of the test.
- Planned instruction to focus on the concepts covered in the test.
- None of these actions are unethical.

APPENDIX D

MEANS COMPARISON OF CALI SCORES

Table 9

Means Comparison of CALI Scores

Standard	Plake et al. (1993)	Mertler (2003)	Current Study
Standard 1 <i>Choosing Appropriate Assessment Methods</i>	3.46	3.74	3.44
Standard 2 <i>Developing Appropriate Assessment Methods</i>	3.22	3.18	3.51
Standard 3 <i>Administering, Scoring, and Interpreting the Results of Assessments</i>	3.96	3.95	2.68
Standard 4 <i>Using Assessment Results to Make Decisions</i>	3.40	3.36	3.65
Standard 5 <i>Developing Valid Pupil Grading Procedures</i>	3.19	2.06	2.86
Standard 6 <i>Communicating Assessment Results</i>	2.70	2.57	2.69
Standard 7 <i>Recognizing Unethical or Illegal Practices</i>	3.26	3.10	1.4
Total Score	23.20	21.96	20.23
Participant Number	N = 555	N = 197	N = 63

APPENDIX E

SUMMARY OF CALI DATA OF COMPARISON GROUPS

Table 10

Summary of CALI Data of Comparison Groups

Highest completed educational degree	Bachelor degree (n=32)	Master degree + (n=31)
	<ul style="list-style-type: none"> • 19.86 items correctly (M=19.86) • Scored higher than Master participants on 4 Standards (2,5,6,7) • Standard 7 is lowest standard out of all 7 	<ul style="list-style-type: none"> • 19.88 items correctly (M=19.88) • Scored higher than Bachelor participants on 3 Standard (1,3,4) • Standard 7 is lowest standard out of all 7
Years of experience	0-5 years' experience (n=39)	6 + years' experience (n=24)
	<ul style="list-style-type: none"> • 19.44 items correctly (M = 19.44) • only scored higher than comparison group on two standards (2,5) • Standard 7 is lowest standard out of all 7 	<ul style="list-style-type: none"> • 20.91 items correctly (M = 20.91) • outscored 0-5 experience group on five Standards (1,3,4,6,7) • Standard 7 is lowest standard out of all 7
Experience in a stand-alone course in assessment	Stand-alone course (n=18)	No stand-alone course (n=45)
	<ul style="list-style-type: none"> • 19.71 items correctly (M=19.71) • Participants scored higher on all 7 Standards • Standard 7 is lowest standard out of all 7 	<ul style="list-style-type: none"> • 19.66 items correctly (M=19.66) • Participants scored lower on all 7 Standards • Standard 7 is lowest standard out of all 7
Type of Teacher Ed. Program	Traditional Teacher Ed. Program (n=27)	Alternative Teacher Ed. Program (n=36)
	<ul style="list-style-type: none"> • 20.78 items correctly (M=20.87) • Participants scored higher on 5 out of the 7 Standards (1,3,4,6) • Standard 7 is lowest standard out of all 7 	<ul style="list-style-type: none"> • 20.32 items correctly (M=20.32) • Participants scored higher than comparison group on 2 Standards (2,5) • Standard 7 is lowest standard out of all 7