



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2010

Elementary Teachers' Assessment Beliefs and Practices

Sarah Calveric
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Educational Leadership Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/2332>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

Elementary Teachers' Assessment Beliefs and Practices

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Sarah B. Calveric

Bachelor of Science in Elementary and Special Education, SUNY Geneseo, 1997
Master's in Administration and Supervision, Virginia Commonwealth University, 2001

Director: Dr. James McMillan,
Department Chair, School of Education

Virginia Commonwealth University
Richmond, Virginia
November 22, 2010

Acknowledgement

The author wishes to extend heartfelt gratitude to several people. First and foremost, I would like to thank my husband, Joe, for his unwavering belief in my abilities, selfless and numerous sacrifices, and keen focus on what is most valuable in life. I would like to thank my three children, Ethan, Luke, and Emma for centering me through regular injections of desperately needed and appreciated laughter and unconditional love. To my parents, Bob and Sheila, the completion of this degree is a reflection of the fundamental teachings you have provided me with over the years: devotion, dedication, and diligence. Your immeasurable emotional and physical support has enabled me to capitalize on every opportunity life has presented. Finally, to Dr. McMillan who has spurred me forward and enlightened me to the many intricacies associated with this venture. This degree serves as the tangible and culminating product associated with the impact each one of you has had on me over the last four years. Thank you for being my champions.

Table of Contents

List of Tables.....	vi
List of Figures.....	ix
Abstract.....	x
Introduction.....	1
Purpose of the Study.....	3
Professional Significance of the Study.....	4
Overview of the Methodology.....	5
Limitations.....	7
Definitions.....	9
Organization of the Dissertation.....	9
Literature Review.....	11
Introduction.....	11
Assessment History.....	12
Assessment Literacy.....	13
Conceptions of Assessment.....	15
Assessment Practices.....	22
Defining Formative Assessment.....	22
Models of Formative Assessment.....	25
Defining Summative Assessment.....	27
Models of Summative Assessment.....	28

Formative and Summative Assessment Practices.....	29
Assessment Professional Development.....	37
Summary of the Literature Review.....	40
Methodology.....	41
Introduction.....	41
Design.....	42
Population and Sample.....	43
Instrumentation.....	44
Data Collection.....	48
Data Analysis.....	49
Summary of the Methodology.....	50
Results.....	53
Overview.....	53
Rate of Return of Surveys.....	54
Missing Data.....	55
Descriptive Data for Demographic Information.....	55
Research Question 1.....	59
Research Question 2.....	62
Research Question 3.....	65
Years of Experience.....	66
Grade Level Assignment.....	71

Degree Attainment.....	75
Level of Assessment Training.....	81
Research Question 4.....	86
Summary.....	87
Conclusions and Implications.....	91
Overview.....	91
Discussion.....	92
Assessment Beliefs.....	92
Value of Assessment Practices.....	95
Demographics and Assessment Beliefs and Importance of Practices.....	98
Assessment Beliefs and Importance of Practices.....	101
Limitations.....	103
Recommendations.....	104
Implications for Practice.....	104
Implications for Further Study.....	106
Concluding Thoughts.....	107
List of References.....	109
Appendix A.....	117
Appendix B.....	123
Appendix C.....	124
Vita.....	125

List of Tables

1. Comparison of Lower- and Upper-Grade Teachers' Use of Assessment Approaches.....	30
2. Means and Standard Deviations of All Items Measuring Assessment Practices for Elementary Teachers.....	31
3. Assessment Practice Frequency.....	33
4. Teachers' Conceptions of Assessment.....	35
5. Research Questions and Data Analyses.....	49
6. Rate of Return of Teacher Surveys.....	54
7. Descriptive Statistics for School Districts.....	56
8. Descriptive Statistics for Gender.....	56
9. Descriptive Statistics for Age Range.....	56
10. Descriptive Statistics for Teaching Assignment (Grade Level).....	57
11. Descriptive Statistics for Years of Experience.....	57
12. Descriptive Statistics for Assessment Training.....	58
13. Descriptive Statistics for Degree Attainment.....	58
14. COAIII Survey Item Sub-categories.....	60
15. Descriptive Statistics for Beliefs Subgroups.....	61
16. Correlations of Assessment Belief Subgroups.....	62
17. Descriptive Statistics for Frequency and Percent for Value of Assessment Practices.....	64
18. Descriptive Statistics for Value of Assessment Practices by Mean.....	65

19. Comparison of Assessment Belief Mean Scores for Years of Teaching Experience.....	66
20. Comparison of Assessment Practice Means by Years of Teaching Experience.....	68
21. ANOVA of Assessment Practices for Years of Teaching Experience.....	70
22. Comparison of Classroom Assessment Belief Mean Scores for Grade Level.....	71
23. Assessment Practices Means by Grade Level Assignment.....	72
24. ANOVA for Assessment Practices by Grade Level.....	74
25. Bonferroni Post Hoc for Assessment Practice (Projects by Team and Grade Levels).....	75
26. Comparison of Classroom Assessment Belief Mean Scores for Degree Attainment.....	76
27. ANOVA of Assessment Beliefs for Degree Attainment.....	77
28. Bonferroni Post Hoc for Assessment Belief (Student Accountability) and Degree Attainment.....	78
29. Assessment Practices Means by Degree Attainment.....	79
30. ANOVA for Assessment Practices by Degree Attainment.....	80
31. Bonferroni Post Hoc for Assessment Practice (Authentic Assessment) and Degree Attainment.....	81
32. Comparison of Classroom Assessment Belief Mean Scores for Types of Assessment Training.....	82
33. Independent Samples t-tests for Assessment Beliefs and Types of Assessment Training.....	83
34. Comparison of Means for Assessment Practices by Assessment Training.....	84

35. Independent Samples t-tests for Assessment Practices and Types of Assessment Training.....	85
36. Correlations of Assessment Belief Subgroups and Value of Assessment Practices.....	87
37. Comparison of Belief Subgroups' Correlation Coefficients: 2007 versus 2010.....	94

List of Figures

1. Models of Assessment Practices.....

Abstract

ELEMENTARY TEACHERS' ASSESSMENT BELIEFS AND PRACTICES

By Sarah B. Calveric, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2010

Director: Dr. James McMillan
Foundations of Education
School of Education

Increased state and federal accountability measures have made the assessment of student performance one of the most critical responsibilities of classroom teachers; yet, inadequate opportunities for preservice and inservice training leave many teachers feeling ill-prepared for this task. Adding to the complexity of building teachers' assessment literacy is the relationship between assessment beliefs and classroom assessment practices. This quantitative study utilizes a validated, online survey to examine how elementary teachers ($n = 79$) define their assessment beliefs (conceptions) and how these beliefs influence which assessment practices are valued within the classroom. Findings suggest that despite teachers' limited exposure to assessment training, four distinct assessment beliefs exist within the elementary classroom: assessment for school accountability, assessment for student certification, assessment for improvement of

teaching and learning, and assessment as irrelevant. Assessment for the improvement of teaching and learning yielded the highest composite mean and was negatively correlated with the irrelevance belief and positively related to school accountability. An analysis of the importance of assessment practices revealed authentic assessments, short answers, teacher-made assessments, and performance assessments as the most valued, while publisher assessments and major exams had the lowest means. Significant relationships were identified between demographics and beliefs and practices, with the most practical findings related to exposure to assessment training and level of degree attainment. Significant relationships were also noted between all beliefs and the value of specific assessment practices, with the exception of the irrelevance belief. No significant relationships were noted between the irrelevant belief and value of assessment practices; however, many negative correlations were documented. Results are discussed in light of other research, indicating that a greater understanding of assessment beliefs and importance of practices can contribute to the development of relevant professional development aimed at the improvement of teachers' assessment pedagogies and practices can contribute to greater educational success.

Key words: assessment, beliefs, conceptions, formative assessment, summative assessment, assessment literacy, professional development

Chapter 1

Introduction

With the passage of the *No Child Left Behind Act of 2001*, accurate assessment of student achievement is becoming increasingly vital at the district, state, and national levels (Popham, 2005). As a result, public and political interests demand that teachers be held accountable for increased student achievement (Campbell, Murphy, & Holt, 2002). Despite this apparent emphasis, Black and Wiliam 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). This disconnect between national mandates and teacher assessment practices provides necessary evidence for the promotion of “professional development in assessment that acknowledges the place of both classroom assessment and official assessment in supporting teaching and learning” (O’Leary, 2008, p. 109).

To produce meaningful professional development related to assessment, it is necessary for research to document a common understanding of what constitutes assessment literacy. Assessment literacy is defined by Fullan (2001) as the teacher’s capacity to examine student performance evidence and discern quality work through the analysis of achievement scores and disaggregation of data. Additionally, Fullan summarizes the need for teachers to be knowledgeable in the formation of action plans aimed at increased student achievement. Fullan’s final capacity associated with teacher assessment literacy relates to educators’ contributions to

political agendas associated with high-stakes testing and achievement data use. Chapman (2008) defines assessment literacy as the possession of essential knowledge and understanding of test characteristics and properties, which is developed from the practice, use, and interpretation of outcome data in making educational decisions.

As the importance of assessment literacy increases, expectations regarding teachers' classroom practices have undergone a paradigm shift (Hargreaves, Earl, & Schmidt, 2002). The recent focus on the concept of assessment literacy has drawn attention to the importance of educators incorporating various assessment practices such as formative (assessment for learning) and summative (assessment of learning) methodologies (Stiggins, 2002b). Educators are expected to be skilled assessment practitioners, designing and interpreting more student-involved classroom assessments, often termed as assessment to improve learning (Guskey, 2003). If competent, teachers can then utilize "assessment-gathered evidence" (Popham, 2008, p. 7) to better gauge the effectiveness of instruction and student progress (Campbell et al., 2002).

Ayalla et al. (2008) found that assessment literacy and assessment reform require significant preparatory measures. To meet the demands of the current accountability era, more needs to be clarified regarding educational measurement and assessment's fundamental underpinnings (Stiggins, 1991a; Stiggins & Bridgeford, 1985; Stiggins & Conklin, 1988). Specifically, researchers note the need for further exploration of the relationship between teachers' beliefs and their assessment practices (Adams & Hsu, 1998; Winterbottom et al., 2008). As a result, there is a pressing need for researchers to gather information from practicing educators about their conceptions (beliefs) of assessment, current classroom assessment practices, and the resulting relationship among the two variables. It is anticipated that the results of this quantitative study will contribute to a comprehensive understanding of teachers'

assessment beliefs and their relationship to assessment practices deemed important by elementary classroom.

Purpose of the Study

Assessment is considered to be a critical component in the process of teaching and learning as it enables educators to evaluate student learning and utilize the information to improve learning and instruction (Harris, Irving, & Peterson, 2008). As a result, Brookhart (1999) emphasizes the importance of teachers using assessments that are valid, reliable, meaningful, and accurate to guide instruction. Mertler (2006) suggests that lack of exposure to assessment fundamentals helps to explain why teachers do not readily recognize the importance of assessment to improved instruction, student motivation, and level of student achievement. Educators must acquire a more sophisticated understanding of assessment literacy necessary for utilizing data to diagnose needs of individual students (Zwick et al., 2008). Despite its seemingly obvious relation to the enhancement of instruction, a lack of training in assessment fundamentals has been documented by researchers and may be the weak link in driving America toward improving education (Airasian, 1991; Cizek, Fitzgerald, & Rachor, 1996; Stiggins, 1991b).

The purpose of this study was to investigate perceptions of a group of elementary teachers from within the Commonwealth of Virginia regarding their assessment conceptions (beliefs) and practices. The researcher analyzed the data to seek greater understanding of how 3rd through 5th grade teachers' assessment beliefs relate to assessment practices. The independent variables of years of teaching experience, age, grade level assignment, level of education, and exposure to an assessment training were used to further identify the relationships between the variables of teachers' assessment beliefs and practices.

The research questions driving this study were:

1. What are elementary teachers' conceptions (beliefs) about assessment?
2. What classroom assessment practices are valued by 3rd -5th grade elementary teachers?
3. What is the relationship between years of experience, grade level assignment, level of education, and assessment training and teachers' assessment beliefs and practices?
4. How do teachers' assessment beliefs relate to the value of assessment practices?

The Professional Significance of the Study

Despite the increased emphasis placed on testing and data-driven decision-making, assessments of teacher preparatory programs underscore gross inadequacies that have led to an ill-prepared pool of teacher candidates (Kirkpatrick, Lincoln, & Morrow, 2006). Campbell et al. (2002) report that many colleges of education and state education agencies continue to require pre-service teachers to complete minimal, if any, specific coursework in classroom assessment. Resulting research documents that teacher assessment skills are traditionally inadequate (Campbell et al., 2002; Cizek et al., 1996), and that many educators claim feelings of ill-preparedness in association with assessment literacy (Kirkpatrick et al., 2006; Mertler, 1999). The resulting number of classroom teachers stating they exited college education programs unprepared to assess student learning, leaves Kirkpatrick et al. and Stiggins (1999) reiterating the need for continued analysis of recent graduates' feedback to discern what preparatory program changes are necessary to enhance assessment literacy.

Cizek et al. (1996) conducted a survey of 143 elementary and secondary teachers to gather data on several assessment-related practices. Similar to Kirkpatrick et al. (2006) and

Stiggins (1999), Cizek et al. found teachers and administrators entering the educational field without systematic training in assessment. More specifically, this study confirmed the generally acknowledged weakness in pre-service and in-service preparation of teachers in classroom assessment and that additional assistance is necessary.

This research study explored teachers' conceptions of assessment and assessment practices and how these dependent variables related to one another and participants' demographic descriptors (independent variables). The study's findings are intended to more clearly define teachers' beliefs associated with assessment and how these beliefs relate to teachers' assignment of value to various assessment practices. Study results may inform a variety of stakeholders who play a role in the education of the Commonwealth's children.

Understanding current assessment beliefs and practices and formulating relevant professional development aimed at the improvement of teachers' assessment pedagogies and practices can contribute to greater educational success.

Overview of Methodology

The teacher participants in this study will be selected through nonprobability sampling to ensure participants are accessible, representative of the population, and represent certain selection criteria: elementary instructors of grades three through five. Specifically, the researcher will use purposive sampling to ensure she identifies information rich subjects who are regularly charged with responsibilities associated with the topic of interest, elementary assessment beliefs and practices.

In order to obtain quantitative data regarding elementary teachers' assessment beliefs and practices, a validated survey was conducted to generate data regarding teachers' assessment beliefs and practices and the following independent variables: years of teaching experience, age,

grade level assignment, level of education, and exposure to an assessment training. Survey Monkey was selected as a survey tool to collect data from teachers on the revised scales: Conceptions of Assessment Abridged III (Brown, 2006) and Classroom Assessment Practices (McMillan, Myran, & Workman, 2002) (Appendix A).

The Human Subjects Research Protocol was submitted for approval by the Institutional Review Board (IRB) at Virginia Commonwealth University prior to the survey being emailed. After obtaining IRB approval and having requested by letter and receiving permission from central administration of each locality, the survey was emailed to principals, either by the researcher or a district representative, for their preview (Appendix B). Approximately one week later, a second email was distributed, either by the researcher or a district representative, inviting administrators of participating buildings to forward the email to the identified sample population. This letter outlined the purpose of the study, confidentiality procedures, and directions associated with the completion of the online survey (Appendix C).

In the initial analysis of data, descriptive measures were compiled and between group tests were completed. Specifically, research questions one and two were analyzed using descriptive statistics such as frequencies, means, and standard deviations. Data were calculated for each subscale related to teachers' conceptions of assessment and assessment practices. Inferential statistics were conducted to test for differences among teachers' assessment beliefs and practices and demographic information (independent variables), specifically age, grade level assignment, years of experience, level of education, and type of assessment training. The fourth question required the researcher to run correlations aimed at determining the relationship between beliefs and practices. The specifics of the methodology are discussed more completely at a later point in the dissertation.

Limitations

The present study focused on a target population consisting of approximately 762 teachers at fifty-nine elementary schools within two school divisions. A limited sample population consisting of ten elementary schools with 124 third through fifth grade teachers posed certain limitations that need to be taken into account when considering the study and its contributions. Although the overall response rate was 64%, this figure is representative of only 79 teachers and lessens the generalizability of results.

An innate limitation of the study's results is that they rely on teachers' self-reported data. Social desirability may have influenced teachers' responses despite the promotion of anonymity during the survey's administration. Participants may also never have participated in self-reflection in relation to their assessment beliefs and practices, which can result in responses which represent something the participants may not fully know. As a result, obeying demand characteristics or supplying answers the respondents believe the researcher desires may have resulted.

The researcher must also consider the impact associated with the time frame identified for survey distribution and completion within each district. Specifically, both participating school districts place a moratorium on all research studies during the Virginia Standards of Learning (SOL) assessment window, mid-March through early June. As a result, to access identified participants prior to departure for summer break, the researcher needed to distribute the survey in one county beginning the week after the conclusion of SOL testing and two weeks prior to the start of summer vacation. The second participating district's study window fell during the last week of school and during the first week of summer vacation. Although the researcher verified

teachers' ability to access county-provided computers during this final week, low school participation (sample population) rates may reflect the impact of the distribution window.

Beyond the proximity of the study window to the end of the school year, conducting the research immediately following the Virginia SOL assessments can also unearth certain limitations. In the sense that the participants had just spent the preceding weeks executing standardized testing with students, the researcher must consider the impact this had upon the teachers' response style and assessment beliefs. It could be suggested that future research on this topic may reveal different results, especially related to survey items reflecting more traditional, summative assessment practices.

Another limitation of this study is the constraints on generalizability and utility of findings. External validity in this research could have been maximized by securing responses from a larger sampling of participants. Secondly, because people's behaviors may change or be biased depending on the setting or situation, results may not hold true in an alternate environment. This concern is heightened in this study due to the time frame in which the study was implemented, and could result in results obtained under this study's implementation period not generalizing to a setting in which a high-stakes assessment has recently concluded or an extended vacation period is imminent. Therefore, to maximize external validity, future researchers may repeat the study using a different instrument distribution window.

A further limitation of this study is that assessment research indicates that teachers' conceptions are described in a one dimensional perspective. Generally, teachers are believed to have one particular assessment belief; however, it is probable that teachers endorse multiple conceptions of assessment and that these intermingle with one another (Brown, 2003). Additionally, respondents' multi-faceted views of assessment beliefs may also have caused

confusion regarding response style. Teachers may have struggled with determining whether their responses should reflect what they personally believe, what assessment should be, or what assessment currently is.

Definitions

Within the context of this study, the following definitions will be used:

Assessment – 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

Assessment literacy – the possession of essential knowledge and understanding of test characteristics and properties, which is developed from the practice, use, and interpretation of outcome data in making educational decisions

Assessment for learning – use of the classroom assessment process and resulting information to advance, not merely check on, student learning.

Conceptions – a framework or mental structure, encompassing beliefs, through which a teacher views, interprets and interacts with the instructional environment; in this study the words conceptions and beliefs are used interchangeably

Formative assessment – formative assessment data provide measurements of student progress toward a particular goal within a curricular unit and are used by students and instructors to guide further instruction and learning

Professional development – formal learning opportunities provided to teachers to improve their knowledge, skills, and classroom practices

Summative assessment – assessment conducted at the end of a period of learning to determine if students have learned what was taught to them

Organization of the Dissertation

This quantitative study was designed to explore, describe, and examine third through fifth grade teachers' conceptions of assessment and assessment practices and determine whether relationships exist between or among the dependent (beliefs and practices) and independent variables (age, grade level assignment, teaching experience, level of education, and exposure to assessment training). Because high-stakes testing and accountability have been the catalyst for enhanced assessment literacy, researchers have revealed the need for extensive teacher preparation and training in educational measurement. Data from this study offered a greater understanding of how 3rd-5th grade teachers' assessment beliefs and practices relate to one another and can serve to better inform the development of assessment professional development.

The first chapter of the dissertation includes an introduction, purpose of the study, the professional significance of the study, overview of the methodology, limitations, definitions of key terms, and organization of the dissertation. The second chapter of the dissertation expands upon the review of literature associated with theoretical and empirical research relating to assessment history, assessment literacy, conceptions (beliefs) of assessment, and their resulting assessment practices. Additionally, the literature review explores the effectiveness of preservice and inservice teachers' assessment development, as well as research related to determinants of effective assessment professional development. A summary of the literature review concludes chapter two.

The third chapter of the dissertation describes the methodology used in the study. It includes an overview of the methodology, design of the study, context of the research, instrumentation, data collection, analysis of the data, and summary of the methodology. The fourth chapter presents a review of the research design, instrumentation, data collection, analysis of the data by

research question, and summary of the results. The final chapter, chapter five, contains an overview, summary of the results, discussion of the results, and implications for practice and further research. Concluding the dissertation are references and appendices.

Chapter 2

Literature Review

Introduction

Just as research related to classroom teachers' instructional practices and beliefs is intertwined with various factors, the literature associated with assessment is also interwoven with other facets of education. As a result, this review of literature included studies and readings in the following areas: assessment history, assessment literacy, conceptions (beliefs) of assessment, models of assessment practices, and assessment preparatory measures (preservice experiences and professional development).

The review of literature was conducted through various means. Electronic searches were conducted through ERIC EBSCOhost, Education Abstracts, and Dissertation Abstracts, as well as ProQuest and Education Policy Analysis Archives. Books, dissertations, and journal references were obtained at the James Branch Cabell Library at Virginia Commonwealth University (VCU). Brown, designer of the abridged Conceptions of Assessment-III (2003) survey, responded to multiple questions about the instrument via e-mail and provided multiple articles relevant to the instrument and study. Websites of educational organizations, such as

Mid-continent Research for Education and Learning (McREL), National Association of Test Directors (NATD), and American Educational Research Association (AERA) were used. A footnote and reference search of sources cited within the reviewed studies and articles revealed additional pathways for further research. Several books, studies, and publications were ordered through the Association for Supervision and Curriculum and Development (ASCD), the Metropolitan Educational Research Consortium (MERC), and the National Association for Elementary School Principals (NAESP). Additionally, the researcher consulted with numerous education research experts throughout the literature review process.

Assessment History

In viewing the evolution of assessment over time, changes in the perspective of assessment and the introduction of the *No Child Left Behind Act of 2001* (NCLB) have required K-12 education to increase its focus on accountability measures. The grand scale and aim of NCLB raised tremendous debate amongst politicians, educators, and the general public. Passed in 2001 with bipartisan consensus, this federal mandate set forth revolutionary methods for high achievement through the promotion of steadily progressing achievement standards, frequent testing to ascertain progress, and accountability of subgroups (Cowan, 2004).

The advent of this more centralized assessment system added numerous federal requirements to existing local and state assessment programs. Although states have customarily controlled educational happenings, NCLB demonstrates a significant expansion of federal authority and the daunting and complex difficulties associated with understanding the federal mandates. Localities have faced the challenge of devising systems that comply with NCLB, while ensuring that their assessment systems remain in alignment with state and local objectives. Additionally, NCLB's accompanying restrictions and constraints are perceived by many as

hindrances to success. Sanctions such as loss of funding, public embarrassment for not meeting the 2013-2014 proficiency deadline, restructuring, take-overs, school choice, and voucher supplemental services have bred desire to abandon federal funding; however, for most public education institutes, forgoing supplemental federal aid is not a realistic option.

As a result, heightened district and teacher accountability has required districts to align state standards and tests while investigating alternative assessment formats to gain more frequent data to drive instructional decisions and financial appropriations (Bangert & Kelting-Gibson, 2006; Delandshere & Jones, 1999). To successfully attain mandated achievement targets, educational organizations must investigate what teachers' conceptions of assessment are and how these beliefs relate to assessment practices. The resulting relationship will inform researchers on how to best proceed with the development of more meaningful and relevant professional development related to heightened assessment literacy.

Assessment Literacy

The *No Child Left Behind* era requires that all educators, at local, state, and national levels, have a sophisticated understanding of assessment (Popham, 2005). Since it has been estimated that teachers spend up to 50 percent of their time on assessment-related activities (Plake, 1993; Stiggins, 1999), researchers continue to emphasize the importance of principals and teachers being adequately trained to use data to modify daily instruction, individualize assistance for identified students, and communicate results to educational stakeholders (Zwick et al., 2008). The following assessment literacy research describes ways in which teachers should use assessment results to make ongoing instructional adjustments and inform decision-making (Campbell et al., 2002; Popham, 2005; Stiggins, 2002; Zwick et al., 2008).

In 1995, Stiggins publicized the importance of assessment literacy for improving the current status of classroom assessment. Stiggins (1991a) defined assessment literacy as a deep understanding of the uses and limitations of the full range of assessment options and the knowledge to select the most appropriate methods to describe the development of young children. Stiggins (1998) referred educators to the quality standards for assessment design, which indicate that effective classroom assessments stem from and serve clear purposes, reflect well-defined and appropriate achievement goals, rely on proper assessment methods, sample student achievement appropriately, and control for all related sources of bias and distortion. More specifically, he stated that assessment literates know what constitutes a high-quality assessment in alignment with a clearly defined learning target (1991a). Additionally, Stiggins (1991a) maintains that educators with sound assessment literacy understand the importance of fully assessing performance, identify potential biases or extraneous variables which may impact results, and recognize the importance of data being in meaningful forms and readily identify when the results are inadequate.

To improve instruction and raise student achievement, Boudett, City, and Murnane (2006), outline eight steps for effective use of assessment data. Boudett et al. (2006) described step two as the building of assessment literacy through the development of a working knowledge of common concepts related to test score results and the acquisition of appropriate skills to interpret assessment data. Demystifying assessment and testing enables teachers to more deeply understand the strengths and limitations associated with the range of assessment options (Jones, 2004). When appropriate assessment strategies are consistently implemented within the classroom student achievement is increased (Black & Wiliam, 1998). Ultimately, increased assessment competency can enhance teachers' abilities to inform stakeholders and hold policy

makers accountable for supporting sound assessment practices for children and the programs that serve them (Jones, 2004).

Conceptions of Assessment

The purpose of this section was to outline what teachers' identify as their conceptions of assessment. Throughout this dissertation, the words conception and beliefs were used interchangeably to represent four distinct assessment beliefs documented in the research findings: improvement of teaching and learning, assessment for student certification, assessment for school and division accountability, and assessment as irrelevant (Brown, 2003). The delineation of the characteristics associated with each conception of assessment are issues that have been discussed and studied and have yielded many articles over the last couple of decades.

Just as society and education have changed over the years, the study of opinions, beliefs, and policies regulating assessment pedagogies and practices reveal multiple transformations. Making a specific impact upon assessment are teachers' conceptions of assessment (Brown, 2003). Conceptions are defined as a framework or mental structure, encompassing beliefs (Thompson, 1992), through which a teacher views, interprets, and interacts with the instructional environment (Pratt, 1992). Despite a conceptions' individualistic appearance, Van den Berg (2002) determined conceptions to be interrelated and complex reflections of socially and culturally shared phenomena. Additionally, Abelson's (1979) research depicts a person's conceptions as individual assertions about reality, which the individual believes as truth at that moment. Since these beliefs are developed through people's experiences, researchers conclude that the conceptions are pervasive and will influence the individual's subsequent interactions with the world (Abelson, 1979; Delandshere & Jones, 1999).

It is important to study teachers' conceptions of assessment due to previously cited research documenting the impact educators' conceptions of learning and teaching have had upon instruction and achievement (Calderhead, 1996; Delandshere & Jones, 1999; Remesal, 2007; Thompson, 1992). Cizek et al. (1996) studied a sample of 143 elementary to secondary teachers to investigate any potential relations between differences in assessment practices and background characteristics such as gender, grade level, and years of teaching experience. The quantitative results uncovered noteworthy diversity among teachers' assessment perspectives and practices. Cizek et al. associated these discrepancies in practice with individual assessment policies that reflected teachers' own individualistic values and beliefs about teaching.

In another study regarding teacher conceptions, Kahn (2001) conducted research aimed at examining teacher-created assessment materials to determine what conceptions or models of teaching and learning were reflected. Kahn found his subjects', 10th grade English teachers, assessment materials to be an "eclectic mixture of approaches" (p. 284). Further analysis of the data and teacher comments revealed that some materials adopted a constructivist methodology, requiring students to construct and interpret meaning, while other assessment modalities represented a more traditional process of recalling information. Kahn concluded that teachers' assessment practices were influenced by individual beliefs or conceptions related to what constitutes learning and concerns about "maintaining student attention, cooperation, and classroom control" (p. 286).

The complexity of constructs, specifically assessment constructs, and the resulting effects upon educational pedagogies have been studied by many researchers (Brown, 2003, 2004, 2006, 2007; Brown & Hattie, 2009; Brown & Lake, 2006; Remesal, 2007). In 2003, Brown studied teachers' assessment conceptions' relationship to learning, teaching, curriculum, and teacher

efficacy. Results from a survey of 525 New Zealand primary teachers were analyzed and correlation coefficients assisted Brown (2003) with the identification of four main assessment conceptions or beliefs of assessment: improvement of teaching and learning, certification of students' learning, accountability of schools and teachers (Torrance & Pryor, 1998; Warren & Nisbet, 1999; Remesal, 2007), and the irrelevance or rejection of assessment (Airasian, 1997; Brown, 2004). It is critical for educators and policy makers to have a sound understanding of these assessment conceptions as research has documented their impact upon teaching and learning (Brown, 2004; Remesal, 2007).

Numerous studies have outlined the fundamentals associated with the conception of assessment for improvement of learning and teaching (Black & Wiliam, 1998; Delandshere & Jones, 1999; Brown, 2003; Popham, 2008). When learning is viewed as continuous development enhanced by structured and meaningful educational experiences, the resulting assessment selection is more likely to yield documentation and feedback associated with the improvement belief (Delandhsere & Jones, 1999, p. 219). Brown (2003) details this improvement conception, promoted by Black and Wiliam (1998) as assessment for learning, by describing two key indicators; (a) students' achievement or performance is depicted through assessment results and (b) the assessments yielded reliable and valid data necessary for accurately describing student performance. Under these circumstances, the purpose of assessment requires wide-ranging use of varied assessment tools, both formal and informal teacher-based, aimed at succinctly capturing students' academic profiles, "with the explicit goal of improving the quality of instruction and student learning" (p. 4).

Brown's (2003) second conception of assessment, certification of students' learning, contends that students are individually accountable for their performance and achievement on

assessments. Assessment for the purpose of determining acquisition of facts and skills is “more likely to be viewed as serving the function of sanction and verification: the student either has or has not learned the content” (Delandshere & Jones, 1999, p. 219). Due to the increasing number of student accountability measures at the secondary level and the high stakes nature of many of these assessment activities, Brown specifically emphasizes the positive and negative consequences related to students’ performance results such as graduation, grade retention, grades, and tracking.

The third conception of assessment, accountability of teachers and schools, underscores society’s use of data to determine school and teacher quality (Brown, 2003). Because much of the focus of NCLB has been around sanctions and rewards as means to increase student achievement, Englert, Fries, Martin-Glenn, and Michael (2005) discuss the importance of informing parents and the community about student progress and school status. Englert et al., measured to what degree their research participants, superintendents, principals, and teachers, were required to meet data-driven performance goals and to what degree they were evaluated based on changes in student achievement. Results indicate that superintendents largely hold the accountability of addressing the public at large regarding performance. “They are accountable for answering questions about how tax dollars are spent, answering to an elected school board, and ensuring that their district meets federal requirements” (p. 18-19). As a result, accountability measures are critical to superintendents’ daily lives and result in their need to explain their districts’ progress toward meeting NCLB’s adequate yearly progress (AYP) criteria.

In 1999, Delandshere and Jones conducted a qualitative study aimed at identifying elementary teachers’ beliefs about assessment. At the completion of 14 individual interviews with the three participants over a three month period, the researchers engaged in an analytic

induction process to generate a set of assertions that emerged from the data. Similar to other documented research (Brown, 2003; Calderhead, 1996; Cizek et al. 1996; Remesal, 2007; Thompson, 1992), teachers' beliefs about assessment are influenced by external functions and purposes. Researchers' final analysis yielded three key assertions or beliefs about the function of assessment: to place students in the accurate leveled curriculum; to formally describe students' achievement and provide justification for grades; and to serve as preparation for mandated testing.

Similar to Brown's (2003) second and third conceptions, certification of students' learning and accountability of teachers and schools, Delandshere and Jones (1999) determined the three participants' assessment views as predominantly summative and external in nature. Teacher interview responses regarded assessment as "a required means of conveying information to external audiences (parents, district, state, other teachers), and rarely as a way to understand learning and inform teaching" (p. 229). Teachers' perceptions of an externally defined assessment pedagogy, limits their assessment practices to summative approaches that imitate the state and federal-mandated testing. As an unintended consequence, Delandshere and Jones point out "teachers are left dissatisfied and unable to learn about their teaching or how their students learn" (p. 238). Additionally, the researchers surmise that teachers' assessment practices play an integral role in the preservation of their conceptions about assessment and its functions and usefulness.

Assessment as irrelevant, the fourth assessment conception, represents teachers who view assessment as unrelated to the work of educators and students (Brown, 2003). Typically associated with formal testing, educators who adopt this assessment conception reject assessment due to its perceived harmful impact upon teacher autonomy and professionalism (Brown, 2003).

Followers of the irrelevance conception believe assessment detracts from student learning and excludes the inclusion of teachers' intuitive evaluations, student-teacher relationships, and in-depth knowledge of curriculum and pedagogy (Brown, 2003).

Remesal (2007) presented research detailing thirty primary and twenty secondary teachers' conceptions of assessment. This study built upon Black and Wiliam's 2005 study which documented four nations' experiences related to teachers' conceptions of assessment and pedagogical reform. Remesal's research focus was to contribute to Black and Wiliam's previous research findings: the acknowledgement of teacher beliefs about various aspects of the instructional practice being another significant contributor to differences in assessment practices "(such as beliefs about what constitutes learning, about value of competition between students or between schools, or about the meaningfulness of tests results as indicators of school effectiveness)" (p. 28).

Remesal (2007) chronicled assessment in the Spanish educational system through the use of qualitative techniques to individually interview fifty teachers (thirty primary and twenty secondary). In addition to the interviews, the researcher conducted a content analysis of assessment materials determined by the participants to be representative of their typical classroom assessment practices. Analysis of the data demonstrated assessment conceptions similar to previously noted research (Brown, 2003; Englert et. al., 2005; Gipps, Brown, McCallum, & McAlister, 1995; Hill, 2000; Stamp, 1987). For example, Remesal detailed a continuum representative of teachers' conceptions of assessment, which flowed from one assessment extreme to the opposite extreme. Specifically, the pedagogical conception of assessment embraces the more formative assessment measures. Similar to Brown's (2003)

improvement conception, Remesal's pedagogical conception views assessment as "an instrument for improvement of teaching and learning" (p. 31).

Comparable to Brown's (2003) second and third conceptions, certification of students and accountability of teachers and schools, Remesal's (2007) accounting conception defines assessment as an instrument of social control. Opposite of the pedagogical conception, the purpose associated with the accounting conception is to certify students' final results, which characterizes this conception as a public method of monitoring teachers' instructional competencies. In between these two extremes, the researcher identified three mixed conceptions: "a mixed pedagogical conception of assessment, in which the pedagogical components predominate over the accounting ones; a mixed undefined conception of assessment, with no clear preference for one wing of the continuum or the other; and a mixed accounting conception of assessment, in which the accounting components predominate over the pedagogical ones" (p. 31).

Results of Remesal's (2007) study document the fifty teacher participants' distribution of the five assessment conceptions. Initial analyses uncover what appears to be a slightly more frequent (54%, [=38% + 16%]) global adoption of the accounting conception of assessment, both extreme and mixed, than the pedagogical conception of assessment (40% [=16% + 24%]). However, Remesal conducted an independent analysis of each educational level. Leveled results indicate that primary teachers assume the pedagogical conception of assessment (60% [=20% + 40%]), either extreme or mixed, with similar conceptions remaining more rare among secondary educators (10%). 75% of secondary teachers demonstrate inclination toward the accounting conception ([=45% = 30%]), while similar conceptions appear only 40% [=33.33% + 6/67%] among primary.

Remesal's (2007) study depicts a more balanced distribution of assessment conceptions among primary educators, while secondary teachers demonstrate a stronger inclination toward the accounting conception. This imbalance of assessment conceptions, both globally and between Spain's primary and compulsory secondary education, appears to support the researcher's concluding thoughts related to educational organizations' need to continue exploring "teachers' conceptions of assessment within and across each particular system" in order to advocate for assessment strategies "that are likely to be understood, accepted and assumed by the teachers" (p.36).

Consistently the results of these studies suggest that four main conceptions of assessment exist within the elementary classroom: improvement of teaching and learning, certification of students' learning, accountability of schools and teachers (Torrance & Pryor, 1998; Warren & Nisbet, 1999; Remesal, 2007), and the irrelevance or rejection of assessment (Airasian, 1997; Brown, 2004). Despite the varying terms used to describe these four assessment beliefs, researchers indicate that teachers' individualistic ideas and thoughts regarding assessment impact their acceptance of various assessment methodologies. To gain further information pertaining to the relationship between teachers' assessment beliefs and practices, this study incorporated survey items specifically aimed at the dependent variables. The data was analyzed to document greater understanding of how teachers' conceptions of assessment impact how educators rate various assessment methods' degree of importance within third through fifth grade classrooms.

Assessment Practices

Having identified four basic conceptions or beliefs regarding assessment, researchers have formulated models of assessment conceptions which represent potential assessment

practices or uses. The following literature expands upon researchers' investigations of teachers' classroom assessment methodologies.

Defining formative assessment (assessment for learning).

In attempt to respond to federal mandates, school districts have researched numerous assessment methodologies identified as instrumental in increasing student achievement (Stiggins, 2002a; Rhodes & Robnolt, 2007). While seeking the greatest student academic gains, educational organizations have investigated what literature has commonly termed formative assessment practices. Formative assessment is the systematic process of continuously gathering evidence about learning (Heritage, 2007). Heritage suggests that formative assessment, also known as assessment for learning (Hargreaves, 2005; Popham, 2008), utilizes data to accurately prescribe or "measure" (Hargreaves, 2005) a student's instructional level of learning and to alter lessons to assist students with attaining an identified learning goal. Additionally, formative assessment actively engages both teachers and students in learning goal development, progress monitoring, and preparation of future learning steps.

Formative assessment data provide measurement of student progress toward a particular goal within a curricular unit and are used by students and instructors to guide further instruction and learning (Black and Wiliam, 1998; Gipps, McCallum, & Hargreaves, 2000; Hargreaves, 2005; Harris et al., 2008). To more closely examine teachers' conceptions of assessment for learning, Hargreaves (2005) conducted a survey of eighty-three teachers' understanding of the phrase "assessment for learning" (p. 214). Anonymous responses were submitted and analyzed by Hargreaves to identify and group together responses with similar emphasis. Teacher quotations and classroom observation data were examined to increase validity of participants' responses and develop six summary definitions: assessment for learning means monitoring

students' performance against targets or objectives; using assessment to guide the next steps associated with teaching and learning; teachers giving feedback for improvement; teachers learning about students' learning; children taking some ownership over their own learning and assessment; and turning assessment into a learning event.

Within these definitions of assessment for learning, Hargreaves (2005) identified implicit conceptions of assessment through the identification of two distinct meanings for assessment: "assessment as measurement" and "assessment as inquiry" (p. 218). The researcher defines measurement as the act or process of determining the amount or extent of each child's learning, which is typically assessed through the use of a test. A vital aspect of measurement in assessment for learning is the marking, checking, reporting process referenced by all eighty-three study participants. The second meaning of the word assessment, assessment as inquiry, referenced action verbs such as "reflecting, reviewing, finding out, discovering, diagnosing, learning about, examining, looking at, engaging with, understanding" (p. 218). At the conclusion of this investigatory process, a heightened awareness of students as learners, not just performers, is gained. Although the assessment techniques may remain the same in the measurement and inquiry paradigms, the inquiry model underscores not only who and what is being tested, but also the assessor and the inquirer.

Dixon and Haigh (2009) reference the significant attention that has been paid to conceptualizing how teaching, learning, and assessment are interwoven and the resulting discourse related to assessment for learning. These discursive shifts have redefined students and teachers' learning and assessment roles and responsibilities (Dixon & Haigh, 2009). In the early years, formative assessment was defined as the process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learners are in their learning, where they

need to go and how best to get there (Assessment Reform Group, 2002). Current demands now require teachers to help every student develop conceptual understanding through experiential learning, ongoing assessment, and the continuous offering of meaningful feedback about work quality and the methods used to produce it (Delandshere & Jones, 1999).

Models of formative assessment (assessment for learning).

Numerous studies document the implementation of various formative assessment models to move beyond the summative documentation of students' understanding of a program of study (Henderson, Petrosino, Guckenbunrg, & Hamilton, 2007). As Figure 1 reflects, Stamp (1987) used multivariate techniques to identify three major conceptions of assessment among a sample of Australian, pre-service teachers. Stamp correlated the three conceptions of assessment with specific assessment uses or practices. Specifically, the first conception, cater for the need and progress of individual pupils, appears to be in close alignment with Brown's (2003) improvement of teaching and learning and necessitates the use of a formative assessment methodology to identify individualized learning needs.

Similarly, Gipps et al. (1995) classified educators into three main types of assessment users: intuitives, evidence gatherers, and systematic planners. Systematic planners were defined as collectors of strategically planned data reflective of curricular objectives and specific instruction for the purpose of guiding instructional decision-making. Also documented in Figure 1, Hill's (2000) model of assessment practices identifies integrated systematic assessment as the assessment type that most closely adopts the improvement conception. Hill defines integrated systematic assessment as a process including systematically planned and collected data for the purpose of documenting progress and making instructional adjustments.

Gipps, et al. (1995)	Hill (2000)	Stamp (1987)	Heritage (2007)	Current assessment terms
Intuitive	Head note assessors	n/a	On-the-fly assessment	Observation
Evidence gatherers	Unit assessors	Traditional, academic summative examination	n/a	Summative
Systematic Planners	Integrated systematic assessors	Cater for need and progress for individual students	Planned-for interaction; curriculum-embedded assessments	Formative
n/a	n/a	Assessment blocks teachers' initiative	n/a	Irrelevant

Figure 1. Models of assessment practices. Rows demonstrate similarities among various researchers' assessment practice findings. Columns depict one researcher's work associated with the spectrum of assessment practices. Adapted from "What Makes a Good Primary School Teacher? Expert Classroom Strategies," by C. Gipps, M. Brown, B. McCallum, and S. McCalister, 1995, "Intuition or Evidence? Teachers and National Assessment of Seven-year Olds," Copyright 1995 by the Open University Press; "Formative Assessment: What do Teachers Need to Know and Do?" by M. Heritage, 2007, *Phi Delta Kappan*, 89(2), 140-145; "Remapping the Assessment Landscape: Primary Teachers Reconstructing Assessment in Self-monitoring Schools" by M. F. Hill, 2000, Unpublished Ph.D. Thesis, University of Waikato, NZ; "Evaluation of the Formation and Stability of Student Teacher Attitudes to Measurement and Evaluation Practices," by D. Stamp, 1987, Unpublished Doctoral Thesis, Macquarie University, Australia.

Heritage (2007) referenced three categories of formative assessment methods: "on-the-fly assessment, planned-for interaction, and curriculum –embedded assessment" (p. 141). On-the fly

assessment describes spontaneous assessment that occurs during the delivery of a lesson. On-the-fly assessment typically develops as the teacher observes student learning and determines the need for altered instruction before proceeding with previously planned activities. Planned-for interaction represents the implementation of previously planned and embedded assessment techniques, such as questioning, for the purposes of encouraging student exploration and eliciting informal assessment information. The third formative assessment category, curriculum-embedded assessments, can serve two functions: to solicit feedback at key points in a learning sequence and those that are an ongoing part of the classroom's activities.

A study by Delandshere and Jones (1999) reports distinct assessment practices associated with assessment aimed at continuous development. Assessment tasks related to this experiential perspective entail less curriculum prescribed responses reflective of classroom experiences. Delandshere and Jones report that this process necessitates teachers' desire to continuously appraise, rather than simply measure, the "quality and validity of the knowledge being demonstrated" (p. 219). Specifically emphasized by the researchers, is the need for the teachers' feedback to be rich with educative value to enable students to embrace greater responsibility for their learning and achievement.

Collectively, the aforementioned research indicates that teachers' who perceive assessment for improvement of teaching and student learning adopted a strategically planned formative assessment practice; however, it was the goal of this study to more clearly define the relationship between teachers' conceptions of assessment and the most valued assessment practices.

Defining summative assessment (assessment of learning).

Summative assessment, also known as assessment of learning (Black & Wiliam, 1998), is a means for documenting the nature and level of students' achievement at various times throughout their academic career (Hill, 2000). Within the summative assessment realm, researchers have identified three main purposes: to report student achievement and progress, to summarize achievement for the purpose of selection and qualification, and to offer data utilized for determining teacher, school, and system effectiveness (Brown, 2003; Hill, 2000; McNair, Bhargava, Adams, Edgerton, & Kypros, 2003). Typically used at the conclusion of an instructional unit or course, summative assessment provides the basics for comparisons between individuals, groups within a school, or between schools (Hill, 2000). This assessment methodology is one of monitoring learning for the purpose of certification and accountability (Hill, 2000; Brown, 2003).

Models of summative assessment.

Similar to Brown's accountability conceptions, Stamp (1987), Gipps et al. (1995), and Hill (2000) each document within Figure 1, assessment practices reflective of summative data used for the purposes of measuring, documenting, and reporting student, teacher, and school progress. Specifically, Stamp's description of the traditional-academic summative examination describes teachers' use of summative information to foster student competition for grades, possibly related to Brown's (2003) student accountability conception. Gipps et al. presented evidence gatherers' collection of evidence, typically obtained at the end of a unit, as a method for determining students' mastery of prescribed achievement objectives.

In 2000, Hill defined unit assessors (see Figure 1) as teachers who, "mainly described their assessment practices as occurring at the end of a unit of work and usually in terms of how well the children had met particular achievement objectives within that unit of work" (p. 225).

She indicated that the teachers within her study who adopted this assessment practice planned from the New Zealand curriculum documents and generally viewed assessment as separate from instruction. As a result, research results indicated that teachers used their collection of assessment evidence to measure achievement, write reports for parents, and group students by similar instructional needs.

Formative and summative assessment practices.

Teachers use a wide array of assessment tools within their classrooms, including standardized tests, district-developed assessments, textbook tests and quizzes, commercially developed tests and quizzes, and informal classroom assessment strategies (Adams & Hsu, 1998; McMillan & Nash, 2000; Trepanier-Street, McNair, & Donegan, 2001). A study conducted in 2001 examined the views and reported practices of lower (K-2 [N = 172]) and upper (3-5 [N = 126]) elementary teachers to determine teachers' use and value of multiple measures of assessment (Trepanier-Street et al., 2001).

After analyzing the 298 participants' responses, Trepanier-Street et al. (2001) concluded that in addition to being aware of a broad range of assessments, teachers also reported use of varied assessment techniques. Specifically, both lower and upper elementary teachers used and valued similar assessment measures; however, some differences and preferences were evident. Table 1 indicates lower elementary used and valued one-on-one assessments, written observational notes, and checklists, ratings scales, and portfolios, while upper elementary teachers placed greater emphasis on paper-pencil assessments, teacher-made tests, conferencing with students, and tests published from reading series and textbooks. Trepanier-Street et al. suggested that differences between the groups may be due to the developmental levels of the students they are teaching.

Table 1

Comparison of Lower- and Upper-Grade Teachers' Use of Assessment Approaches

Approach	Lower-grade		Upper-grade		Total		P
	f	%	f	%	f	%	
Observational notes	152	88.4	99	78.6	251	84.2	.025*
Review written work	167	97.1	125	99.2	292	98.0	.407
Baseline performance	144	83.7	95	75.4	239	80.2	.079
Discuss progress with child	124	72.1	112	88.9	236	79.2	.000**
Checklists/rating scales	122	70.9	72	57.1	194	65.1	.014*
Notes/reports to parents	155	90.1	122	96.8	277	93.0	.037*
Request parent view	103	59.9	80	63.5	183	61.4	.549
Textbook tests	74	43.0	102	81.0	176	59.1	.000**
Mandated standardized tests	47	27.3	73	57.9	120	40.3	.000**
Individual skill assessment	157	91.3	90	71.4	247	82.9	.000
Teacher-made tests	127	73.8	116	92.1	243	81.5	.000**

Note. f = frequency; P = probability of the relationship determined by Fisher Exact test. Adapted from "The View of Teachers on Assessment: A Comparison of Lower and Upper Elementary Teachers," by M. L. Trepanier-Street, S. McNair, and M. M. Donegan, 2001, *Journal of Research in Childhood Education*, 15(2), p. 237.

* $p < .05$.; ** $p < .001$.

McMillan et al. (2002) used a 6-point Likert scale to survey 901 third through fifth grade elementary teachers regarding their individual assessment and grading practices. Table 2 shows means and standard deviations of all items measuring assessment practices and indicates that rather than relying upon a singular form of assessment, third through fifth grade teachers embrace various tools and techniques to assess math and language arts. For example, the researchers noted objective assessments as the most frequently used assessment for both subject

areas (math mean of 3.43 and language arts, 3.43), with performance assessments (mean of 3.43) and projects (mean of 3.59) used almost as regularly as objective assessments in language arts. Mathematics responses included less reliance upon performance and project assessments (means of 2.84 and 2.51, respectively). Mathematics and language arts data indicated greater use of teacher-made (means of 3.63 and 3.90, respectively) and publisher supplied assessments (means of 3.54 and 3.22, respectively). The standard deviations (approximately 1 point on the scale) documented noteworthy variation within elementary teachers' assessment practices.

Table 2

Means and Standard Deviations of All Items Measuring Assessment Practices for Elementary Teachers

Types of Assessments	Mathematics		Lang. Arts	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Major examinations	3.21	1.39	3.05	1.38
Oral presentations	2.37	1.11	3.03	.88
Objective assessments (e.g. multiple choice, matching, short answer)	3.82	1.07	3.75	1.01
Performance assessments (e.g. structured teacher observations or ratings of performance, such as a speech or paper)	2.84	1.14	3.43	.93
Assessments provided by publishers or supplied to the teacher (e.g. in instructional guides or manuals)	3.54	1.05	3.22	1.06
Assessments designed primarily by yourself	3.63	.95	3.90	.98
Essay-type questions	2.42	1.15	3.39	1.03
Projects completed by teams of students	2.51	1.03	2.91	.99
Projects completed by individual students	3.06	1.24	3.59	.96
Performance on quizzes	3.93	.91	3.80	.98
Authentic assessments (e.g. real world performance tasks)	2.95	1.08	2.89	1.06

Note. $N = 901$. M = mean. SD = standard deviation. Adapted from "Elementary Teachers' Classroom Assessment and Grading Practices," by J. McMillan, S. Myran, and D. Workman, 2002, *The Journal of Educational Research*, (95)4, p. 207.

Although the McMillan et al., (2002) research was limited by teacher self-report, demographics, and location (Virginia initiating statewide assessment program consisting of all multiple choice tests, except writing), the large sample size provided strong external validity. The researchers concluded that few relationships existed between assessment practices and grade level, but that later grades did place a greater emphasis on “homework, extra credit, constructed-response assessments, objective assessments, and major examinations” (p. 212).

McNair et al. (2003) studied assessment practices of 157 elementary teachers from southeastern Michigan to determine use of varied assessment tools. As the second phase of a three phase study, the researchers used results from the 1997 statewide survey of Michigan teachers to determine their study’s focus. Because previous data indicated clear patterns of teachers’ assessment preferences but did not clearly identify what teachers actually did in the classroom, McNair et al. conducted follow-up interviews to document “the types, frequency, and utility of assessment techniques used by classroom teachers” (p. 24).

Researchers from five of the six universities involved in phase 1 used the 1997 statewide survey data to develop interview questions aimed at gaining greater insight regarding assessment tools. Data collected from primary teachers from various school districts representing a mix of urban and rural and high and low socioeconomic status were coded according to assessment strategy use, frequency of use, source of the assessment tool, and the purpose of the assessment data gathered from the use of a particular method. Data were divided into two groups, (66% of total sample) preschool through second grade (PreK-2) and (34% of total sample) third through fourth (3-4) grade teachers.

The McNair et al. (2003) study addressed results associated with paper-and-pencil tests, observations, checklists, and portfolios. Differences between pre-kindergarten and elementary

teachers, pre-kindergarten through second grade, and teachers at grade three and higher revealed that the frequency with which tests are used differs significantly by grade, specifically paper-and-pencil tests (see Table 3). Additionally, the results indicate a significant difference between the two groups for the source from which the tests are obtained (own, commercial, or both) and used (formative or summative). Data also revealed that the utility of paper-and-pencil tests does not differ by grade level since 92% in lower grades and 98% in upper grades relate their use of these tests to summative purposes.

Table 3

Assessment Practice Frequency

	PreK-2	3 rd -4 th
Paper-and-Pencil Tests	36%	92%
Observations	79%	91%
Checklists	47%	52%
Portfolios	95%	88%

Note. Adapted from “Teachers Speak Out on Assessment Practices,” by S. McNair, A. Bhargava, L. Adams, S. Edgerton, and B. Krypos, 2003, *Early Childhood Education Journal*, (31)1, pp. 25-27.

Results for checklists and portfolios (see Table 3) indicated no significant difference between frequency of use (McNair et al., 2003). Teachers in both grades frequently used checklists but indicated their preference for self-created tools. Additionally, results documented that despite checklists and portfolios traditional association with formative assessment, participants in the study used these tools primarily in a summative manner for the purpose of external accountability and reporting.

Despite observation’s essential role within a valid assessment system, the results of this study indicated that this assessment tool is primarily being used for a summative purpose rather than formative (McNair et al., 2003). Observation is used to gather information on students’

performance to support the ongoing differentiation of instruction. Although participants within the study indicated observation was a favored assessment strategy, the data revealed it was most often used to gather behavioral data rather than academic (73% of early level teachers and 76% of grade 3-4). Pearson's chi-square analyses yielded no significant differences between the two groups' frequency of use and utility of observations (see Table 3); however, a discrepancy between teacher comments and interview question responses revealed potential for greater identification with a formative assessment pedagogy, but a lack in understanding and implementation of assessment techniques that supported the "improvement conception" (Brown, 2003).

Similar to the McNair et al. (2003) study, Adams and Hsu (1998) explored 744 first through fourth grade mathematics teachers' conceptions of assessment and assessment practices and their relationship with grade level assignments. Despite a 36% return rate (269 surveys), the researcher deemed the sample representative of the research population. Results of Adams and Hsu's study indicated that teachers' conceptions of assessment encompass a wide array of assessment techniques and strategies. Specifically (see Table 4), on a 5-point Likert scale ranging from 5 = *Very important* to 1 = *Not important*, item means ranged from 2.65 for essays to 4.75 for teacher observations. The importance of observations was noted not only by the greatest mean but also the smallest standard deviation (0.48) and represents the teachers' strong agreement regarding the importance of this item. Additionally, "student performances, had the next highest mean, 4.70 (0.46) and the smallest standard deviation, also indicating strong agreement between teachers. The results suggested that teachers view their own actions and student actions as credible means for gathering assessment evidence.

Table 4.

Teachers' Conceptions of Assessment

Item	<i>M</i>	<i>SD</i>	<i>n</i>	χ^2
C1. Portfolios of students' work	3.895	1.181	267	17.366
C2. Interviews of students	3.641	1.078	265	8.799
C3. Student performances	4.704	0.457	267	1.179
C4. Student journals	3.340	1.210	267	13.870
C5. Essays	2.650	1.163	266	23.839
C6. Open-ended responses	3.784	0.958	265	27.679*
C7. Teacher observations	4.753	0.488	268	18.958
C8. Homework	3.403	1.174	268	33.928*
C9. Students' self-assessment	3.787	0.973	268	12.827
C10. Direct questioning	4.233	0.736	266	16.258
C11. Standardized test	3.037	1.244	268	14.727
C12. Teacher-made test	4.146	0.908	267	30.172*
C13. Student exhibitions	3.843	0.966	268	11.884
C14. Class discourse/discussion	4.220	0.749	268	16.418
C15. Students' disposition/attitudes	4.134	0.936	267	19.632
C16. Students' modeling of math	4.495	0.703	268	9.685
C17. Students' application of math	4.694	0.508	268	7.235
C18. Problem solving explorations	4.544	0.649	268	15.802
C19. Student calculator use	3.459	0.995	268	22.759
C20. Student computer use	3.916	0.949	263	15.854

Note. *n* = Number of cases in subsamples; *M* = Mean; *SD* = Standard deviation; χ^2 = Chi-square. Based on a 5-point Likert scale with 5 = *Very important* and 1 = *Not important*. *Table χ^2 = 26.296 in all cases except for C3, where the table χ^2 = 9.488. Adapted from Classroom Assessment: Teachers' Conceptions and Practices in Mathematics," by T. Adams and J. Hsu, 1998, *School Science and Mathematics*, 98(4), p. 176.

Standardized tests yielded the greatest variability among teacher responses (Adams & Hsu, 1998). With a mean score of 3.04 and a standard deviation of 1.24, the level of variation documented teachers' disparity in response: some assigned neutral, some assigned slight importance, and others disagreed. Within this study, the use of standardized tests to assess math

knowledge appeared to be representative of the ongoing debate in the education community. However, despite the debate related to the use of standardized tests to assess math knowledge, teachers generally rated assessment practices as neutral or important, which Adams and Hsu suggest represents teachers' agreement with the need for a variety of assessment techniques (McMillan et al., 2002).

When exploring the relationships between grade level and assessment conceptions and practices, Adams and Hsu (1998) used chi square analyses (see Table 4) to ascertain information pertaining to significance. Significant differences were noted for grade level and open-ended responses (27.68), homework (33.93), and teacher-made tests (30.17). Within this examination, the researchers documented more third and fourth grade teachers held homework as very important than did first and second grade teachers. However, more first and second grade teachers held very important conceptions for the use of "teacher-made tests as a means of assessment than did third and fourth grade teachers" (p. 179). Adams and Hsu concluded that these results support the assertion that teacher beliefs impact assessment practices, particularly by grade level (Stiggins & Bridgeford, 1985).

The existing research on assessment practices clearly documents numerous assessment methodologies identified as instrumental in increasing student achievement. While formative measures are represented by researchers as promoting the improvement of teaching and learning, summative instruments are viewed as more competitively structured to address accountability mandates for students, schools, and districts. Additionally, the large amount of assessment research prominently notes usage of various assessment techniques within the classroom; however, it is unclear how teachers' assessment beliefs relate to assessment practices level of importance. As a result, research indicated a need for this study to include survey items related to

assessment methods level of importance. The researcher used the teachers' results to determine how educators value various assessment techniques, and ultimately how the data related to assessment beliefs.

Assessment Professional Development

Despite the 1990 publication of the *Standards for Teacher Competence in Educational Assessment of Students*, calling for widespread staff development in the area of assessment, numerous researchers continue to document further evidence regarding the need for extensive training of all educators (Plake & Impara, 1993; Stiggins, 1991, 2002a; Zwick et al., 2008). A study conducted in the 1990's by the Joint Committee on Competency Standards in Student Assessment for Educational Administrators, surveyed over 1,700 administrators associated with three professional organizations. Participants were surveyed on 37 different assessment-related skills, of which three rated as most needed by educational administrators: knowing terminology associated with standardized tests, knowing the purposes of different kinds of testing, and understanding the connection between curriculum content and various tests (Impara, 1993). A couple of years later, the National Council on Measurement in Education (NCME) published the *Code of Professional Responsibilities in Educational Measurement*, requiring all professionals involved in any facet of educational assessment to "maintain and improve...professional competence in educational assessment" (NCME, 1995, p.1).

In spite of these national endeavors, Stiggins (2002a, 2002b) reports that only approximately twelve states require assessment competency for licensure attainment; however, no state licensing examination incorporates assessment skills for verification of competence. As

a result, higher education institutes housing teacher preparation programs have taken little note of the need to produce assessment literate teachers capable of engaging in assessment for learning (Stiggins, 2002a, 2002b). A recent report sponsored by the Wallace Foundation (Adams & Copland, 2005), succinctly documents skills required of administrators for state licensure. Adams and Copland (2005) noted that completely missing from the licensing framework was any mention of the meaning and use of assessments. In a 2003 study by the National Board on Educational Testing and Public Policy at the Lynch School of Education at Boston College, researchers analyzed 4,200 teacher survey responses to gain insight regarding the adequacy of professional development associated with standardized test interpretation. Almost a third of the respondents reported that professional development in this area was inadequate or very inadequate (Pedulla et al., 2003).

The evidence presented in Hill's (2000) educational case study involving twenty teachers within two primary schools in New Zealand documented that teachers understand assessment and the associated accountability obligations differently. Through Hill's transcription of interviews, analysis of observations, and reviewing of school records, the researcher was able to gather information pertaining to the teacher participants' assessment practices and beliefs.

Hill (2000) surmised that teachers frequently did not associate formative assessment practices with assessment, resulting in important implications for policy makers and professional developers. This lack of recognition by primary teachers may also be related to the McNair et al. (2003) study as results suggested "teachers may use appropriate assessment terminology and prefer more authentic classroom strategies, yet may lack the knowledge or skills crucial for assessing children systematically and meaningfully" (p. 30). Providers of professional development and teacher preparation programs need to elicit educators' ideas about assessment

and consider how these beliefs may impact their understanding of assessment in relation to teaching and learning.

Zwick et al. (2008) utilized the Instructional Tools in Educational Measurement and Statistics (ITEMS) survey to assess participants' understanding of educational measurement and statistics. At the conclusion of the field test and revised survey administration, researchers used results from both administrations to document substantial gaps in respondents' knowledge of educational measurement and statistics. The findings of Zwick et al. noted, "Only 10 of 24 UCSB respondents were able to choose the correct definition of measurement error, and only 10 new that a Z-score represents the distance from the mean in standard deviation units" (p. 15). ITEMS results provided the impetus for change, which Popham (2006) suggests will occur slowly and may hinge upon the inclusion of assessment competencies within state licensure requirements.

Black and Wiliam's (1998) research documents large student achievement gains on summative assessments, such as standardized tests, when partnered with well-crafted formative measures that are used diagnostically to adjust instruction and remediate students' weak skill areas. However, due to educators' minimal opportunities to acquire assessment literacy skills, available test data most frequently serve accountability purposes only (Zwick et al., 2008). As educational leaders conduct professional development opportunities associated with assessment, it is important to provide instruction on a wide range of techniques and tools in relation to teachers' grade levels (Adams & Hsu, 1998).

As research has documented (Adams & Hsu, 1998; Brown, 2003), teachers' distinct conceptions of assessment require knowledge of a spectrum of assessment tools to effectively assess student learning within the classroom. In general, studies have documented educators'

varying understanding and application of assessment practices, which has been linked to inadequate exposure to meaningful assessment professional development. It is the researcher's hope that results of this study will emphasize the critical need for the development of relevant professional development opportunities in the area of assessment as this information holds powerful implications related to student learning and achievement.

Summary of the Literature Review

This literature review provided a brief historical overview in relation to assessment within the last two decades and reviewed current literature about teachers' assessment beliefs and practices, particularly formative and summative assessment. The review highlighted national and international research and spotlighted investigations into the relationship between elementary teachers' conceptions of assessment and assessment practices as well as the influence of other mediating factors.

Chapter 3

Methodology

Introduction

This quantitative study seeks to gather practicing elementary teachers' current beliefs regarding assessment; the value assigned to specific classroom assessment practices; the relationship among demographic information (independent variables) and teachers' assessment conceptions and practices (dependent variables); and the relationship between elementary teachers' conceptions of assessment and their assessment practices. During the literature review, the original research questions were revised to facilitate the data collection and analysis. A survey will be administered to grade 3-5 educators to investigate the resulting research questions:

1. What are elementary teachers' conceptions (beliefs) about assessment?
2. What assessment practices are valued by 3rd through 5th grade elementary teachers?
3. What is the relationship between years of experience, grade level assignment, level of education, and assessment training and teachers' assessment beliefs and practices?
4. How do teachers' assessment beliefs relate to the value of assessment practices?

This chapter will review the design of the study, context of the research, population, instrumentation, data collection, data analysis, and the summary of the methodology.

Design

A quantitative design approach was used in an effort to describe the current perceptions of elementary teachers regarding conceptions about assessment and assessment practices to determine to what degree relationships exist among the variables. According to Gay and Airasian (2000) quantitative research is “based on the collection and analysis of numerical data” (p. 8) and is used to “describe current conditions, investigate relationships, and study cause-effect phenomena” (p.11). McMillan and Schumacher (2006) described essential elements of sound quantitative design as including subject selection, identification of data collection techniques, articulation of data gathering procedures, and procedures for treatment implementation and noted the importance of the researcher addressing “principles in each component that enhance the quality of the research” (p. 117).

This exploratory non-experimental study used a validated survey as the testing instrument. Mitchell and Jolley (2007) outlined three objectives that the researcher carefully planned for in order to conduct sound survey research. First, Mitchell and Jolley described the importance of the researcher having a clearly defined research hypothesis so that what is to be measured is evident. Second, they communicated the need for the selected instrument, in this study a survey, to accurately measure “the thoughts, feelings, or behaviors that you want to measure” (p. 208). Third, research results must be easily generalized to the identified population, which in this study is grade 3-5 elementary teachers.

The conceptual framework adopted in this study for selecting variables and organizing relationships among the variables was based on the previous studies of teachers’ conceptions of teaching, learning, and assessment and assessment practices utilized in the elementary classroom. It is intended that the survey data will provide a better understanding of teacher, school, and

district-based assessment practices and more adequately detail any existing relationships among the dependent and independent variables. Further, the information will aid in identifying teacher, school, and district-wide needs for professional development training, contribute to the development and use of more effective assessment practices, and ultimately yield improved student learning and teaching effectiveness.

Population and Sample

The target population in this study included third through fifth grade teachers working across two suburban and somewhat rural divisions in the Commonwealth of Virginia. The selected divisions had a combined third through fifth grade student census of 15,169 and 59 campus sites during the 2009-2010 academic year. These divisions were selected based on convenience sampling which McMillan and Schumacher (2006) noted is less costly and time consuming, provides for ease of administration, can provide a high participation rate, and makes generalization of results possible to similar subjects.

The participating counties collectively had 762 third through fifth grade teachers. One hundred twenty-four teachers comprised the sample population of which 84 responded to the survey. Five respondents' data were removed from the overall results due to partial survey completion, which resulted in a 64% response rate. Fifty-six respondents were from district A while 23 were employed by district B. Of the 74 female and 5 male participants 33 ranged in age from 43 and above, 16 were 34-42, 25 were 26-33, and 5 participants were 21 to 25. Twenty-five participants indicated they were teaching third grade; 31 were teaching fourth grade; and 22 teaching fifth grade. Of the participants, 10.1% indicated that they have less than 3 years of teaching experience; 36.7% have 4-10 years; 25.3% have 11-20 years; and 27.8% have greater than 20 years.

All respondents were asked to provide additional demographic information: level of education and type of completed assessment training. The level of education of the participants included 44.3% at the bachelor's level, 12.7% at the postgraduate certificate level, and 43% at the graduate level. Of the participants, 12.8% responded that they had no previous training in assessment. 68 respondents answered that they had received some level of training. Specifically, of the 87% who indicated participation in previous assessment training, 30.8% had taken an undergraduate course in assessment; 30.8% had taken a graduate course; and 63% had attended a workshop provided by their district or school or through an outside agency.

Instrumentation

The quantitative design of this study includes an online survey of participants. The survey was administered through Survey Monkey, an online survey 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 the service is available to the researcher at minimal cost. Another beneficial feature of Survey Monkey is the option to group respondents' answers in particular ways. For example, each school site serves as an individual collector enabling all participants' survey results to be sorted by school. Additionally, administering an online survey reduces the potential for interviewer and social desirability bias as well as provides participant anonymity (Mitchell and Jolley, 2007).

The survey consists of three sections: the first section includes demographic questions about the participants' background (gender, age, years of experience, grade level teaching assignment, level of education, and participation in assessment training); the second section is comprised of 27 Likert-type items scored on a scale from 1 to 5 (1 = *strongly disagree* and 5 = *strongly agree*) which address conceptions of assessment (assessment for learning or

improvement, assessment for student certification, assessment for school accountability, and assessment is irrelevant); and the third section is a set of 11 items regarding elementary assessment practices. The third section's Likert-type scale ranges from 1 to 5 with 1 equaling *not important* and 5 equaling *very important*.

After seeking permission from the author of the instrument, Gavin Brown's 2003 *Teachers' Conceptions of Assessment Abridged Survey* (COA-III) was adapted to serve the needs of this study. The original instrument included 50 items; however, for this study only 27 items will be used (see Appendix A). Additional scales related to conceptions of assessment were located, such as Adams and Hsu's (1998) 20 item survey on conceptions of assessment; however, no other scale dealt solely with the four main conceptions of assessment research findings: improvement of teaching and learning, assessment for student certification, assessment for school and division accountability, and assessment as irrelevant. Brown's COA-III Abridged items were designed to measure the structural relationships of the four main assessment conceptions and teachers' level of agreement or support for each conception.

When an instrument is partnered in conjunction with other batteries or requires a restricted response time, shorter surveys may prove more desirable (Brown, 2006). As a result, Brown (2006) investigated whether the abridged version of the COA-III provided results of similar quality. A confirmatory approach was adopted by Brown to determine whether this model measured the same conceptual framework in a substantial manner. He selected the three strongest statements related to each factor while being careful to avoid content redundancy. These identified statements were then reanalyzed using the data from the full battery. Brown recorded sufficient item loadings for the two jurisdictions' responses to first and second order factors and completed a confirmatory factor analysis to determine fit. Results indicated that the

intercorrelated Conceptions of Assessment-III Abridged noted “good fit characteristics ($\chi^2 = 841.02$; RMSEA = .057; TLI = .87)” (p. 169) and that the factors (school accountability, student accountability, assessment improves education, and assessment is irrelevant) “had very similar direction and values” (p.169) as the full scale reported by Brown in 2004.

Brown (2006) used an independent confirmatory study with two jurisdictions, Queensland and New Zealand. Results for the 692 primary only teachers had acceptable fit ($\chi^2 = 1492.61$; $P < .001$; RMSEA = .074; TLI = .80) and sufficient loadings on each factor. Despite these interfactor correlations differing from New Zealand’s primary results, the direction remained similar. Brown surmised that the differences in factor correlations were related to how the two jurisdictions’ primary teachers view the relationship among the four main assessment purposes.

Regardless of the variance within the two jurisdictions’ interfactor correlations, Brown (2006) demonstrated that the COA-III Abridged instrument provided valid factor scale scores. Therefore, the shortened inventory was deemed an efficient and valid measure of teachers’ conceptions of assessment and was selected as a measure for this study.

The assessment practices portion of the instrument consists of 11 items which were drawn from the McMillan et al. (2002) 34 item questionnaire designed to explore factors considered by teachers when grading, such as student effort, improvement, academic performance, types of assessments used, and the cognitive level of assessments. A six-point scale ranging from *not at all* to *completely* was used by McMillan et al. to enable teachers to document usage without the restrictions associated with a commonly used ipsative scale. After gaining permission to edit the instrument from the author of the scale, the researcher limited the inclusion

of survey items in this study to those relevant to types of assessments used by teachers. The original scale was revised to include a five-point scale ranging from *not important* to *very important* to assist participants with documenting levels of importance versus the original scale's goal of reporting results associated with assessment usage. The resulting 11 items related to assessment practices can be seen in Appendix A.

McMillan et al. (2002) constructed the original 47 item scale from previous questionnaires noted in the literature, as well as research summarizing teachers' grading and assessment practices. To strengthen the content-related evidence for reliability, the researchers conducted a pilot study consisting of 15 teachers. Participants were asked to review the 47 items "for clarity and completeness in covering most, if not all, assessment and grading practices used" (p. 206). After completing item revisions, twenty-three teachers from outside of the study's sample population were secured for a second pilot test. Participants were charged with reviewing the items for "feedback on clarity, relationships among items, item-response distributions, and reliability" (p. 206). Item statistics documenting weak reliability and items with minimal variation or correlations greater than .90 ($r > .90$) were eliminated, resulting in 27 remaining items.

Approximately 4 weeks after the completion of the second pilot test, the same twenty-three teachers were asked to retake the questionnaire (McMillan et al., 2002). Reliability was determined by the researchers' use of stability estimates to review the percentage of matches for the items. Items documenting exact matches of 60% or less were removed or combined with other items. Results confirmed that an average of 46% of participants' responses to items had an exact match, while "89% of the matches were within 1 point on the 6 point scale" (p. 206). The revised questionnaire consisted of 34 items clustered into three categories: items assessing

different factors used to determine grades (19), items assessing different types of assessments used (11), and items assessing the cognitive level of the assessments (4).

Data Collection

Before contacting the school division regarding participation in this study, the researcher submitted the required materials to the Institutional Review Board (IRB) at Virginia Commonwealth University. The materials included the completed protocol for the research project and the survey materials. Upon receipt of IRB approval, the Director of Research for each school division represented by the 59 schools was sent a letter requesting permission to conduct this study within all elementary schools. A copy of the survey (Appendix A), principal letter (Appendix B), and teacher letter (Appendix C) were provided to the Directors.

Upon receiving permission from the school divisions' representatives, a list of elementary principals was obtained through one county's research and technology department. District B required the researcher to send all documents to her via email. She in turn would act as a liaison of information between the researcher and the principals. An initial email was sent in late May, inviting each administrator to preview the survey to determine participation of third through fifth grade teachers. The purpose of the study, importance of voluntary participation, and confidentiality assurance was included in this correspondence. This email solicitation is shown in Appendix B, and the online survey is listed in Appendix A. A second email was sent one week later to administrators, encouraging all principals of participating buildings to forward the survey to the target population. This second email contained the letter of participation to teachers (Appendix C) which included a live link to the validated survey. Survey responses were then collected for a two week period for each district.

It should be noted that prior to conducting a mass distribution of the survey, the researcher piloted the instrument on two occasions with five colleagues in order to elicit commentary and feedback. At the conclusion of the pilot tests, the researcher made minor corrections to word choice and proceeded with plans for mass distribution of the revised survey.

Data Analysis

The participants' responses to the survey were entered into the statistical software program, PASW, upon which descriptive measures were compiled and between group tests completed. Specifically, research questions one and two (see Table 5) were analyzed using descriptive statistics such as frequencies, means, standard deviations, and percents. Data were calculated for each subscale related to teachers' conceptions of assessment and all items for assessment practices.

Table 5.

Research Questions and Data Analyses

Research Question	Statistics	Data Analysis
1. What are elementary teachers' conceptions (beliefs) of assessment?	Descriptive	Means, Standard Deviations, Frequencies, and Percents
2. What assessment practices are valued by 3rd through 5th grade teachers?	Descriptive	Means, Standard Deviations, Frequencies, and Percents
3. What is the relationship between years of experience, grade level, level of education, and assessment training and teacher beliefs and practices?	Inferential	Analysis of Variance (ANOVA); <i>t</i> tests; Post hoc (if needed)
4. How do teachers' assessment beliefs relate to the value of assessment practices?	Inferential	Correlation; Scatter Plots (if needed)

To gather data related to question three (see Table 5), the researcher conducted an Analysis of Variance (ANOVA) to test for differences among the demographic information (independent variables), specifically grade level assignment, years of experience, level of

education, and participation in assessment training and teachers' assessment beliefs and practices. Mitchell and Jolley (2007) noted that ANOVA's are especially useful when a study has "more than one independent variable or more than two levels of an independent variable" (p. 589). When ANOVA results yielded a significant F statistic, the researcher completed a follow-up test to determine specifically which group(s) differed. Post hoc t tests enabled the researcher to reduce the impact of *Type I* error and determine which means differ from one another.

The fourth research question investigated how teachers' assessment conceptions or beliefs related to assessment practices. Inferential statistics (see Table 5) were conducted to examine the correlation between the two variables, assessment conceptions and practices. Mitchell and Jolley (2007) recommended the use of a Pearson Correlation test to more closely analyze the correlation among our variables. Results from this statistical analysis were used by the researcher to determine whether a positive, zero, or negative correlation existed between teachers' assessment conceptions and practices.

Summary of the Methodology

To summarize the methodology for this study, chapter three explained the researcher's use of the non-experimental quantitative design approach. The study focused on the relationships between independent variables such as years of experience, level of education, grade level assignment, and participation in assessment training and 3-5 elementary teachers' assessment conceptions and practices. The independent and dependent variables were assessed from late May through June using constructs from the survey's three sections: demographics, assessment conceptions, and assessment practices. The data were analyzed through descriptive and inferential statistical analyses using PASW software, and the final two chapters spotlighted the study's results, discussion, and implications.

The following chapter presents the results obtained from the data analysis.

Chapter 4

Results

Overview

The purpose of this study was to elicit self-ratings from third through fifth grade elementary teachers regarding their assessment beliefs and importance of practices. Specifically explored were teachers' perceptions related to the four main purposes of assessment: assessment makes schools accountable, assessment makes students accountable (student certification), assessment improves instruction and learning, and assessment is irrelevant. The abridged, 27-item *Conceptions of Assessment Inventory* (CoA-III A) from Brown (2003) was used in this research by teachers to indicate their level of agreement using a Likert scale ranging from 1 to 5 (1 = *strongly disagree* and 5 = *strongly agree*). The study also utilized a revised five-point, assessment practices scale (1 = *not important* and 2 = *very important*) from McMillan et al. (2002) to assist participants with documenting assessment practices' level of importance. Additionally, collected demographic information enabled the researcher to consider the variables of gender, age range, level of education (highest degree), years of experience, grade level assignment, and level of previous assessment training.

Following are results from demographic information and survey responses which are presented within the framework of the following research questions.

1. What are elementary teachers' beliefs (conceptions) about assessment?

2. What classroom assessment practices are valued by 3rd -5th grade elementary teachers?
3. What is the relationship between years of experience, grade level assignment, level of education, and assessment training and teachers' assessment beliefs and practices?
4. How do teachers' assessment beliefs relate to the value of assessment practices?

Rate of return of surveys. The rate of return of the surveys from teachers varied between the two districts, with an overall return rate of 63%, as seen in Table 6. Of the 79 respondents, 72.2% stemmed from District A while 27.8% of surveys were completed by District B. To encourage participation in the study, an introductory letter with an embedded survey link was forwarded by building principals to eligible participants. Approximately one week later, a reminder email was sent to principals and forwarded to teachers reminding them of the pending closing survey window (1 remaining week). The 63% overall rate of return aligned nicely with the preferred rate of return of 50%-60% noted by several researchers (Diem, 2003; Rudestam & Newton, 2007).

Table 6

Rate of Return of Teacher Surveys

District	Completed Surveys	Percent Completion
District A	57	72.2%
District B	22	27.8%

Missing data. Some of the eighty-four originally submitted online surveys contained missing data which resulted in its exclusion from overall results. Bryman and Cramer (1997) recommended more than 10 percent of missing data as a criterion which can be applied to what represents too much missing data. In this study, five participants had greater than 10 percent of their responses missing and therefore all related data was omitted. The researcher determined the need to utilize Valid Percent columns when analyzing remaining results to account for any remaining participants' data sets which had less than 10 percent of missing values (Rudestam & Newton, 2007).

The remainder of this chapter is comprised of three major sections which include a presentation of the demographic descriptive statistics, the descriptive and inferential data analyses for each of the four research questions, and an overall summary of the research findings.

Descriptive Data for Demographic Information

This section of the chapter reflects the demographic information provided by the study's 79 participants. The data are presented in tabular form for the following demographic characteristics: school district, gender, age range, level of education, years of experience, grade level assignment, and types of training in educational assessment.

The participating districts' target population collectively had 762 third through fifth grade teachers. With an overall response rate of 63%, 84 out of 124 (sample population) third through fifth grade teachers participated in the study, five of which were removed due to partial survey completion. Descriptive statistics were used to determine 57 respondents worked in district A while 22 were employed by district B (see Table 7). Of the 74 female and 5 male participants (see Table 8) 33 ranged in age from 43 and above, 16 were 34-42, 25 were 26-33, and 5

participants were 21 to 25 (see Table 9). One respondent did not provide information pertaining to his/her grade level teaching assignment. As a result, 78 responses yielded the following results: 24 participants teaching third grade; 31 teaching fourth grade; and 23 teaching fifth grade (see Table 10).

Table 7

Descriptive Statistics for School Districts

Variable	<i>n</i>	%
District A	57	72.2
District B	22	27.8
Total	79	100.0

Table 8

Descriptive Statistics for Gender

Variable	<i>n</i>	%
Female	74	93.7
Male	5	6.3
Total	79	100.0

Table 9

Descriptive Statistics for Age Range

Variable	<i>n</i>	%
21-25	5	6.3
26-33	25	31.6
34-42	16	20.3
43 and above	33	41.8
Total	79	100.0

Table 10

Descriptive Statistics for Teaching Assignment (Grade Level)

Variable	<i>n</i>	%
3 rd Grade	24	30.8
4 th Grade	31	39.7
5 th Grade	23	29.5
Total	78	100.0

Table 10 documents the number of educators per grade level. Of the participants, 10.1% indicated that they have less than 3 years of teaching experience; 36.7% have 4-10 years; 25.3% have 11-20 years; and 27.8% have greater than 20 years (see Table 11). The level of education of the participants included 44.3% at the Bachelor's level, 12.7% at the postgraduate certificate level, and 43% at the graduate level (see Table 12).

Table 11

Descriptive Statistics for Years of Experience

Variable	<i>n</i>	%
< 3 years	8	10.1
Between 4-10	29	36.7
Between 11-20	20	25.3
More than 20	22	27.8
Total	79	100.0

Table 13 outlines the sample population's level of education. 35 participants (44.3%) have attained a Bachelor's degree. 10 teachers (12.7%) have earned a postgraduate certificate, while 34 (43%) have earned a Master's degree. None of the respondents had earned a Doctorate. Of the 79 participants, 12.7% responded that they had no previous training in assessment. 69

respondents answered that they had received some level of training. Specifically, of the 87.3% who indicated participation in previous assessment training, 30.4% had taken an undergraduate course in assessment; 30.4% had taken a graduate course; and collectively, 62% had attended a workshop provided by their district or school or through an outside agency (see Table 12).

Table 12

Descriptive Statistics for Assessment Training

Variable	<i>n</i>	%
None	10	12.7
Undergraduate course	24	30.4
½ to 1 day workshop provided by current or previous employer	42	53.2
½ to 1 day workshop provided by outside agency	7	8.9
Graduate course	24	30.4
Other	0	0
Total		100.0

Table 13

Descriptive Statistics for Degree Attainment

Variable	<i>n</i>	%
Bachelor's	35	44.3
Postgraduate certificate	10	12.7
Master's	34	43.0
Doctorate	0	0.0
Total	79	100.0

Question 1: What are elementary teachers' beliefs (conceptions) about assessment?

In response to the first research question, “What are elementary teachers’ (3rd-5th) assessment beliefs?” the researcher used descriptive statistics to determine the means, standard deviations, frequencies, and percents of the four main assessment beliefs: assessment for school accountability, assessment for student certification/accountability, assessment is irrelevant, and assessment for improvement. Due to the *COAIII* (Brown, 2007) instrument consisting of 27 items, the researcher determined the need to create subgroups for the purpose of analysis. Prior to running the descriptive statistics, the researcher clustered instrument items by Brown’s (2007) previously identified assessment subgroups. Table 14 documents how the 27 survey items were clustered in Brown’s previous study and this study. These new variables were used when determining the descriptive statistics associated with respondents’ assessment beliefs.

Table 14

COAIII Survey Item Sub-Categories

Sub-Categories	Survey Items
Irrelevance	Interferes with teaching Unfair to students Forces against beliefs Filed and ignored Little use of results Little impact on teaching Imprecise process Measurement error Account error and imprecision
Accountability of Students	Assign grade/level to work Meet qualification standards Place students into categories
Accountability of Schools	Good way to evaluate school How well schools are doing Accurate indicator of school quality
Improvement	Dependable Consistent Trustworthy What learned Higher order thinking How much learned Modifies ongoing teaching Integrated with teaching Allows different instruction Feedback about performance Informs of learning needs Helps improve

Note. Adapted from “Conceptions of Assessment-III” by Brown, G. T. L. (2007, December). *Teachers’ conceptions of assessment: Comparing measurement models for primary and secondary teachers in New Zealand*. Paper presented at the New Zealand Association for Research in Education, Christchurch, NZ.

Results reported in Table 15 reveal calculations of the frequency, mean, and standard deviation of the four variables associated with assessment beliefs. The mean standard scores ranged from 3.43 to 4.25 suggesting that average levels of assessment beliefs revealed some variability. The assessment for improvement mean ($M = 4.18$) yielded the highest result while

assessment as irrelevant ($M = 3.43$) reflected the lowest average score. Each standard deviation indicated the average variability of the scores from the mean within a normal distribution. School accountability ($SD = 1.07$) had the greatest level of variance as approximately 68% of responses fell within one standard deviation of the mean. The three remaining subgroups, improvement ($SD = .58$), student accountability ($SD = .77$), and irrelevant ($SD = .71$) revealed minimal variation in comparison to school accountability.

Table 15

Descriptive Statistics for Beliefs Subgroups

Variable	<i>n</i>	<i>M</i>	<i>SD</i>
Improvement	74	4.25	.58
Student Accountability	76	4.18	1.07
Irrelevant	75	3.43	.71
School Accountability	78	3.68	.77

To determine if belief subgroups related to one another, the researcher conducted a Pearson Correlation analysis to identify any levels of significance (see Table 16). Results revealed school accountability as having a moderately significant association ($r = .58$) with the improvement assessment belief. A moderate correlation was also noted between school accountability and student accountability ($r = .55$). Additionally, the irrelevant assessment belief was found to have a mild, negative correlation ($r = -.307$) with the improvement belief.

Table 16

Correlations of Assessment Belief Subgroups

Assessment Beliefs		Student Accountability	Irrelevant	School Accountability	Improvement
Student Accountability	<i>r</i>	1	.08	.55**	.21
	Sig. (2-tailed)		.51	.000	.08
	<i>n</i>	76	73	75	71
Irrelevant	<i>r</i>	.08	1	-.14	-.31**
	Sig. (2-tailed)	.51		.23	.01
	<i>n</i>	73	75	74	70
School Accountability	<i>r</i>	.55**	-.14	1	.58**
	Sig. (2-tailed)	.00	.23		.00
	<i>n</i>	75	74	78	73
Improvement	<i>r</i>	.21	-.31**	.58**	1
	Sig. (2-tailed)	.08	.01	.00	
	<i>n</i>	71	70	73	74

Note. **. Correlation is significant at the 0.01 level (2-tailed).

Question 2: What assessment practices are valued by grade 3-5 elementary teachers?

The descriptive analyses of the second research question, “What assessment practices are valued by 3rd through 5th grade teachers?” is noted in Tables 17 and 18. Table 17 documents the frequency and percentage of participants’ responses to assessment value by survey item. When looking at the percentages associated with value of assessment practices, teacher ratings indicate that approximately 51% of the study’s participants feel authentic assessments are “*Very Important*”, while publisher assessments (11.5%) and major exams (6.1%) were viewed as “*Not Important*” by participants. Surprisingly, teachers identified all of the following assessment types as having some level of value within the classroom: assessments designed by self, performance quizzes, objective assessments such as multiple choice and matching, short answer assessments,

performance assessments, authentic assessments, and oral presentations. Not one of the aforementioned assessment types received a participant rating of “1 – *Not Important*”.

When collectively reviewing percentages associated with “*Not Important*” and “*Slightly Important*”, 41% of participants reported that publisher assessments had little value within the 3rd through 5th grade classroom. A joint review of assessments rated as “*Quite Important*” and “*Very Important*” showed 81.3 percent of participants placed significant value on authentic assessments such as “real world” performance tasks. Additionally, while approximately one-quarter of teachers responded that projects in teams (26%) and major exams (24.3%) had little value as an assessment type, approximately three-fourths of the study’s respondents identified short answer (74.4%) and performance assessments (76.9%) such as structured teacher observations or ratings of a performance such as a speech or paper as highly valuable.

Table 17

Descriptive Statistics for Frequency and Percent for Value of Assessment Practices

Variable	Not Important	Slightly Important	Fairly Important	Quite Important	Very Important	Total
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Designed by self	0 (0)	7 (8.9)	17 (21.5)	37 (46.8)	18 (22.8)	79 (100)
Performance quizzes	0 (0)	5 (6.5)	27 (35.1)	37 (48.1)	8 (10.4)	79 (100)
Objective assessments	0 (0)	10 (12.7)	33 (41.8)	31 (39.2)	5 (6.3)	79 (100)
Short answer	0 (0)	4 (5.1)	16 (20.5)	45 (57.7)	13 (16.7)	78 (100)
Performance assessment	0 (0)	1 (1.3)	17 (21.8)	40 (51.3)	20 (25.6)	78 (100)
Projects by self	1 (1.3)	5 (6.5)	22 (28.6)	38 (49.4)	11 (14.3)	79 (100)
Major exams	5 (6.4)	14 (17.9)	27 (34.6)	29 (37.2)	3 (3.8)	78 (100)
Authentic assessments	0 (0)	2 (2.6)	11 (14.1)	25 (32.0)	40 (51.3)	78 (100)
Projects in teams	3 (3.9)	17 (22.1)	24 (31.2)	25 (32.5)	8 (10.4)	77 (100)
Publisher assessments	9 (11.5)	23 (29.5)	30 (38.5)	15 (19.2)	1 (1.3)	78 (100)
Oral presentations	0 (0)	8 (10.3)	28 (35.9)	33 (42.3)	9 (11.5)	78 (100)

Note. Scale ranges from 1 (*Not Important*) to 5 (*Very Important*); Adapted from “Assessment Practices Instrument” by McMillan, J., Myran, S., & Workman, D. (2002). Elementary teachers’ classroom assessment and grading practices. *The Journal of Educational Research*, (95)4, 203-213.

Table 18 shows the means with respect to how third through fifth grade teachers value assessment practices. Teachers reported that publisher assessments yielded the lowest assessment value mean ($M = 2.69$) while performance assessments ($M = 4.01$) and authentic assessment ($M = 4.32$) means were the highest. Assessments designed by the teachers and short answer assessments revealed a similar level of high importance with approximate means of 3.8 for both types.

Table 18

Descriptive Statistics for Value of Assessment Practices by Mean

Variable	<i>n</i>	<i>M</i>	<i>SD</i>
Designed by self	79	3.84	0.88
Performance quizzes	77	3.62	0.76
Objective assessments	79	3.39	0.79
Short answer	78	3.86	0.75
Performance assessments	78	4.01	0.73
Projects by self	77	3.69	0.85
Major exams	78	3.14	0.98
Authentic assessments	78	4.32	0.81
Projects in teams	77	3.23	1.04
Publisher assessments	78	2.69	0.96
Oral presentations	78	3.55	0.83

Note. Adapted from “Assessment Practices Instrument” by McMillan, J., Myran, S., & Workman, D. (2002). Elementary teachers’ classroom assessment and grading practices. *The Journal of Educational Research*, (95)4, 203-213. Means range from 1 (*Not Important*) to 5 (*Very Important*).

Question 3: What is the relationship between years of experience, grade level assignment, level of education, and assessment training and teachers’ assessment beliefs and practices?

Composite scores for assessment beliefs were disaggregated according to each independent variable: years of experience, grade level assignment, level of education, and assessment training. Descriptive analyses were completed in order to conduct a mean comparison among the independent variables (years of experience, grade level assignment, type of assessment training, and level of education) and assessment beliefs and practices. The means were compared for each level of independent variable to determine if there was significant

variation between teachers' ratings of assessment beliefs and importance of practices and the varying demographic characteristics.

Years of experience. Mean composite scores for each assessment belief subgroup were compared for the four different levels of the independent variable, years of teaching experience. The four levels of this variable were: 0-3 years of experience, 4 to 10 years of experience, 11 to 20 years of experience, and greater than 20 years of experience. Table 19 summarizes the mean scores for each category of years of experience by the belief subgroups: student accountability, irrelevant, school accountability, and improvement. The data indicated a general trend for teachers with the least amount experience. As shown, teachers with 0 to 3 years of experience have the lowest mean in three out of the four belief subgroups. In comparison to their less-experienced colleagues, teachers with 4 to 10 years of experience had the highest means for school accountability and assessment for improvement. Standard deviations for each subgroup indicated that the most variability in responses was associated with school accountability, while the least variability in responses was related to the improvement belief.

Table 19

Comparison of Assessment Belief Mean Scores for Years of Teaching Experience

Variables	0-3 years			4-10 years			11-20 years			>20 years		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Student Accountability	8	3.92	.79	29	4.06	.75	20	4.28	.77	21	4.27	.80
Irrelevant	8	3.04	.63	29	3.42	.71	20	3.41	.76	19	3.64	.66
School Accountability	8	3.46	1.15	30	3.81	.76	19	3.67	1.11	22	3.53	1.40
Improvement	8	4.17	.72	29	4.32	.49	18	4.27	.53	20	4.10	.68

Note. Means range from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*).

To determine if there were any significant differences among the levels of teaching experience, an ANOVA was conducted. The ANOVA results did not reveal any significant difference according to years of teaching experience for assessment beliefs. As a result a *post hoc* analysis was not needed to identify the specific differences among the four levels of the independent variable.

Mean composite scores for each assessment practice were compared for the four different levels of the independent variable, years of teaching experience. The four levels of this variable were: 0-3 years of experience, 4 to 10 years of experience, 11 to 20 years of experience, and greater than 20 years of experience. Table 20 summarizes the mean scores for assessment practice by each level of years of experience. The data indicated a general trend for teachers with the least amount experience. As shown, teachers with 0 to 3 years of experience have the highest mean for all assessment practices with the exceptions of major exams, authentic assessments, and publisher assessments. Collectively, publisher assessments scored the lowest average among each of the four age ranges. An additional trend that can be seen in Table 20 is the decline in means as the years of experience increase. For example, when looking at performance assessments, projects by self, major exams, authentic assessments, projects and teams, publisher assessments, and oral presentations the highest means can typically be associated with the least experienced teacher population. As years of experience increases, the averages tend to decrease.

Table 20

Comparison of Assessment Practice Means by Years of Teaching Experience

Variable	0-3			4-10			11-20			>20		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Designed by self	8	4.00	.93	29	3.76	.95	20	3.70	.92	22	4.00	.76
Performance quizzes	8	3.88	.64	28	3.46	.69	20	3.65	.81	21	3.71	.85
Objective assessments	8	3.63	.52	29	3.38	.73	20	3.10	.64	22	3.59	1.01
Short answer	8	4.13	.35	28	4.00	.72	20	3.70	.57	22	3.73	.99
Performance assessment	7	4.43	.54	29	4.10	.72	20	4.00	.80	22	3.77	.69
Projects by self	8	4.13	.64	29	3.97	.87	19	3.47	.96	21	3.33	.58
Major exams	8	3.25	1.17	29	3.38	.86	20	3.10	.91	21	2.81	1.08
Authentic assessments	8	4.25	.71	29	4.62	.49	20	4.25	.91	21	4.00	1.00
Projects in teams	8	4.13	.64	28	3.61	.96	220	2.90	1.12	21	2.71	.78
Publisher assessments	8	2.88	.99	29	2.62	1.05	20	2.95	.89	21	2.48	.87
Oral presentations	8	3.88	.64	29	3.76	.74	20	3.35	.93	21	3.33	.86

Note. Means range from 1 (*Not Important*) to 5 (*Very Important*).

Table 21 shows results from an ANOVA of assessment practices for years of experience. Results showed significant differences in select assessment practices by years of experience. Specifically, significant differences were noted for projects by self, authentic assessments, and projects by teams. A Bonferroni *post hoc* analysis showed a significant mean difference for projects by self between teachers with 0-3 and greater than twenty years of experience. A significant mean difference for authentic assessments between teachers with 11-20 years of

experience and those with greater than 20 years was identified. Additionally, two significant mean differences were documented for projects by teams for teachers with 0-3 years of experience and the two independent variable levels of 11-20 years and greater than 20 years of experience. A final significance for projects by teams was noted for teachers with 4-10 years of experience and those with greater than 20 years.

Table 21

ANOVA of Assessment Practices for Years of Teaching Experience

Practices		<i>Df</i>	<i>F</i>	<i>p</i>
Designed by self	Between Groups	3	.57	.64
	Within Groups	75		
	Total	78		
Performance quizzes	Between Groups	3	.80	.50
	Within Groups	73		
	Total	76		
Objective assessments	Between Groups	3	1.65	.186
	Within Groups	75		
	Total	78		
Short Answer	Between Groups	3	1.20	.37
	Within Groups	74		
	Total	77		
Performance assessments	Between Groups	3	1.75	.16
	Within Groups	74		
	Total	77		
Projects by self	Between Groups	3	3.75	.02*
	Within Groups	73		
	Total	76		
Major exams	Between Groups	3	1.45	.24
	Within Groups	74		
	Total	77		
Authentic assessments	Between Groups	3	2.63	.05*
	Within Groups	74		
	Total	77		
Projects by team	Between Groups	3	6.94	.00*
	Within Groups	73		
	Total	76		
Publisher assessments	Between Groups	3	.99	.40
	Within Groups	74		
	Total	77		
Oral presentations	Between Groups	3	1.94	.13
	Within Groups	74		
	Total	77		

Note. * $p < .05$

Grade level assignment. The means for assessment beliefs as measured by the *COAIII* (Brown, 2003) showed slight variation according to the independent variable, grade level assignment. Table 22 summarizes the mean scores by grade level.

Table 22

Comparison of Classroom Assessment Belief Mean Scores for Grade Level

Variables	3 rd			4 th			5 th		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Student Accountability	23	4.39	.65	30	4.07	.89	24	4.04	.71
Irrelevant	23	3.42	.68	29	3.41	.71	23	3.42	.74
School Accountability	25	3.55	1.28	30	3.71	.93	23	3.7	1.06
Improvement	21	4.09	.59	30	4.25	.49	23	4.33	.68

Note. Means range from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*).

Multiple Analysis of Variance (ANOVA) analyses were conducted to test for significant differences in mean assessment belief scores according to the three levels of the independent variable (third, fourth, and fifth grades). ANOVA results indicated no statistically significant differences between grade levels and assessment beliefs.

Teachers' reported importance of assessment practices was examined to determine if value varied according to the three independent variable levels for grade level assignment: third, fourth, and fifth. Table 23 shows a comparison of means for assessment practice by grade level assignment. Standard composite scores for third grade ranged from 2.79 to 4.38. Fourth grade composite scores for assessment practices ranged from 2.73 to 4.35, and fifth grade means ranged from 2.57 to 4.23. When comparing the ranges for all levels of the independent variable, it suggests that average levels for assessment practices were relatively similar; however, the standard deviation within each level suggests greater variability within groups, especially for

major exams (3rd grade), projects in teams, and publisher assessments. This variability indicates that despite mean scores, participants maintain a wide perspective on assessment practice importance. For each remaining assessment practice by grade level, the smaller standard deviations indicate the clear majority of respondents scored near the mean which resulted in a more even distribution of scores. This more even division suggests a less discrepant perspective on importance of assessment types.

Table 23

Assessment Practices Means by Grade Level Assignment

Variable	3 rd			4 th			5 th		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Designed by self	24	3.83	.87	31	3.71	.86	23	4.00	.95
Performance quizzes	24	3.75	.68	30	3.47	.82	22	3.73	.77
Objective assessments	24	3.75	.78	31	3.23	.72	23	3.48	.90
Short answer	23	4.00	.71	31	3.81	.79	23	3.83	.78
Performance assessment	23	4.17	.83	31	4.03	.66	23	3.83	.72
Projects by self	23	3.83	.78	30	3.70	.99	23	3.52	.73
Major exams	24	3.25	1.15	30	3.20	.93	23	3.00	.85
Authentic assessments	24	4.38	.88	31	4.35	.76	22	4.23	.87
Projects in teams	24	3.17	1.05	30	3.63	.93	22	2.82	1.01
Publisher assessments	24	2.79	1.10	30	2.73	.91	23	2.57	.90
Oral presentations	23	3.65	.78	31	3.71	.74	23	3.26	.96

Note. Means range from 1 (*Not Important*) to 5 (*Very Important*).

An ANOVA was conducted to determine if there were significant differences in mean assessment practice scores according to grade level. As shown in Table 24, the results of the

ANOVA indicated a significant difference among projects completed by teams and grade level assignment. A Bonferonni *post hoc* (see Table 25) analysis indicated the mean score for projects completed by teams was significantly different between 4th and 5th grade teachers (mean difference = .82). 4th grade teachers average composite mean for projects completed in teams was 3.17 compared to 2.82 for 5th grade teachers.

Table 24

ANOVA for Assessment Practices by Grade Level

Practices		<i>Df</i>	<i>F</i>	<i>p</i>
Designed by self	Between Groups	2	.70	.50
	Within Groups	75		
	Total	77		
Performance quizzes	Between Groups	2	1.20	.32
	Within Groups	73		
	Total	75		
Objective assessments	Between Groups	2	1.04	.36
	Within Groups	75		
	Total	77		
Short Answer	Between Groups	2	.28	.76
	Within Groups	74		
	Total	76		
Performance assessments	Between Groups	2	1.32	.27
	Within Groups	74		
	Total	76		
Projects by self	Between Groups	3	.74	.48
	Within Groups	73		
	Total	75		
Major exams	Between Groups	2	.43	.65
	Within Groups	74		
	Total	76		
Authentic assessments	Between Groups	2	.22	.81
	Within Groups	74		
	Total	76		
Projects by team	Between Groups	2	4.43	.02*
	Within Groups	73		
	Total	75		
Publisher assessments	Between Groups	2	.35	.71
	Within Groups	74		
	Total	76		
Oral presentations	Between Groups	2	2.18	.12
	Within Groups	74		
	Total	76		

Note. * $p < .05$

Table 25

Bonferonni Post Hoc for Assessment Practice (Projects by Team) and Grade Levels

Assessment Practice	(I) What grade level do you teach?	(J) What grade level do you teach?	Mean Difference (I-J)	Std. Error	Sig.	95% Conf. Int.	
						Lower Bound	Upper Bound
Projects by team	3 rd	4 th	-.47	.27	.27	-1.13	.20
		5 th	.34	.29	.71	-.37	1.06
	4 th	3 rd	.47	.27	.27	-.20	1.13
		5 th	.82*	.28	.01	.13	1.50
	5 th	3 rd	-.35	.29	.71	-1.06	.37
		4 th	-.82	.28	.01	-1.50	-.13

Degree attainment. Teachers' assessment beliefs were also analyzed by levels of education. Within this independent variable, the researcher identified four levels: Bachelor's, Master's, postgraduate certificate, and Doctorate. It should be noted that no participant documented successful attainment of a Doctoral degree at the time of survey completion. The means for each belief by degree attainment are listed in Table 26. For student accountability, the mean score for teachers with a bachelor's degree was higher than the mean score for those with postgraduate and Master's degrees. That trend was consistent for assessment as irrelevant; however, means for teachers with Bachelor's degrees were lower than Master's recipients for both assessment for school accountability and improvement. Table 26 shows teachers with Master's degrees as having the highest mean for assessment as improvement and assessment as irrelevant, while teachers with postgraduate certificates did not have the highest mean for any assessment belief.

Table 26

Comparison of Classroom Assessment Belief Mean Scores for Degree Attainment

Variables	Bachelor's			Postgraduate Certificate			Master's			Doctorate		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Student Accountability	33	4.54	.71	10	3.77	.65	33	3.94	.73	0	0	0
Irrelevant	33	3.35	.73	10	3.36	.66	32	3.55	.73	0	0	0
School Accountability	34	3.87	1.20	10	3.27	1.05	34	3.61	.92	0	0	0
Improvement	33	4.20	.62	9	4.00	.60	32	4.37	.49	0	0	0

An ANOVA was conducted to determine if any of the differences were statistically significant (see Table 27). The results of the ANOVA showed a statistically significant difference between degree attainment and student accountability, with no other significant difference between the mean scores for the three remaining dependent variables. Table 26 documents mean scores for teachers with Bachelor's degrees of 4.54, while postgraduate certificate recipients' mean score was 3.77 and Master's was 3.94.

Table 27

ANOVA of Assessment Beliefs for Degree Attainment

Beliefs		<i>Df</i>	<i>F</i>	<i>p</i>
Student				
Accountability	Between Groups	2	7.73	.001*
	Within Groups	73		
	Total	75		
Irrelevant	Between Groups	2	.69	.51
	Within Groups	72		
	Total	74		
School				
Accountability	Between Groups	2	1.38	.26
	Within Groups	75		
	Total	77		
Improvement	Between Groups	2	1.81	.17
	Within Groups	71		
	Total	73		

Note. * $p < .05$

A Bonferroni *post hoc* analysis (see Table 28) was run to determine where within the four levels of degree attainment the statistically significant difference existed. The post hoc analysis indicated the mean score for student accountability for teachers with Bachelor's degrees was statistically different from those of the teachers earning postgraduate certificates and Master's degrees. The average composite for teachers with Bachelor's degrees was 4.54 compared to means of 3.77 for postgraduate certificate and 3.94 for Master's.

Table 28

Bonferonni Post Hoc for Assessment Belief (Student Accountability) and Degree Attainment

Assessment Belief	(I) What grade level do you teach?	(J) What grade level do you teach?	Mean Difference (I-J)	Std. Error	Sig.	95% Conf. Int.	
						Lower Bound	Upper Bound
Student accountability	Bachelor's	Post. Cert.	.79*	.26	.01	.14	1.40
		Master's	.56*	.18	.00	.17	1.02
	Postgraduate Certificate	Bachelor's	-.77*	.26	.01	-1.40	-.14
		Master's	-.17	.26	1.00	-.80	.46
	Master's	Bachelor's	-.60*	.17	.00	-1.02	-.17
		Post. Cert.	.17	.26	1.00	-.46	.80

Teachers' reported levels of importance for assessment practices were examined to determine if value varied according to the four independent variable levels of degree attainment: Bachelor's, postgraduate certificate, Master's, and Doctorate. No data is reported for the independent variable doctorate level due to no respondents having attained this degree at the time of the survey. Table 29 shows a comparison of means for assessment practices by degree attainment. Standard composite scores for Bachelor's degree ranged from 2.82 to 4.00. Postgraduate certificate composite scores for assessment practices ranged from 3.00 to 4.40, and Master's means ranged from 2.47 to 4.62. When comparing the ranges for all levels of the independent variable, it suggests that average levels for assessment practices were relatively varied. Furthermore, the standard deviations for each practice within each level suggest levels of variability within and across independent variable levels.

Table 29

Assessment Practices Means by Degree Attainment

Variable	Bachelor's			Post. Cert.			Master's		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Designed by self	35	3.74	.92	10	3.90	.88	34	3.91	.87
Performance quizzes	35	3.66	.77	10	3.80	.92	32	3.53	.72
Objective assessments	35	3.46	.78	10	3.30	.95	34	3.35	.77
Short answer	34	3.76	.82	10	4.00	.67	34	3.91	.71
Performance assessment	35	3.97	.71	10	3.90	.57	33	4.09	.81
Projects by self	34	3.59	.78	10	4.00	.67	33	3.70	.95
Major exams	34	3.35	1.01	10	3.10	.88	34	2.94	.95
Authentic assessments	34	4.00	.82	10	4.40	.70	34	4.62	.74
Projects in teams	35	3.03	.89	9	3.56	.88	33	3.36	1.19
Publisher assessments	34	2.82	1.00	10	3.00	.82	34	2.47	.93
Oral presentations	34	3.50	.83	10	3.50	.97	34	3.62	.82

Note. Means range from 1 (*Not Important*) to 5 (*Very Important*).

Tables 30 and 31 reflect ANOVA and Bonferonni *post hoc* results for assessment practices by degree attainment. The results of the ANOVA showed a significant difference according to authentic assessment practices and degree attainment ($p = .01^*$). A Bonferonni *post hoc* analysis revealed the specific difference between the three levels of this independent variable. A significant difference was found for authentic assessments in relation to teachers who have attained a Bachelor's versus a Master's degree. The mean difference was a -.62 with teachers having earned a Bachelor's degree having a mean composite score of 4.00 and Master's recipients' mean score of 4.62. The most highly educated teachers scored significantly higher on the importance of authentic assessments than teachers with Bachelor's degrees.

Table 30

ANOVA for Assessment Practices by Degree Attainment

Practices		<i>Df</i>	<i>F</i>	<i>p</i>
Designed by self	Between Groups	2	.34	.71
	Within Groups	76		
	Total	78		
Performance quizzes	Between Groups	2	.53	.59
	Within Groups	74		
	Total	76		
Objective assessments	Between Groups	2	.22	.80
	Within Groups	76		
	Total	78		
Short Answer	Between Groups	2	.52	.60
	Within Groups	75		
	Total	77		
Performance assessments	Between Groups	2	.36	.70
	Within Groups	75		
	Total	77		
Projects by self	Between Groups	2	.91	.41
	Within Groups	74		
	Total	76		
Major exams	Between Groups	2	1.54	.22
	Within Groups	75		
	Total	77		
Authentic assessments	Between Groups	2	5.54	.01*
	Within Groups	75		
	Total	77		
Projects by team	Between Groups	2	1.39	.26
	Within Groups	74		
	Total	76		
Publisher assessments	Between Groups	2	1.78	.18
	Within Groups	75		
	Total	77		
Oral presentations	Between Groups	2	.19	.83
	Within Groups	75		
	Total	77		

Note. * $p < .05$

Table 31

Bonferonni Post Hoc for Assessment Practice (Authentic Assessment) and Degree Attainment

Assessment Belief	(I) What grade level do you teach?	(J) What grade level do you teach?	Mean Difference (I-J)	Std. Error	Sig.	95% Conf. Int.	
						Lower Bound	Upper Bound
Student accountability	Bachelor's	Post. Cert.	-.40	.28	.46	-1.08	.28
		Master's	-.62*	.19	.00	-1.07	-.16
	Postgraduate Certificate	Bachelor's	.40	.28	.46	-.28	1.08
		Master's	-.22	.28	1.00	-.90	.46
	Master's	Bachelor's	-.62*	.19	.00	.16	1.07
		Post. Cert.	.22	.28	1.00	-.46	.90

Level of assessment training. To determine the descriptive and inferential statistics associated with teachers' level of assessment training and beliefs and practices, the researcher conducted five different independent sample t-tests to analyze the following question, "What training in educational assessment have you had (Tick all that apply)?" Respondents could select all that applied from five responses (none, completed an undergraduate assessment course, ½ to 1 day workshop provided by current or previous employer, ½ to 1 day workshop provided by outside agency, and completed a graduate assessment course. An "Other" text box was offered; however, no responses were provided. Table 32 documents the mean, standard deviation, and frequency for each response item.

Table 32

Comparison of Classroom Assessment Belief Mean Scores for Types of Assessment Training

Beliefs	None			Undergraduate Course			Workshop by Current or Previous Employer			Workshop by Outside Agency			Graduate Course		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Student Accountability	11	4.36	1.10	22	4.21	.61	42	4.30	.68	7	4.33	.58	23	3.83	.74
Irrelevant	9	3.15	.67	23	3.63	.82	43	3.44	.70	7	3.46	.67	22	3.49	.74
School Accountability	11	3.91	1.27	24	3.60	.94	41	3.73	1.15	7	4.14	.74	24	3.60	.92
Improvement	10	4.40	.54	23	4.13	.67	38	3.58	.60	7	4.38	.48	23	4.41	.55

Note. Scale ranges from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*).

Table 33 reflects results from the five Independent t-tests, which were conducted by the researcher to determine if there were significant differences among assessment beliefs by types of training. All composite averages yielded no statistically significant differences with the exception of teachers who had received a graduate course in assessment training ($p = .01$). Additionally, although the p value of .075 is not statistically significant, the researcher notes the practical importance workshops provided by current and previous employers appear to have upon third through fifth grade teachers.

Table 33

Independent Samples t-tests for Assessment Beliefs and Types of Assessment Training

Beliefs	None			Undergraduate Course			Workshop by Current or Previous Employer			Workshop by Outside Agency			Graduate Course		
	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>
Student Accountability	.96	76	.34	.39	76	.70	1.80	76	.08	.63	76	.53	-2.55	76	.01*
Irrelevant	-1.29	74	.20	1.60	74	.12	.13	74	.90	.11	74	.92	.45	74	.66
School Accountability	.82	77	.42	-.35	77	.73	.59	77	.56	1.24	77	.22	-.35	77	.73
Improvement	.98	73	.33	-1.08	73	.29	-.03	73	.98	.70	73	.48	1.77	73	.08

Note. *Sig.* = 2 tailed test.

The means and standard deviations for assessment practices by types of assessment training are reported in Table 34. Table 34 documented teachers with no assessment training yielded the highest composite means for performance assessments and major exams ($M = 3.89$) and authentic assessments and assessments designed by self ($M = 3.80$). The remaining four levels of assessment training, undergraduate course, workshop by current or previous employer, workshop by outside agency, and graduate course, reveal three assessment practices with the highest means within their independent variable level: short answer, assessments designed by self, and authentic assessments. The lowest means across all assessment training levels suggest a trend related to projects in teams and publisher assessments (see Table 34).

The results of twenty independent *t*-tests yielded two significant differences among assessment practices by types of assessment training (see Table 35). The use of *t*-tests enabled the researcher to compare the two samples (yes or no to types of training) so inferences could be made about the population from which the sample was drawn from. Similar to results for assessment beliefs and teachers who earned post-graduate or master's degrees, Table 35 shows how teachers who completed assessment training at the graduate level revealed a significant

difference for the student accountability belief. The results indicate that advanced assessment training may impact a third through fifth grade teacher's belief in relation to assessment for student accountability. Additionally, when analyzing assessment practices by assessment training, results indicate significant differences between teachers who have had no assessment training and major exams and teachers who have completed a graduate assessment course and authentic assessments.

Table 34

Comparison of Means for Assessment Practices by Assessment Training

Beliefs	None			Undergraduate Course			Workshop by Current or Previous Employer			Workshop by Outside Agency			Graduate Course		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Designed by self	10	3.80	.92	24	3.79	.88	42	3.81	.86	7	4.14	.90	24	3.79	.98
Performance quizzes	10	3.60	1.00	24	3.83	.64	42	3.64	.76	7	3.71	.76	24	3.45	.67
Objective assessments	10	3.40	1.00	24	3.42	.83	42	3.29	.71	7	3.71	.49	24	3.33	.76
Short answer	10	3.60	1.08	24	3.96	.64	41	3.88	.78	7	4.14	.69	24	3.92	.72
Performance assessment	9	3.89	.60	24	4.04	.81	42	4.00	.83	7	4.00	.82	24	4.21	.66
Projects by self	9	3.78	.83	24	3.63	.97	41	3.63	.80	7	3.43	.98	24	3.92	.88
Major exams	9	3.89	.60	24	3.04	1.08	42	3.21	1.00	7	3.29	1.11	24	2.83	.96
Authentic assessments	10	3.80	.79	24	4.21	.88	41	4.32	.82	7	4.00	.82	24	4.83	.48
Projects in teams	10	3.20	1.03	24	3.21	1.06	41	3.24	1.02	7	3.00	1.00	23	3.48	1.12
Publisher assessments	9	3.22	1.09	24	2.71	.86	42	2.62	.83	7	3.00	.82	24	2.46	.93
Oral presentations	10	3.60	.97	24	3.33	.76	42	3.57	.80	6	3.33	1.03	24	3.63	.71

Note. Means range from 1 (*Not Important*) to 5 (*Very Important*).

Table 35

Independent Samples t-tests for Assessment Practices and Types of Assessment Training

Beliefs	None			Undergraduate Course			Workshop by Current or Previous Employer			Workshop by Outside Agency			Graduate Course		
	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>	<i>t</i>	<i>df</i>	<i>Sig.</i>
Designed by self	.14	77	.89	.2759	77	.77	.28	77	.78	-.96	77	.34	.29	77	.77
Performance quizzes	.10	75	.92	-1.65	75	.10	-.24	75	.81	-.33	75	.74	1.24	75	.22
Objective assessments	-.03	77	.97	-.18	77	.86	1.28	77	.20	-1.13	77	.26	.44	77	.66
Short answer	-1.17	76	.25	-.74	76	.46	-.24	76	.82	-1.05	76	.30	-.45	76	.65
Performance assessment	.54	76	.59	-.23	76	.82	.17	76	.87	.05	76	.96	-1.60	76	.12
Projects by self	-.34	75	.74	.44	75	.66	.60	75	.55	.85	75	.40	-1.61	75	.11
Major exams	-2.53	76	.01*	.60	76	.55	-.71	76	.48	-.41	76	.68	1.89	76	.06
Authentic assessments	2.22	76	.03	.81	76	.42	.04	76	1.00	1.09	76	.28	-4.07	76	.00*
Projects in teams	.11	75	.91	.14	75	.89	-.09	75	.93	.65	75	.54	-1.34	75	.18
Publisher assessments	-1.79	76	.08	-.10	76	.92	.73	76	.47	-1.03	76	.40	1.45	76	.15
Oral presentations	-.20	76	.84	1.56	76	.12	-.23	76	.82	.55	76	.51	-.52	76	.61

Note. *Sig.* = 2 tailed test.

Question 4: How do teachers' assessment beliefs relate to the value of assessment practices?

Survey results from Brown's (2003) Conceptions of Assessment III scale and McMillan, et al. (2002) assessment practices instrument were compared to determine if the four belief subgroups (assessment is irrelevant, assessment for school accountability, assessment for student accountability or certification, and assessment for improvement) had any relationship to assessment practices' level of importance for third through fifth grade teachers. For both surveys, respondents used a five-point scale. The COAIII (Brown, 2003) scale ranged from *Strongly Disagree* to *Strongly Agree*, while the assessment practice survey ranged from *Not Important* to *Very Important*.

A Pearson product-moment correlation was computed between all assessment practices and the four assessment belief subgroups. Statistically significant relationships were detected (see Table 36) between the student accountability assessment belief subgroup and the value of the following assessment practices: performance quizzes ($r = .35$), major exams ($r = .40$), assessments provided by publishers ($r = .37$). Moderate relationships were also revealed between the assessment for school accountability belief subgroup and major exams ($r = .40$) and the importance of assessments provided by the publisher ($r = .40$). Additionally, the improvement assessment belief was found to have the weakest significant relationship ($r = .33$) with the value of major exams. There were no statistically significant relationships detected between assessment as irrelevant and the assessment practice items.

Table 36

Correlations of Assessment Belief Subgroups and Value of Assessment Practices

Assessment Practices		Student Accountability	Irrelevant	School Accountability	Improvement
Designed by self	<i>r</i>	.06	-.10	.08	.14
	Sig.	.59	.40	.47	.24
	<i>n</i>	76	75	78	74
Performance quizzes	<i>r</i>	.35**	-.10	.21	.19
	Sig.	.00	.38	.07	.12
	<i>n</i>	74	74	76	72
Objective assessments	<i>r</i>	.15	-.13	.11	-.02
	Sig.	.21	.26	.34	.85
	<i>n</i>	76	75	78	74
Short Answer	<i>r</i>	.05	.14	.05	-.11
	Sig.	.68	.22	.65	.35
	<i>n</i>	.76	74	77	73
Performance assess.	<i>r</i>	-.02	.08	-.08	.01
	Sig.	.89	.50	.47	.91
	<i>n</i>	75	74	77	73
Projects by self	<i>r</i>	.07	-.07	-.01	.11
	Sig.	.54	.53	.93	.38
	<i>n</i>	74	73	76	73
Major exams	<i>r</i>	.40**	-.15	.40**	.33**
	Sig.	.00	.19	.00	.00
	<i>n</i>	75	74	77	73
Authentic assessments	<i>r</i>	-.07	.12	-.02	.16
	Sig.	.58	.30	.88	.18
	<i>n</i>	75	74	77	73
Projects by team	<i>r</i>	.04	-.20	.11	.11
	Sig.	.72	.09	.35	.35
	<i>n</i>	74	74	76	73
Publisher assessments	<i>r</i>	.37**	-.20	.40**	.16
	Sig.	.00	.09	.00	.17
	<i>n</i>	75	74	77	73
Oral presentations	<i>r</i>	.15	-.08	.20	.14
	Sig.	.26	.51	.08	.25
	<i>n</i>	75	74	77	73

Note. Sig. = 2 tailed test.

Summary

The target population in this study included third through fifth grade teachers working across two divisions in the Commonwealth of Virginia. The participating counties collectively

had 762 third through fifth grade teachers. One hundred twenty-four teachers comprised the sample population of which 84 responded to the survey. Five respondents' data were removed from the overall results due to partial survey completion, which resulted in an overall response rate of 64%.

Demographics. This study sought to determine what assessment practices are valued by third through fifth grade teachers, what assessment beliefs third through fifth grade teachers hold, how demographic characteristics impact beliefs and importance of practices, and how assessment beliefs relate to the value of assessment practices. Descriptive statistics conducted for demographic characteristics indicated that the largest percentage of assessment training (53.2%) occurred within the context of a half or whole day workshop provided by a current or former employer. Noteworthy are the 12.7% of respondents who indicated they have had no assessment training and the relatively small number of participants who received assessment training via their undergraduate programming (30.4%).

Question 1. Descriptive statistics for the four assessment belief subgroups (improvement, student accountability, school accountability, and irrelevant) yielded a moderate range of composite averages and standard deviations. Overall mean scores ranged from 3.43 (irrelevant) to 4.25 (improvement), on a 5 point scale. A Pearson Correlation analysis of the four belief subgroups revealed mildly significant correlation coefficients for improvement and irrelevance beliefs (negative correlation) and moderately significant correlation coefficients for school accountability and student accountability beliefs and a school accountability and improvement assessment beliefs.

Question 2. When determining what assessment practices are valued by teachers, the researcher discovered third through fifth grade educators find importance in various assessment

practices. Specifically, 51% of respondents identified authentic assessments as “Very Important”. Conversely, large percentages of participants reported the following assessment types as either “Not Important” or “Slightly Important”: publisher assessments (41%), projects in teams (26%), and major exams (22%). Means for performance assessment, assessments designed by self, and short answer assessments revealed a similar level of high importance with approximate means of 3.8 for both types.

Question 3. Though significant differences were found between belief subgroup means and among various teacher characteristics (degree attainment and student accountability and types of assessment training and student accountability), the statistical differences did not necessarily suggest a practical one. Differences in mean scores for belief subgroups, whenever significant, were just slightly over half a point on a scale of 1 to 5. Standard deviations for mean scores for each significant relationship also did not indicate wide variability within each belief by characteristic.

Statistical differences for assessment practices by demographics revealed significant data associated with years of experience, grade level assignment, degree attainment, and level of assessment training. Two significant differences among practices and assessment training were identified: no assessment training and major exams and graduate course in assessment and authentic assessments. Similarly, a significant difference was found for authentic assessments in relation to teachers who have attained a Bachelor’s versus a Master’s degree. In relation to assessment practices and grade level assignment, data indicate a statistically significant difference between 4th and 5th grade teachers and projects completed by teams. Finally, when analyzing by the grade level variable, significant differences were noted for projects by self, authentic assessments, and projects by teams.

Question 4. In response to the fourth research question, “How do teachers’ assessment beliefs relate to the value of assessment practices” scores for the four assessment belief subgroups were compared to each assessment practice item. Mild statistically significant relationships were identified for the student accountability belief subgroup and performance quizzes, major exams, and assessments provided by publishers and the improvement belief and major exams. Moderate relationships were also revealed between assessment for school accountability belief subgroup and the value of major exams and publisher assessments. No statistically significant relationships were shown for the irrelevant belief and the value of assessment practices; however, many negative correlations are noted in Table 36.

Chapter V

Conclusions and Implications

Overview

The primary aims of this study were to determine what third through fifth grade teachers' endorsed as their assessment beliefs and valued as assessment practices. A quantitative, non-experimental design using survey research was employed by the researcher to address these objectives. Using third through fifth grade elementary teachers in two mostly suburban school districts in central Virginia, a web-based survey was performed to determine teachers' assessment beliefs and valued assessment practices. Belief subgroups and assessment practices were analyzed by demographic characteristics to identify any statistically significant results. The researcher also conducted correlation analyses of the four assessment beliefs with assessment practices to determine if any significant relationships existed. Four overarching questions guided this study:

1. What are elementary teachers' conceptions (beliefs) about assessment?
2. What assessment practices are valued by grade 3-5 elementary teachers?
3. What is the relationship between years of experience, grade level assignment, level of education, and assessment training and teachers' assessment beliefs and importance of practices?
4. How do teachers' conceptions of assessment relate to the value of assessment

practices?

To attend to these questions, previously validated survey instruments underwent minor adaptations to best determine what assessment beliefs third through fifth grade teachers hold, which assessment practices are most important to teachers, how demographic characteristics relate to beliefs and practices, and how teachers' assessment beliefs relate to assessment practices.

Discussion

Assessment beliefs. The Conceptions of Assessment III (COA-III) Inventory (Brown, 2003) was used to measure teachers' assessment beliefs. After conducting this 27-item inventory, the researcher used the author's previously identified belief subgroups (assessment for improvement, assessment for student accountability, assessment for school accountability, and assessment as irrelevant) to analyze the data (2007).

Not surprisingly, composite averages for assessment beliefs by subgroup reflected assessment for improvement of learning and instruction as the highest mean score. Almost the same number of respondents reported assessment for student accountability as a primary belief of third through fifth grade teachers; however the discrepancy among standard deviations indicate much more teacher response variability associated with the student accountability belief. These results may be related to the participating districts' mandate for the regular use of assessment practices, such as benchmark assessments, which can assist with identifying the need for instructional adjustments and placement of students within educational programming.

Numerous researchers noted the importance of assessment as a critical factor in the process of teaching and learning as it enables educators to evaluate student learning and utilize information to improve learning and instruction (Campbell et al., 2002; Popham, 2005; Stiggins,

2002; Harris et al., 2008; Zwick et al., 2008). This study's data for the assessment for improvement belief parallel this research and additional research by Black and Wiliam (1998), Delandshere and Jones (1999), and Brown (2003). It appears teachers who reported assessment for improvement as a major belief view the purpose of assessment as improving the quality of instruction and student learning.

The lower composite means associated with assessment for school accountability and assessment as irrelevant may indicate rather impartial endorsement of the two beliefs. School accountability results may be related to a study conducted by Englert et al. (2005) which focused on superintendents, principals, and teachers' requirement to meet data-driven performance goals and to what degree they were evaluated based on changes in student achievement. Results from the study indicated that superintendents largely hold the accountability of addressing achievement to the public. Additionally, Delandshere and Jones (1999) determined when teachers' assessment view is predominantly summative and external in nature, teachers regard assessment as a required means of conveying information to an external audience. Collectively, composite means and standard deviations for both belief subgroups indicate teachers hold slightly neutral views of these two beliefs. Minimal response variation and averages which fall between slightly agree and slightly disagree provide practical significance in that third through fifth grade participants may require further discernment among assessment beliefs in order to more effectively depict their personal assessment beliefs.

In determining relationships among subgroups, results revealed school accountability as having a moderately significant association ($r = .58$) with the improvement assessment belief. The researcher concluded that teachers' belief that assessment is about improvement of learning and teaching is also about the improvement of schools and showing school accountability.

Conversely, the irrelevant assessment belief was found to have a mild, negative correlation ($r = -.307$) with the improvement belief. Table 37 shows a comparison of Brown's data from the 2007 administration of the *COAIII* to this study's results. Brown's data is similar to the 2010 administration of the abridged *COAIII* in that despite different populations, both data sets identify a negative correlation between assessment as irrelevant and assessment for improvement. Although the 2007 results yielded a stronger negative relationship, current data from this study also indicate a mild, negative relationship. Generally, relationship trends document that those who believe in either the irrelevance or the improvement belief will not traditionally endorse the other. This pattern could potentially indicate what Brown suggested (2007) that "teachers associate improvement with what schools and teachers do and can be made accountable for" (p. 15).

Table 37

Comparison of Belief Subgroup' Correlation Coefficients: 2007 Versus 2010

Belief Subgroups	Irrelevant		Student Accountability		School Accountability		Improvement	
	2007	2010	2007	2010	2007	2010	2007	2010
Irrelevant		1	.40	.08		-.14	-.75	-.30*
Student Accountability	.40	.08		1	.50	.55**	.19	.23
School Accountability		-.14	.50	.55**		1	.41	.58**
Improvement	-.75	-.30	.19	.23	.41	.58**		1

Note. **. Correlation is significant at the 0.01 level (2-tailed); Adapted from "Conceptions of Assessment-III" by Brown, G. T. L. (2007, December). *Teachers' conceptions of assessment: Comparing measurement models for primary and secondary teachers in New Zealand*. Paper presented at the New Zealand Association for Research in Education, Christchurch, NZ.

Value of assessment practices. Third through fifth grade assessment practice means indicated that there is not one sole assessment that is valued far beyond others. However, two major types of assessment were identified by third through fifth grade teachers as having the most importance within the teachers' assessment repertoire - performance assessments and authentic assessments. Although performance and authentic assessments yielded the highest composite means scores, relatively high averages for assessments designed by the teachers and short answer assessments revealed their importance to teachers. Publisher assessments, major exams, and projects in teams reflected the lowest level of importance to teachers.

Results from this study reveal distinct similarities and differences in comparison to data gathered in 2002 by McMillan, Myran, and Workman. When interpreting these data, it is important to recognize the differences in survey purposes for the 2002 study and the current research. Specifically, McMillan et al. utilized the validated scale to analyze types of assessment used in determining grades. Frequency of use was the focus versus the current study's focus of assessment practice value or importance within the classroom.

These distinct differences in the use of the assessment practice instrument were considered by the researcher when relating previous research results to current findings. Although McMillan et al. separately assessed assessment practices for math and language arts, results indicated elementary teachers most frequently used objective assessments (math and language arts) and performance assessments and projects (language arts). Assessments in math included fewer performance assessments and projects. In comparison, the current study's findings related to objective assessments such as multiple choice and matching document some teacher value ($M = 3.39$), however not as extensive as 2002 frequency of use results.

Although the variation in previous and current results associated with multiple choice objective assessments were initially surprising to the researcher, further analysis and application to current assessment context helped the researcher develop possible conclusions. Specifically, since McMillan et al. were determining usage of assessment practices in determining grades, the rise of accountability measures in 2002 may have resulted in a high composite mean for objective assessments. One could reason that with the influx of mandated objective assessments as the primary measure of school and district accountability within Virginia, teachers would also utilize this assessment format more regularly to assign grades. Conversely, current findings for importance of assessment practices within the classroom revealed relatively minimized importance of objective assessments ($M = 3.39$) such as multiple choice tests. Beyond the two studies' disparate results and purposes of instrumentation (usage versus importance), the researcher concludes that despite Virginia Standards of Learning being assessed regularly through the use of objective assessments, third through fifth grade teachers assign greater value to a much broader spectrum of assessment types such as oral presentations, performance quizzes, projects by self, assessments designed by self, short answer, performance assessments, and authentic assessments.

Another difference between the two studies is in relation to the use of publisher assessments. While current data indicate teachers find publisher assessments fairly valuable ($M = 2.69$), 2002 results indicate much greater use of publisher assessments. Potential explanations for the heightened use of publisher assessments in 2002, may be related to counties' participation in reading textbook adoptions and subsequent basal series trainings and minimal availability of other assessment resources. The importance of this type of assessment may be reduced in 2010, as the study's current results indicate, because of the introduction of numerous assessment tools

and techniques since 2002. Teachers have far greater access to a wider variety of evaluation tools, which data reveal are valued to a more significant degree. Additionally, at least one of the two participating counties has embraced the use of varied instructional tools for the purposes of differentiating instruction, which may have lessened the use of publisher materials as primary resources for teaching, learning, and assessing. Further research on this topic could determine more formally, how newly adopted instructional techniques and resources may relate to the value of assessment practices within the current educational classroom.

Additionally, current data document heightened value for performance assessments (2010, $M = 4.01$ versus 2002, Math – $M = 2.84$ and Language Arts - $M = 3.43$), especially when considering McMillan et al used a 6-point scale versus the amended 5-point scale for the current study. Despite instrumentation purposes (usage versus importance), this finding suggests, either formally or informally, that as educators gain distance from the commencement of Standards of Learning assessment, they see greater value in performance assessments as a measure of student achievement. With further research, a more practical understanding of the relationship between value and usage of assessment practices could assist with the development of more literate assessment practitioners.

One final commonality among the two studies supports the need for teachers' continued exposure to a spectrum of assessment tools for the effective assessment of student learning within the classroom. Despite considerable variation noted among standard deviations, McMillan et al. (2002) noted great reliance on assessments prepared by the teachers. Similar findings associated with assessments designed by teachers indicate despite changes in testing accountability from 2002 to 2010, educators continue to value teacher made assessments. This data emphasize the importance of continued evaluation of teachers' assessment literacy and

exposure to preparatory coursework and ongoing training to ensure proper development of reliable and valid teacher-made assessments.

Overall, within this study third through fifth grade elementary teachers generally rated assessment practices as fairly important to very important. This suggests, like previous research by Adams and Hsu (1998) and McMillan et al. (2002) indicated, teachers agree with the need for a variety of assessment techniques.

Demographics and assessment beliefs and importance of practices. Means were compared for each level of independent variable (years of experience, grade level assignment, level of education, and completion of assessment training) to determine if there was a significant variation between teachers' ratings of assessment importance and beliefs and varying demographic characteristics. No relationship between years of experience and assessment beliefs was noted; however, there were statistically significant relationships identified between this independent variable and three assessment practices: projects by teams, projects by self, and authentic assessments. Almost all of the statistically significant relationships involved teachers with greater than twenty years of experience. However, this pattern does not appear to have any practical significance. One relationship worth noting is the highly variable relationships identified among years of experience and projects completed by teams. Specifically, every level of independent variable had a significant relation, some of which were negative. For example, when comparing teachers with less than three years of experience to those with eleven years or more, their thoughts on the value of projects in teams reflected a significant negative correlation. This suggests that teachers with less experience find this assessment practice more valuable than those with 11 or more years of experience.

One significant relationship was identified when conducting tests for significant differences among grade level assignment and the two dependent variables, assessment beliefs and practices. The mean score for projects completed by teams was significantly different between 4th and 5th grade teachers.

Teachers' assessment beliefs and practices were also analyzed by the four levels of education or degree attainment: Bachelor's, Master's, postgraduate certificate, and Doctorate. When analyzing the mean score for the student accountability assessment belief, significant differences were identified between teachers earning Bachelor's degrees and those earning postgraduate certificates and Master's degrees. The relationships between the levels of independent variable suggest that those who have not completed education beyond a Bachelor's degree believe to a significant degree that assessment measures serve student accountability purposes. Although the composite means indicate that educators who have earned higher education degrees or certificates also endorse the belief that assessment is for student accountability, it is interesting to note that small standard deviations among all three levels indicate little variability in response style. Additionally, when examining assessment practices by levels of degree attainment, similar to beliefs, a significant difference was found between Bachelor's and Master's recipients, specifically for authentic assessments. The most highly educated respondents scored significantly higher on the importance of authentic assessments than teachers with bachelor's degrees.

Prior to conducting inferential analyses of assessment beliefs by the independent variable, types of assessment training, descriptive data were calculated. Frequencies and percents for each of the five levels of this variable were tabulated and revealed data closely aligned with previous research findings. For example, approximately 13% of participants indicated that they had not

received any training in assessment, while only 30.8% completed an undergraduate assessment course. These results were surprising to the researcher for two reasons. First, Plake (1993) and Stiggins (1999) estimated that teachers spend up to fifty percent of their time on assessment-related activities. Secondly, state and federal mandates place rigid achievement benchmarks upon schools, which require teachers to remain vigilant with progress monitoring and data analysis. However, having identified these results, it appears that despite these factors, current educators continue to reflect previous researchers' findings related to teachers' inadequate levels of assessment literacy and professional development related to assessment (Plake & Impara, 1993; Stiggins, 1991, 2002a; Zwick et al., 2008).

The results of five independent t-tests yielded three significant differences among assessment beliefs and assessment practices by types of assessment training. Similar to results for assessment beliefs and teachers who earned post-graduate or Master's degrees, teachers who completed assessment training at the graduate level revealed a significant difference for the student accountability belief. The results indicate that advanced assessment training may impact a third through fifth grade teacher's belief in relation to assessment for student accountability. Additionally, when analyzing assessment practices by assessment training, results indicate significant differences between teachers who have had no assessment training and major exams and teachers who have completed a graduate assessment course and authentic assessments. It makes sense given the nature of the independent variable, type of assessment training, that results for the two most polar assessment training options, none and graduate course, yielded significant differences.

Assessment beliefs and importance of practices. For this study, assessment belief subgroup data were compared to the importance of assessment practices data to identify

relationships between the two variables. Statistically significant relationships were detected between the student accountability belief subgroup and performance quizzes, major exams, and assessments provided by publishers. These findings have implications for practice as well as future research. From a practical standpoint, consistent with Brown (2002) and Delandshere and Jones (1999), teachers who utilize assessment for the certification of student learning or to verify student learning believe that students are accountable for their performance and achievement on assessments. Brown specifically emphasized the positive and negative consequences associated with assessment for student accountability, such as tracking, grade retention, and tracking. The current study's results indicate those who endorse the student accountability belief find greater levels of importance in the aforementioned assessment practices. Although additional research could formally explain these findings, the researcher noted that both counties current use of major exams and publisher assessments results in students' placement into appropriate academic programming, such as reading groups and remedial and enrichment instructional programs.

Moderate relationships were also revealed between the assessment for school accountability belief subgroup and major exams and assessments provided by publishers. Similar to the significant relationship between student accountability and major exams and publisher assessments, the school accountability belief also reveals key assessment assertions: to certify students' final results; monitor teachers' instructional competency; and to inform parents and the community about student progress and school status (Brown, 3003; Englert, et al., 2005). These results which suggest teachers endorsing the school accountability belief also find importance in major exams and publisher assessments is not surprising to the researcher. Currently, both federal and state accountability systems, which are direct measures of school and teacher success, utilize these assessment practices to gauge and report achievement. Additionally, as was

noted in relation to the student accountability belief and publisher assessments, accountability of teachers and schools also utilizes publisher assessments, such as Phonological Awareness Literacy Screening (PALS), Qualitative Reading Inventories (QRI), and Developmental Spelling Analysis (DSA), to measure student gains, teacher effectiveness, and school success.

The value of major exams compared to the improvement assessment belief was found to have the weakest significant relationship. This result was surprising to the researcher due to the improvement belief yielding the smallest standard deviation ($SD = .58$) and highest composite mean ($M = 4.25$) among belief subgroups. The researcher expected a larger number of assessment practices to be significantly related to this assessment belief; however, only the one assessment practice was determined to have a mild correlation. Although future research can formally identify why minimal significant relationships exist between the improvement belief and value of assessment practices, Brown (2003) and Black and Wiliam (1998) describe the process of assessment for learning and improvement belief as requiring wide-ranging use of varied assessment tools, both formal and informal teacher-based, aimed at succinctly capturing students' academic profiles. As a result, it could be speculated that this study's results indicate third through fifth grade teachers who endorse this belief value a widespread number of assessment types to plan for instruction, measure student achievement, and identify the need for instructional adjustments.

Assessment as irrelevant, the fourth assessment belief, represents teachers who view assessment as unrelated to the work of educators and students (Brown, 2003). Brown noted in 2003, educators who adopt this assessment conception reject assessment due to its perceived harmful impact upon teacher autonomy and student learning and excludes the importance of teachers' intuitive evaluations, student-teacher rapport, and in-depth knowledge of curriculum

and pedagogy. There were no statistically significant relationships detected between assessment as irrelevant and the assessment practice items.

Limitations

As indicated in a previous chapter, this study experienced limitations associated with a combination of factors. Specifically, external validity in this study was compromised by three factors: participants, settings, and time frames. The schools which comprised the sample population represented only 17% of the targeted population and resulted in a relatively small sample ($n = 84$). Respondents were predominantly females who worked in suburban elementary schools, which made it challenging to determine whether similar results would occur with a different group of people or whether they are solely representative of the “local context”.

Results also reflect teachers’ self-reports of assessment beliefs and value of practices. No data were gathered to validate whether the self-reports were consistent with actual practice in the third through fifth grade classrooms. Additionally, since self-report through a survey required participant motivation, there was potential for a biased sample (Mitchell & Jolley, 2007) with only those with the greatest interest responding.

The small sample size placed constraints on external validity, and therefore, the researcher’s ability to generalize findings to other settings and environments. To complicate matters further, participants in both school districts had just recently completed extensive state-wide testing, which may have impacted teachers’ response styles and/or assessment beliefs. Since similar timing conditions may not be replicated in future survey administrations, one could not automatically assume that the same results would occur. Conclusion validity was also potentially threatened by the use of multiple ANOVAS versus the use of MANOVAS. When a

researcher conducts multiple analyses of the same data and views each analysis' data as independent, the researcher runs the risk of fishing for significant relationships that are not there.

Finally, previous researchers indicated the multi-faceted nature of teachers' assessment beliefs. This study defined assessment beliefs in a one dimensional manner, which did not address the potential for intermingling of beliefs. In a self-administered survey there is also no opportunity to ask for clarification or conduct further exploration of a response, leaving some responses either inaccurate due to a misunderstanding or the survey item's failure to elicit an accurate response. Additional work to sharpen the psychometric measures or the introduction of a qualitative measure could strengthen the research associated with how teachers truly conceptualize their assessment beliefs.

Recommendations

Implications for practice. Five major implications for practice emerged from this study.

These included:

- 1.) Teachers' conceptions of assessment, specifically assessment for improvement of instruction and learning, require knowledge of a spectrum of assessment tools and practices to effectively assess student learning within the classroom.
- 2.) Pre-service and practicing teachers require ongoing exposure to meaningful assessment professional development.
- 3.) Teachers identified performance assessments, authentic assessments, teacher designed assessments, and short answer assessments as holding the most importance within the classroom. Major exams and publisher assessments were identified as having the least value.

- 4.) Types of assessment training and degree attainment reflect the most significant relationships with assessment beliefs and importance of assessment practices.
- 5.) Teachers' assessment beliefs do relate to the importance placed on select assessment practices.

Results from this study indicate that third through fifth grade teachers embrace beliefs associated with improvement of learning and teaching. Similar to previously conducted research by Black and Wiliam (1998), Delandshere and Jones (1999), and Brown (2003), the global importance assigned to a variety of assessment practices emphasized the need for teachers' wide-ranging use of varied tools, both formal and informal, aimed at succinctly capturing students' academic profiles for the purpose of improving instruction and learning. However, the significance of documented deficits in teachers' assessment professional development (Plake & Impara, 1993; Stiggins, 1991, 2002a; Zwick et al., 2008) continues to hinder teachers' ongoing development of assessment literacy. This study's data revealed tremendous differences in teachers' exposure to assessment professional development, which strengthens the outcry for school divisions and institutes of higher education to explore the most efficient means of heightening assessment competency.

When crafting a professional development plan associated with assessment, it would behoove school districts to delve more deeply into teachers' understanding of formative assessment and their identification of performance assessments, authentic assessments, teacher designed assessments, and short answer assessments as holding the most importance within the classroom. Major exams and publisher assessments were identified as having the least value. Interestingly, these results contradict current accountability measures, which regularly measure student achievement through the use of standardized measures. Possibly teachers are perplexed

by contradictory messages from the school or district level. While critical thinking and higher-order skills are being emphasized at the building level, more content continues to be added to grade level expectations which can hinder in-depth instruction. Additionally, while encouraged to utilize rubrics, portfolios, and authentic assessments, teachers, schools, and students continue to receive rewards or sanctions for students' performance on standardized testing. Understanding reasons behind teachers' assignment of assessment value would help with more accurately defining assessment professional development which supports the use of alternative assessment approaches in addition to traditional testing strategies.

Beyond this study's validation of the importance of assessment training, continued degree attainment reflects greater levels of importance for specific assessment practices. This is important for school districts to note as they partner regularly with universities and colleges to offer opportunities for educators to participate in advanced degree attainment. When developing these partnerships, school divisions must stress the importance of offering assessment courses which address all assessment beliefs and a wide array of practices, which is necessary for fostering greater assessment literacy among teachers.

Implications for further study.

Within the context of this study, the researcher looked solely at assessment beliefs, the value of assessment practices, their relationship, and the impact of demographic variables upon both dependent variables. To move this research toward more practical applications, further research related to how assessment beliefs and the importance of assessment practices directly impact the selection and implementation of assessment practices within the classroom must be conducted. Because this study did not determine causal relationships, additional investigation on

how beliefs and assessment value impact the selection and implementation of practices would help to explain decisions made in relation to assessment within the elementary classroom.

Limited assessment training documented within this study underscores previously identified inadequacies in assessment preparatory measures. This study's results reiterate the need for continued analysis of recent graduates' feedback to discern what preparatory program changes are necessary to enhance assessment literacy. A regional effort, such as the Metropolitan Educational Research Consortium (MERC), or statewide study focusing on pre-service teachers' completion of specific coursework in classroom assessment could help expose the absence of assessment fundamentals and in turn diagnose the need for widespread programmatic changes. Additionally, future research could also support the need for quality professional development versus quantity by looking more closely at the nature of assessment training.

Conducting this study with a more narrowed instructional focus may also assist with gathering data relevant to a specific subject. Like McMillan et al. (2002), revealing data associated with assessment practices in relation to a subject may more succinctly and precisely identify significant relationships and differences. Drilling down to subject-specific data could lead to the establishment of more meaningful and relevant assessment training and practice usage. Adapting the survey in the future may also investigate the benefit of expanding the interpretation of types of assessment training to reflect a more practitioner approach, such as data analysis in teams and with administrators.

Concluding Thoughts

This research provided a quantitative study of third through fifth grade teachers' assessment beliefs and value of assessment practices. Analysis of demographic characteristics revealed significant relationships with select beliefs and practices, which should be considered

when developing ways to enhance teachers' assessment literacy. It is surprising that despite the establishment of assessment standards in 1990, this study documents the continued need for widespread staff development in the area of staff development (Plake & Impara, 1993; Stiggins, 1991, 2002a; Zwick et al., 2008). Educational leaders must understand the relationship among beliefs and assessments' value in order to provide the skills needed to effectively select and implement assessments within the classroom. Once accomplished, the school, district, state, and students, above all else, will reap the instructional and learning benefits.

List of References

List of References

- Abelson, R.P. (1979). Difference between belief and knowledge systems. *Cognitive Science*. (3), 355-366.
- Adams, J. E., & Copland, M. A. (2005). *When learning counts: Rethinking licenses for school leaders*. Retrieved from Wallace Foundation website:
http://www.wallacefoundation.org/SiteCollectionDocuments/WF/Knowledge%20Center/Attachments/PDF/When_Learning_Counts.pdf
- Adams, T. & Hsu, J. (1998). Classroom assessment: Teachers' conceptions and practices in mathematics. *School Science and Mathematics*, 98(4), 174-180.
- Airasian, P. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues and Practice*, 10(1), 13-16, 26.
- Airasian, P. (1997). *Classroom Assessment* (3rd ed.). New York: McGraw Hill.
- Assessment Reform Group (ARG). (2002). *Assessment for learning*. Retrieved from <http://www.assessment-reform-group.org/news.html>
- Ayalla, C., Shavelson, R., Ruiz-Primo, M., Brandon, P. R., Yin, Y., Furtak, E. M., & Young, D. (2008). From formal embedded assessments to reflective lessons: the development of formative assessment studies. *Applied Measurement in Education*, 21(4), 315-334.
- Bangert, A., Kelting-Gibson, L. (2006). Teaching principles of assessment literacy through teacher work sample methodology. *Teacher Education and Practice*, 19(3), 351-364.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-144.
- Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal*. 16(2), 249-261.
- Boudett, K., City, E., & Murnane, R. (2006). The data wise improvement process. *Principal Leadership*, 7(2), 53-56.
- Brookhart, S. (1999). Teaching about Communicating Assessment Results and Grading. *Educational Measurement: Issues and Practice*, 18(1), 5-13.

- Brown, G. T. L. (2003). *Teachers' instructional conceptions: Assessment's relationship to learning, teaching, curriculum, and teacher efficacy*. Paper presented at the Joint New Zealand and Australian Associations for Research in Education Conference, Auckland, NZ.
- Brown, G. T. L., (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education, 11*(3), 301-318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99*(1), 166-170.
- Brown, G. T. L. (2007, December). *Teachers' conceptions of assessment: Comparing measurement models for primary and secondary teachers in New Zealand*. Paper presented at the New Zealand Association for Research in Education, Christchurch, NZ.
- Brown, G. T. L., & Hattie, J. A. (2009, April). *Understanding teachers' thinking about assessment: Insights for developing better educational assessments*. Paper presented at the National Council of Measurement in Education Conference, San Diego, CA.
- Brown, G. T. L., & Lake R. (2006, November). *Queensland teachers' conceptions of teaching, learning, curriculum and assessment: Comparisons with New Zealand teachers*. Paper presented at the Annual Conference of the Australian Association for Research in Education, Adelaide, Australia.
- Bryman, A., & Cramer, D. (1997). *Quantitative Data Analysis with SPSS for Windows: A guide for social scientists*. New York: Harper & Row.
- Calderhead, J. (1996). Teachers' beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 709-725). New York: Simon & Schuster/McMillan.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers*. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Chapman, M. (2008). *Assessment literacy and efficacy: Making valid educational decisions* (Unpublished doctoral dissertation). University of Massachusetts Amherst, Massachusetts.
- Cizek, G., Fitzgerald, S., & Rachor, R. (1996). Teachers' Assessment Practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159-

179.

- Cowan, K. (2004). *The New Title I: The Changing Landscape of Accountability*. Washington D.C.: Thompson Publishing Group.
- Delandshere, G. & Jones, J. (1999). Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision*, 14(3), 216-240.
- Diem, K. (2003). *Maximizing response rate and controlling nonresponse error in survey research*. (Fact Sheet 997). [Electronic Version]. New Brunswick, NJ; Rutgers, The State University of New Jersey, N.J. Agricultural Experiment Station, Rutgers Cooperative Extension.
- Dixon, H., & Haigh, M. (2009). Changing mathematics teachers' conceptions of assessment and feedback. *Teacher Development*, 13(2), 173-186.
- Englert, K., Fries, D., Martin-Glenn, M., & Michael, S. (2005). *How are educators? A comparative analysis of superintendent, principal, and teachers' perceptions of accountability systems*. Aurora, CO: Mid-continent Research for Education and Learning.
- Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bass.
- Gardner, J. (2006). *Assessment and learning*. London: SAGE.
- Gay, L. R., & Arasian, P. (2000). *Educational Research: Competencies for Analysis and Application* (6th ed.). USA: Merrill/Prentice Hall.
- Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence? Teachers and national assessment of seven-year-olds*. Buckingham, UK: Open University Press.
- Gipps, C., McCallum, B., & Hargreaves, E. (2000). *What makes a good primary school teacher? Expert classroom strategies*. London: RoutledgeFalmer.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6-11.
- Harris, L., Irving, S., & Petterson, E., (2008, December). *Secondary teachers' conceptions of the purpose of assessment and feedback*. Paper presented at the annual conference of the Australian Association for Research in Education, Brisbane, Australia.
- Hargreaves, E. (2005). Assessment for Learning? Thinking outside of the (black) box. *Cambridge Journal of Education*. (35)2, 213-224.

- Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1), 69-95.
- Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007–No. 039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145.
- Hill, M. F. (2000). *Remapping the assessment landscape: Primary teachers reconstructing assessment in self-managing schools* (Unpublished doctoral thesis). University of Waikato, Hamilton, NZ.
- 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.
- Jones, J. (2004). Framing the assessment discussion. *Young Children*, 59(1), 14-18.
- Kahn, E. (2001). A case study of assessment in a grade 10 English course. *The Journal of Educational Research*, 93(5), 276-286.
- Kirkpatrick, L., Lincoln, F., & Morrow, L. (2006). Assessment of a collaborative teacher preparation program: Voices of interns. *Delta Kappa Gamma Bulletin*, 36-45.
- McMillan, J., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, (95)4, 203-213.
- McMillan, J., & Nash, S. (2000). *Teacher Classroom Assessment and Grading Practices Decision Making*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- McMillan, J., & Schumacher, S. (2006). *Research in Education: Evidence-Based Inquiry* (6th ed.). Boston: Pearson Education, Inc.
- McNair, S., Bhargava, A., Adams, L., Edgerton, S., & Kypros, B. (2003). Teachers speak out on assessment practices. *Early Childhood Education Journal*, 31(1), 23-31.
- Mitchell, M. L., & Jolley, J. M. (2007). *Research Design Explained* (6th ed.). USA: Thomson Wadsworth.

- 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., & Campbell, C. (2006). *Measuring teachers' knowledge and applications of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.
- National Council on Measurement in Education Ad Hoc Committee on the Development of a Code of Ethics (1995). *Code of professional responsibilities in educational measurement*. Retrieved from http://www.ncme.org/about/docs/prof_respons.doc
- O'Leary, M. (2008). Towards an agenda for professional development in assessment. *Journal of In-service Education, 34*(1), 109-114.
- Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Plake, B. S., (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher, 6*(1), 21-27.
- Plake, B. & Impara, J., (1993). *Teacher Assessment Literacy: Development of Training Modules*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.
- Popham, J. (2005). *Classroom assessment: What teachers need to know* (4th ed.). Boston: Allyn and Bacon.
- Popham, J. (2008). *Transformative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly, (42)*4, 203-220.
- Remesal, A., (2007). Educational reform and primary and secondary teachers' conceptions of assessment: the Spanish instance, building upon Black and William (2005). *The Curriculum Journal, (18)*1, 27-38.
- Rhodes, J., & Robnolt, V. (2007). *Alignment of District Assessments With the Virginia Standards of Learning (SOL)*, Paper presented at the Annual Metropolitan Education Research Consortium, Richmond, VA.
- Rudestam, K., & Newton, R. (2007). *Surviving your dissertation*, (3rd ed.). Los Angeles, CA: SAGE Publications.

- Stamp, D. (1987). *Evaluation of the formation and stability of student teacher attitudes to measurement and evaluation practices* (Unpublished doctoral dissertation). Macquarie University, Sydney, Aus.
- 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. (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., & 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.
- Thompson, A. G. (1992). Teacher's beliefs and conceptions: A synthesis of the research. In D. A. Groups (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: McMillan.
- Torrance, H. & Pryor, J. (1998). *Investigating formative assessment: teaching, learning and assessment in the classroom*. Buckingham, UK: Open University Press.
- Trepanier-Street, M. L., McNair, S., & Donegan, M. M. (2001). The view of teachers on assessment: A comparison of lower and upper elementary teachers. *Journal of Research in Childhood Education*. 15(2), 234-241.
- Van den Berg, B. (2002). Teachers' meanings regarding educational practice. *Review of Educational Research*, 72, 577-625.
- Warren, E. & Nisbet, S. (1999). The relationship between purported use of assessment techniques and beliefs about the uses of assessment. In J. M. Truran & K. M. Truran (Eds.). *22nd Annual Conference of the Mathematics Education and Research Group of Australasia* (pp. 515-521). Adelaide, SA: MERGA.
- Winterbottom, M., Brindley, S., Taber, K. S., Fisher, D., Finney, J., & Riga, F. (2008). Conceptions of assessment: Trainee teachers' practice and values. *Curriculum Journal*. 19(3), 193-213.

Zwick, R., Sklar, J., Wakefield, G., Hamilton, C., Norma, A., & Folsom, D. (2008). Instructional tolls in educational measurement and statistics (ITEMS) for school personnel: Evaluation of three web-based training modules. *Educational Measurement: Issues and Practice*, 27(2), 14-27.

Appendix A

Online Survey

Elementary Teachers' Assessment Conceptions (Beliefs) and Practices

Introduction:

June 7, 2010

Dear Teacher:

You have been invited to participate in a research study concerning third through fifth grade teachers' assessment beliefs and practices. Your county representative and building level administrator have granted permission to conduct this study within your school. In an effort to gather all available data, I am asking participants to complete the survey by Friday, June 18, 2010.

Thank you in advance for your support of my study. This research could not be completed without your help. Please feel free to contact me with any questions.

Kindest regards,

Sarah Calveric
Doctoral Candidate
Virginia Commonwealth University
scalveric@hcps.us

Appendix A (continued)

Consent to Participate

On the following screens, you will find a survey that will take you approximately 10-15 minutes to complete. Survey Monkey is a secure site, and all responses are sent over an encrypted connection. Your participation is entirely voluntary, and you may withdraw from this study at any time by clicking the “exit this survey” icon located at the top right hand corner of the screen. You may also choose to omit specific questions if you would prefer not to answer them. Your decision whether or not to participate will in no way jeopardize your future relations with your current employer. Should you decide to exit the study at a later date, you may also withdraw any provided information.

Be assured that any information obtained in connection with this study will remain confidential. By completing the online survey, you will be giving me permission to publish aggregated findings in my dissertation and present findings in professional journals and at professional conferences.

Part I.

Please provide the following demographic information.

A) What is your sex? (*Tick one only*)

- Female
- Male

B) Select the appropriate age range.

- 21-25
- 26-33
- 34-42
- 43 and above

What is your highest degree? (*Tick one only*)

- Bachelor
- Postgraduate Certificate
- Master
- Doctor

B) For how many years have you taught? (*Tick one only*)

- Less than 3
- Between 4 and 10
- Between 11 and 20
- More than 20

C) What grade level do you teach? (*Tick one only*)

- 3rd Grade
- 4th Grade
- 5th Grade

E) What training in educational assessment have you had? (*Tick all that apply*)

- None
- Completed an undergraduate assessment course
- ½ to 1 day workshop provided by your current or previous employer
- ½ to 1 day workshop provided by outside agency
- Completed a graduate assessment course
- Other: (*give details*)

Part II.

Conceptions of Assessment III Abridged Survey

Part II of the survey asks about your beliefs and understandings about ASSESSMENT. Please answer the questions using YOUR OWN understanding of assessment.

1. Please give your rating for each of the following 27 statements based on **YOUR** opinion about assessment. Indicate how much you actually agree or disagree with each statement. Use the following rating scale and choose the one response that comes closest to describing your opinion.

- Strongly Disagree
- Slightly Disagree
- Agree
- Mostly Agree
- Strongly Agree

Note that the ratings are ordered from *Disagree* on the LEFT to *Agree* on the RIGHT.

Please tick one box for each.

Conceptions of Assessment	Strongly Disagree	Slightly Disagree	Agree	Moderately Agree	Strongly Agree
1. Assessment provides information on how well schools are doing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Assessment places students into categories	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Assessment is a way to determine how much students have learned from teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Assessment provides feedback to students about their performance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Assessment is integrated with teaching practice	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Assessment results are trustworthy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Assessment forces teachers to teach in a way that is contradictory to their beliefs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Teachers conduct assessments but make little use of the results	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Assessment results should be treated cautiously because of measurement error	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Assessment is an accurate indicator of a school's quality	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Assessment is assigning a grade or level to student work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Assessment establishes what students have learned	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Assessment informs students of their learning needs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please tick one box for each.

Conceptions of Assessment	Strongly Disagree	Slightly Disagree	Agree	Moderately Agree	Strongly Agree
14. Assessment information modifies ongoing teaching of students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Assessment results are consistent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Assessment is unfair to students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Assessment results are filed & ignored	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Teachers should take into account the error and imprecision in all assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Assessment is a good way to evaluate a school	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Assessment determines if students meet qualifications standards	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Assessment measures students' higher order thinking skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Assessment helps students improve their learning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Assessment allows different students to get different instruction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Assessment results can be depended on	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Assessment interferes with teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. Assessment has little impact on teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. Assessment is an imprecise process	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please continue to Part III...

Appendix A (continued)

Part III.

Elementary Assessment Practices Survey

1. Please give a rating for each of the following 11 statements based on **YOUR** opinion about assessment practices. Use the following rating scale and choose the response that comes closest to describing each assessment's level of importance.

- Not Important
- Slightly Important
- Fairly Important
- Quite Important
- Very Important

Note that the ratings are ordered from *Not Important* on the LEFT to *Very Important* on the RIGHT.

Please tick one box for each

Assessment Practices	Not Important	Slightly Important	Fairly Important	Quite Important	Very Important
28. Assessments designed primarily by yourself	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Performance quizzes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. Objective assessments (e.g., multiple choice, matching, short answer)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31. Essay type questions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32. Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33. Projects completed by individual students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34. Major exams	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35. Authentic assessments (e.g., "real world" performance tasks)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36. Projects completed by teams of students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37. Assessments provided by publishers or supplied to teacher (e.g., in instructional guides or manuals)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38. Oral presentations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Thank you for your help. Your cooperation is greatly appreciated.

Appendix B

Email Survey Solicitation
May 31, 2010

Dear Principal:

As part of the requirements of Virginia Commonwealth University's Educational Leadership doctoral program, I am conducting research for the purpose of analyzing how third through fifth grade teachers' assessment beliefs relate to classroom assessment practices. It is anticipated that teachers representing sixty elementary schools in the Commonwealth of Virginia will participate in this study during the weeks of June 7 to June 18, 2010.

Your county's Director of Research and Planning has reviewed the study and permitted me to contact all principals within your school district. I would welcome your organization's participation in this 10 minute online survey. Each third through fifth grade teacher's participation is entirely voluntary. The promise of strict confidentiality is assured in both the collection and reporting of the data. Any findings obtained in connection with this study will be presented in such a way that no individual school or person will be identifiable. By completing this online survey, your teachers will be giving me permission to publish aggregated results in my dissertation, in peer reviewed journals, and at professional conferences.

As a fellow elementary principal, I am hopeful that the study's findings will assist with more clearly defining how teachers' assessment beliefs relate to the value of classroom assessment practices. Understanding current assessment beliefs and practices and formulating relevant professional development aimed at the improvement of teachers' assessment pedagogies and practices can positively contribute to instructional planning and educational success.

In acknowledgement of the Standards of Learning administration window, a second email will be sent to you on Monday, June 4, 2010. Should you approve your teachers' participation in this research study, please forward the email to the survey to all eligible participants.

Please feel free to review the attached survey instrument. Should you have any questions about this study, please contact me at scalveric@hcps.us. Thank you in advance for your time and consideration. This study could not be completed without your help.

Sincerely,

Sarah Calveric, Principal
Doctoral Candidate
Virginia Commonwealth University

Appendix C

Email Survey Solicitation June 7, 2010

Dear Teacher:

As part of the requirements of Virginia Commonwealth University's Educational Leadership doctoral program, I am conducting research for the purpose of analyzing how third through fifth grade teachers' assessment beliefs relate to classroom assessment practices. It is anticipated that teachers representing sixty elementary schools in the state of Virginia will participate in the study.

I would welcome your participation in this 10 minute online survey. Your participation is entirely voluntary, and you may withdraw from this study at any time. You may also choose to omit specific questions should you prefer to not provide a response. Your decision whether or not to participate will in no way jeopardize your future relations with your current employer. Please note, that should you determine the need to withdraw from the study at a later date, all data associated with the information you provided will be properly discarded.

The promise of strict confidentiality is assured in both the collection and reporting of the data. Any findings obtained in connection with this study will be presented in such a way that no individual will be identifiable. By completing this online survey, you will be giving me permission to publish aggregated results in my dissertation, in peer reviewed journals, and at professional conferences.

To participate in the survey:

Step 1 - Click on the link to the survey: <https://www.surveymonkey.com/>

Step 2 - Follow the instructions, clicking "next" at the bottom of every screen

Step 3 - Remember to click "done" at the end of the survey when you are finished

I am hopeful that results from this study may assist universities and districts with preparing and training teachers to utilize assessment practices in ways that enhance instructional planning and student learning. Should you have any questions about this study, please contact me at scalveric@hcps.us.

Thank you in advance for your time and willingness to share your assessment beliefs and practices. This study could not be completed without your help.

Sincerely,

Sarah Calveric
Doctoral Candidate
Virginia Commonwealth University

Vita

Sarah B. Calveric was born in 1975 in Watertown, New York. After completing her Bachelor's of Science in Elementary and Special Education at the State University of New York at Geneseo in 1997, Mrs. Calveric secured a middle school, special education teaching position in Hanover County, Virginia. While teaching sixth through eighth grades, Sarah received the Sallie Mae Beginning Teacher of the Year award. She began pursuing leadership opportunities which paralleled learning experiences offered through the Master's in Administration and Supervision program at Virginia Commonwealth University. In 2000, Mrs. Calveric requested elementary experience and transferred to a Hanover County Public School as a fourth grade, regular education teacher with a collaborative classroom. During this time, Sarah completed her Master's degree (December, 2001). In May of 2002, Mrs. Calveric was named the Assistant Principal of a neighboring HCPS elementary school. She served three years as Assistant Principal before being named in May, 2005, Principal of Cold Harbor E.S. in Hanover County, a position she still holds. During Mrs. Calveric's time at Cold Harbor, she was recipient of the Business Advisory Committee's Award for Excellence in Educational Leadership and began and completed her Ph.D. in Educational Leadership through Virginia Commonwealth University.