

INVESTIGATING THE LINK BETWEEN CURRENT
CLASSROOM TEACHERS' CONCEPTIONS,
LITERACY, AND PRACTICES
OF ASSESSMENT

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by

Mark R. Snyder

May 2017

Examining Committee Members:

Julie Booth, Ph.D., Advisory Chair, Associate Dean of Undergraduate Education

Joseph P. DuCette, Ph.D., Office of the Dean

Catherine Schifter, Ph.D., Psychological Studies in Education

Jean Boyer, Ph.D., NCSP, School Psychology

ABSTRACT

Teachers' assessment conceptions, assessment literacy, and self-reported assessment practices were investigated using a single administration survey of U.S. classroom teachers. These phenomena were investigated both individually and in their inter relationships. Assessment conceptions were measured with the Teachers' Conceptions of Assessment III – abridged survey and assessment literacy with the Assessment Literacy Inventory. Self-reported classroom assessment practices were analyzed with factor analysis to determine a set of five assessment practice factors that indicate a set of classroom assessment practice behaviors. Analysis suggested certain assessment conceptions held by teachers and aspects of their assessment literacy were significant predictors in their loadings for certain assessment practice factors. One of these significant relationships was that the degree to which the teachers held the conceptions that assessment holds schools accountable and that it aids in student improvement predicted the frequency with which they reported using tests and quizzes in their classroom. There were also significant differences in the assessment practices self-reported based upon the grade level of student instructed, years of teaching experience, as well as other demographic variables. These findings suggest that study and use of the three assessment phenomena would inform practitioners about what may influence classroom teachers' assessment practices, and how they can best be remediated.

ACKNOWLEDGMENTS

A many people were instrumental in the completion of this dissertation. I would first like to thank Dr. Julie Booth for helping me to formulate my ideas, guiding me in my research, and encouraging me to complete the project. I would also like to thank Drs. Joseph DuCette, Catherine Schifter, and Jean Boyer for their time, expertise, and feedback. I am additionally grateful for the time and encouragement shown by many other members of Temple University's faculty of the College of Education. I must also thank Drs. Gavin Brown and Craig Mertler for their consent to use their measures of assessment conceptions (Conceptions of Assessment III abridged) and assessment literacy (Assessment Literacy Inventory) respectively.

My biggest thanks are to my wife, Julia, who was invaluable on this journey and afforded me the opportunity to complete it. I could not have done it without you. Special thanks is necessary for my parents and parents-in-law, who provided much needed encouragement and on many occasions the childcare necessary to use it. Finally, I want to thank my daughters Ellie and Jemmie. You were often simultaneously the greatest source of procrastination and motivation for this endeavor. It is quite likely that one of your hugs might be how I finally finished.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER	
1. INTRODUCTION	1
2. REVIEW OF LITERATURE	4
3. METHODS	33
4. RESULTS	55
5. DISCUSSION	72
REFERENCES	83
APPENDICES	
A. PERSONAL SOLICITATION LETTER	96
B. SCHOOL PARTICIPATION LETTER	97
C. SNYDER SURVEY OF ASSESSMENT PRACTICES, CONCEPTIONS, AND LITERACY	100

LIST OF TABLES

Table	Page
3.1 <i>Frequency Distributions for Categorical Demographic Variables</i>	34
3.2 <i>Descriptive Statistics for Years of Teaching Variables</i>	36
3.3 <i>Summary of First Three Factors from Exploratory Factor Analysis Results of Assessment Practice (N = 76)</i>	43
3.4 <i>Summary of Factors Four and Five from Exploratory Factor Analysis Results of Assessment Practice (N = 76)</i>	46
3.5 <i>Number of Items of Each Scale and Expected Time to Complete</i>	53
4.1 <i>Assessment Conception Scores & Pearson Correlations</i>	56
4.2 <i>Assessment Practice Factor Pearson Correlations</i>	58
4.3 <i>Assessment Literacy Scores in Percent of Items Correct</i>	61
4.4 <i>Pearson Correlation between Assessment Practice Factors and Assessment Conceptions and Assessment Literacy</i>	62
4.5 <i>Pearson Correlation between Assessment Literacy and Teaching Experience</i>	70

LIST OF FIGURES

Figure	Page
2.1 Strength and Inter-correlations of the COA-III Conception of Assessment.....	11
3.1 Scree Plot of Principle Axis Factoring of Assessment Practices.....	41
3.2 Adopted Conceptions of Assessment Model from Brown, G.T.L. 2011.....	48

CHAPTER 1

INTRODUCTION

Since the publication of *A Nation at Risk: The imperative for Educational Reform* in 1983, academic assessment went through a fundamental change within the educational system. In light of the report's assertion that many students were being failed by their educational system and the recommendation that states submit report cards demonstrating achievement levels and progress, academic assessment began a shift from a tool to determine educational progress of specific students to its use as a barometer by which educational systems could be judged (Miller, Linn & Gronlund, 2012). This shift in assessment's role in the classroom created an assessment culture where both greater import and greater scrutiny were placed on all forms of assessment in the classroom.

The necessity for research into the classroom assessment domain is predicated upon several presuppositions. While there is debate about the method and form of assessment that should be used with various students (Alkharusi, 2008), there is consensus that the assessment practices teachers employ do have an impact on their student's achievement (Brown & Hirschfeld, 2008; Segars & Tillema, 2011). Given the impact on student achievement, the individual assessment choices teachers make within their classroom and the underlying reasons these choices are made are worth study. It is also clear that in addition to assessment's role in ultimate student achievement, the overall negative student views of the assessment these students encounter in the classroom (Nartgun, 2009) and the many teachers' low initial self-reported knowledge of and self-efficacy within the domain of classroom assessment (Dekker & Feijs, 2005; Nan et al., 2006) suggest the need for continued research.

As there is no indication that enacting teacher change in the domain of assessment differs significantly from other domains (Schwager & Carlson, 1994), research should inform the current values, attitudes and practices of the teachers in question, prior to suggesting the method or form of intervention. Schwager and Carlson conclude that there are two components to school change: the scholastic environment in which the teacher instructs, which consists of the degree of support and innovation, and the attitudes and beliefs of the teachers themselves. This research further clarifies the later.

Understanding this culture of assessment that has been cultivated in the thirty years since *A Nation at Risk*'s publication is impossible without a clear understanding of the teachers within this culture. This dissertation focused on three distinct phenomena within the assessment culture: the conceptions teachers hold about the assessments that take place in their classroom and school, the assessment practices these teachers choose to employ, and the degree of proficiency with assessment (or, assessment literacy) that teachers possess. This dissertation is an attempt to not only understand what these three phenomena look like in teachers with varying content proficiency and age of students instructed, but to investigate the link between the phenomena as well.

This dissertation used teacher survey data to suggest a set of self-reported practice profiles that should help describe the specific assessment practices that often are employed together. Further, it identified correlations between these practice profiles and teachers' assessment conceptions and assessment literacy. Ultimately, this research is designed to suggest predictive relationships between the assessment conceptions and literacy teachers might possess and the kinds of assessment practices that they report to employ.

Given the import that assessment has in the classroom, if the educational community was aware of what teachers' perceptions, practices, and literacy were with respect to assessment we could engage teachers more effectively. As these phenomena do not operate in isolation, our understanding of them should not be similarly limited. Using the relationships between these phenomena and the knowledge of what demographic characteristics are likely with respect to them in both high and low degree would enable teacher educators to tailor both curriculum and professional development. Gathering this information is the first step by which appropriate professional development and interventions can be designed and implemented in the assessment domain.

The following research questions will be addressed in this study:

1. What are the nature of current teachers' assessment conceptions, assessment practices and assessment literacy?
2. What are the relationships among teacher conceptions of assessment, teacher assessment literacy, and assessment practices?
3. What individual differences in demographics and experience characterize teachers with different conceptions of assessment, teachers that self-report different assessment practices and teachers who hold different levels of assessment literacy?

CHAPTER 2

REVIEW OF LITERATURE

The review of literature will begin with a general discussion of the role the dissertation results will play in informing assessment understanding within classroom teachers and how it will operationally define assessment. The investigation of the literature surrounding the three phenomena surveyed in this dissertation is organized with a discussion of assessment conceptions, followed by assessment literacy and concluding with a discussion of the assessment practices. The rationale for this organization mirrors the intended data analysis, as in addition to an investigation of these phenomena by themselves; their relationship will be explored with the intent to predict assessment practice factors with various assessment conception and assessment literacy scores. Finally, the review of literature will conclude with an investigation of the relationships between these phenomena as they have previously been studied.

Classroom Assessment

Despite assessment's accepted role within the educational process, it has often become an afterthought in the everyday practices of classroom teachers (Coffey, Sato & Thiebault, 2005), resulting in inefficient and incorrect measurement of the student achievement from which important decisions are made. Assessment has long been considered by teachers and administrators to be a crucial component of teacher professional development (Plake, Impara, & Fager, 1993). However, teachers have reported they are ill-equipped to assess at the level that their students need and feel that

they are unlikely to engage in a discussion about appropriate assessment with a peer or superior (Frey & Schmitt, 2007; Sammons et al., 2007).

This information suggests a gap or disconnect between what academics understand with respect to assessment conceptions, literacy and practices, and how classroom teachers view and use assessment in their classroom every day. This gap's significance continues to increase in light of the degree to which teachers and schools are evaluated on and held responsible for the level of achievement their students attain (Miller, Linn & Gronlund, 2012). Even without the need for teachers to demonstrate student achievement to government agencies, it is necessary for teachers to master the role of assessment in the classroom. They need a working understanding of proper assessment procedures in order to effectively measure their students' achievement and thereby conceivably have aided in student learning (Ruiz-Primo & Furtak, 2006).

This disconnect can only be bridged if the assessment conceptions, literacy and practices of teachers within the classroom are understood more fully. It is important to understand the level of current assessment literacy, the degree to which different assessment conceptions are held and how these conceptions and literacy translate into assessment use in the classroom. Frey and Schmitt (2007) point out the disparity in and lack of understanding of the classroom assessment environment when they describe the lack of consensus in even defining the terminology to be used for discussions of assessment and its implementation.

For the purposes of this research, the exact definition of what specifically constitutes assessment is mutable and can vary for individual participants. This is because this research is surveying self-report of specific classroom behaviors, assessment

conceptions, and assessment literacy, the latter two of which can be measured and compared without a common definition of assessment. In general, this research will adopt a global definition of assessment presented by Ysseldyke (1987) and others:

Assessment is the process of collecting data for the purpose of making decisions about individuals. In educational settings, six kinds of decisions are made: referral, screening, classification or eligibility, instructional planning, pupil evaluation, and program evaluation decisions. When we examine the extent to which tests are helpful in teaching, we must do so in the light of the kinds of decisions being made. (p.27)

This research will not focus on the classification of those assessment decisions. Instead, the aim is to further inform the field about the environment in which the assessments are designed, selected, and administered.

This understanding must extend to the entire field of classroom assessment. Once an accurate reflection of self-reported assessment practices, teacher's assessment conceptions, and assessment literacy is attained, then the relationships between the three phenomena and their component parts can be explored. These relationships will lead to a predictive understanding of how teacher assessment conceptions and literacy may suggest the use of specific classroom assessment. The individual nature of these assessment phenomena and their relationship will also be illuminated by an investigation of the individual characteristics of the teachers that possess high and low quantities of each. Potential differences between the adoption and rejection of certain assessment conceptions or practices, or the degree of assessment literacy between different kinds of teachers, may illuminate the amount of influence training and/or classroom experiences have on the formation and use of these conceptions, literacy and practices.

In light of the discussion within the field of assessment, there are a few terms that are in need of clarification. Assessments are often categorized by the intent of their

results use. In this context the terms formative and summative assessment are used. Formative assessment refers to assessment where the results are used to determine the next step of instruction or course of action. The results may help to determine the amount of method or remediation. With all formative assessment, there is an emphasis on the method and immediacy of the result feedback for the student being assessed. Summative assessment conversely is primarily focused on the certification of mastery for a particular domain. Summative assessment results suggest a finality to the instruction and a determination of competency (Miller, Linn, and Gronlund, 2012).

Teacher Assessment Conceptions

While the notion of how teachers understand assessment and its purpose has typically been viewed through the lens of beliefs, belief systems, and belief clusters, Gavin Brown applied the idea of teachers having conceptions of assessment in light of the various practical definitions that exist (Marton, 1981; Pratt, 1992) that more closely approximated the framework of assessment understanding found in teachers. Teacher conceptions are described as “a framework through which a teacher views, interprets, and interacts, with the teaching environment” (Brown, 2002, p.156). Prior to Brown’s research, mention of conceptions of assessment referred more broadly to ideas about assessment use and purpose whereby the results were explored descriptively or qualitatively.

The usage of the terms “belief” and “conception” is not neatly defined and often a cause of confusion (Pajares, 1992). For the purposes of this dissertation, the differences between assessment beliefs and assessment conceptions were adopted from previous

research (Brown, 2002; Remesal, 2011). Beliefs will suggest some basic internal truth about some aspect of one's reality, while not necessarily being objectively true or immutable through one's life (Goodenough, 1990). Beliefs are not disorganized but organized in our mind into a conception. Put differently, a teacher's conception of assessment is an organized system of that teacher's beliefs about assessment. In light of this distinction and the preference of the creator of the measure this study used, this dissertation used the term assessment conception when referring to the organized structure of ideas and beliefs about the role and importance of assessment (Brown, 2002).

Early Conception Measurement

Philippou and Christou (1997) investigated the role and usage of assessment in primary and secondary teachers in Greece and Cyprus. While they used a survey which gave teachers the opportunity to respond on a Likert-type scale, the responses were not analyzed with a more complex theoretical framework or model in mind. Instead, the responses to items dealing with what assessment should do were analyzed by percent agreement among the different groups of teachers. While the specific findings of their research suggest few sentiments about the role of mathematics assessment finding consensus among Greek and Cypriot teachers, the benefit from their work in the domain of assessment conceptions is evident by what is not done and what is not understood. By the authors own admission, this methodology is insufficient for understanding a teacher's assessment conceptions. In addition, limiting survey analysis to simple respondent consensus limits both the understanding within this domain and the possible connections with other assessment domains such as literacy and practice. This critique suggested the

necessity for a complex measure that would be able to gauge a teacher's assessment conceptions within a theoretically sound model (Brown, 2004).

Brown's Teachers' Conceptions of Assessment Inventory

Brown makes the logical assumption that since teacher conceptions of instruction influence how they teach (Brown, Lake & Matters, 2009; Pajares, 1992), their conceptions of assessment should help us understand how they assess and how we might remediate it. He developed and employed a measure that assigns a value for each responding teacher on each of four different assessment conceptions (Brown, 2004).

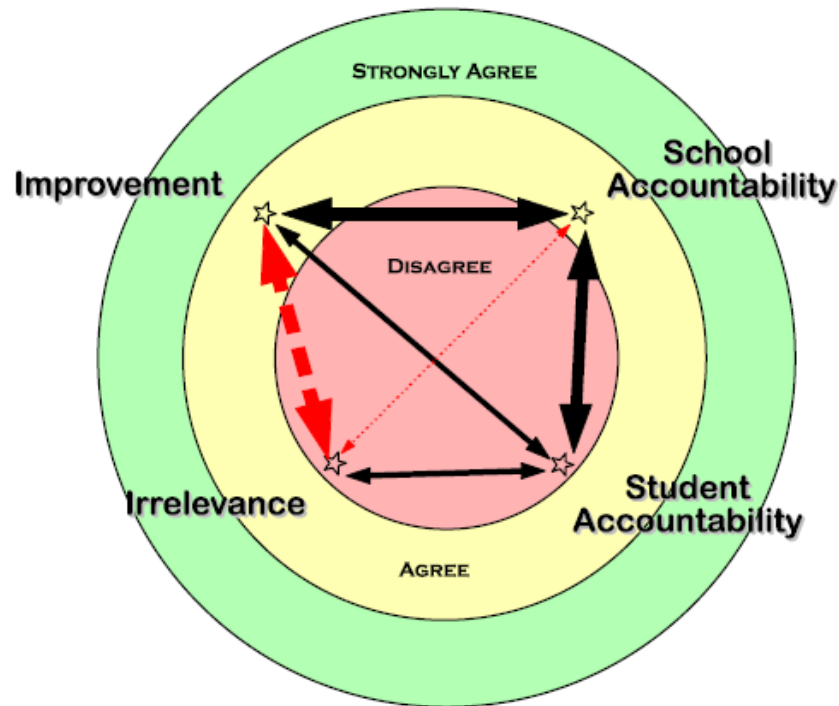
Research has suggested that teachers have four main conceptions of assessment. Brown (2004) used the self-report measure the "Teachers' Conceptions of Assessment Inventory" (CoA-III) given to primary educators and administrators in his native New Zealand to help support the theoretical notion that teachers' understanding of, and attitudes toward, assessment can be categorized in four ways. The original measure consisted of 50 Likert-type items with a 6-point agreement rating scale. An abridged version (Conceptions of Assessment Abridged, CoA-IIIA) consisting of 27 items was later validated (Brown, 2006) and was used in this study. Brown argued that the four different conceptions of assessment were that assessment improves students' teaching and learning, that assessment holds students accountable for learning, that assessment makes schools and teachers accountable, or that assessment is irrelevant and possibly harmful for both teachers and students. Brown (2004) often describes the conceptions as three purposes and one anti-purpose. Brown's theoretical framework for the development of this measure is that teachers can and do hold contradictory conceptions

as teachers often see assessment as serving contradictory purposes. For this reason, each conception will achieve a numeric value denoting the degree to which that conception is held or agreed with.

The conceptions of school and student accountability are more straightforward and are suggested with items such as “Assessment provides information on how well schools are doing” (school accountability) and “Assessment determines if students meet qualifications standards” (student accountability). Both the conceptions that assessment leads to the improvement of the learning environment and assessment is irrelevant are further denoted by sub-factors within the conception. The improvement conception is denoted by items dealing with validity, accurate description of achievement, positive role in teaching, and direct involvement in student attainment. The irrelevance conception is likewise denoted by three different factors with items related to assessment inaccuracy, ignoring of assessment results, and the general negative effect that assessment has on the classroom (Brown & Matters, 2011).

As a corollary, respondents were asked to define the term assessment by selecting the form of assessment that they were most often visualizing while taking the survey. Despite 11 options, most teachers responded that they were using common classroom tests as their example of what assessment meant. As stated previously, for the purpose of this and others measures within this research, it is unnecessary to have a common definition of assessment among participants.

Figure 2.1: Strength and Inter-correlations of the COA-III Conception of Assessment



In addition to the four conceptions of assessment being supported, the degree of inter-correlation of the four conceptions suggests that the four models exist on various continuums of assessment understanding. Figure 1 above describes the degree of correlation among the four conceptions as well as the overall population agreement with each conception individually. Thick lines suggest positive correlation and thin lines weak correlation. The strong negative correlation (red segmented line) between Irrelevance and Improvement reflects a strong positive (black continuous line) correlation between the conception that assessment improves student learning and assessment is relevant. It is worth noting that none of the conceptions in his original sample achieved an overall score suggesting strong respondent agreement in any one conception (Brown, 2004). The value for this study will likewise not be in suggesting that some percentage or consensus of teachers hold a certain assessment conception over another, but in using

the numeric score respondents achieve in the four different conceptions to help understand the relationship of the four conceptions to other assessment domains such as assessment practice and assessment literacy. A detailed description of the assessment conceptions model suggested by the Conceptions of Assessment Inventory will be presented in the methods section of this study.

Qualitative Measurement of Brown's Conceptions' Model

While Brown's CoA - III looks at assessment conceptions quantitatively; there have also been attempts to understand the idea of assessment conceptions as defined by Brown (2002) with a qualitative lens. Remesal (2011) interviewed 30 primary and 20 secondary teachers in Spain twice, one month apart. Remesal found similar conceptions to Brown, although she argued her teachers' conceptions fell into a bi-polar continuum where on one pole the focus was on monitoring teaching and learning (pedagogical - regulation pole) and the other the focus was on teacher and student accountability and the certification of achievement (societal - accreditation pole). She argued that Brown's four conception model existed within the continuum. She then placed teachers onto one of the two poles for each internal conception. Teachers were categorized by those who responded on one of the two poles for each conception, or had some mixture of the two poles, depending on the conception. While the majority of the participants had some mixed pedagogical or societal conceptions of classroom assessment, the important finding for this research is that Brown's assessment conceptions were found in the interviews, but similarly to other qualitative research in this domain, the ability to suggest relationships with other domains was limited.

Wang, Kao, and Lin (2010) used Taiwanese pre-service science teachers to research assessment conceptions with a qualitative methodology as well. Their focus was using the knowledge of the pre-service teachers' conceptions to better instruct and prepare them for effective assessing in the classroom. In-depth interviews with open ended questions were conducted to help understand how the students understood assessments' role and level of import in the classroom by asking participants questions derived from the Likert-style statements in Brown's Teachers' Conceptions of Assessment Inventory. The value in the findings for this research was in their use, as they were used as a way to engage in dialogue with the pre-service teachers and to help explore the conceptions that they held as they pertained to the assessment practices that they would employ. The suggestion by the authors is that similar engagement could be done with the original quantitative measure. This dissertation attempted to engage in that relationship between assessment conceptions and assessment practices by using in-service teachers' self-reported assessment practices.

Primary vs. Secondary Teachers' conceptions

Another important aspect of the Conceptions of Assessment Inventory - Abridged (CoA-III A) is with the model fit and configuration as it pertains to different populations of teachers. This aspect of model fit pertains specifically to what underlying or latent variables the specific items of the CoA-III A may inform about the participants. In this case, the latent variables are the four conceptions of assessment described previously. The concern is to accurately associate the individual inventory items with their appropriate conception of assessment. This also has to do with whether different items

suggest the same thing for different populations of teachers. For its use in this study, the expectation was that acceptable model fit would be achieved with this diverse population of teachers based on the model fit from previous studies. It was beyond the scope of this study to attempt to find a better fitting model of teacher assessment conceptions. The expectation of model fit is explained below.

As primary and secondary teachers have very different teaching environments, it is important to establish a commonality among the conceptions that they hold about educational assessment. It was the expectation of this research that the assessment conceptions of primary and secondary teachers can be investigated using the same model and measurement device. Brown, Lake and Matters (2011) sought to investigate the use of the “Conceptions of Assessment Inventory” (CoA-III A) with both primary (n = 784) and secondary (n = 614) teachers in New Zealand. The study found that the CoA-III A could be used with the same model configurations for both primary and secondary teachers. While the two group’s assessment conceptions proved configurally invariant, the best model fit came when the two group’s regression weights between first order factors and items as well the intercepts were allowed to vary. The authors conclude that within this population, the primary and secondary school teachers should be considered coming from different populations. This is significant for this study in suggesting that the same model be used for both groups of teachers, but that you should expect different assessment conceptions between them.

The actual differences in the conceptions’ mean scores of primary and secondary teachers were limited to the improvement and student accountability conceptions. The mean scores vary from 1 - 6 with “6” suggesting the highest level of agreement and “1”

indicating disagreement. Primary teachers in the study had statistically significantly higher mean scores within the Assessment Promotes Improvement conception than secondary teachers while secondary teachers had higher mean scores and thusly higher agreement with the conception that assessment hold students accountable.

This research suggests that the instrument is adequate and, consequently, this model was used for the target population in the current study. It also suggests that it is reasonable to compare conception mean scores and to expect differences in those conception means based on the context and assessment culture of the state, district, school.

Inventory Use with New Teacher Populations

Brown, Kennedy, Fok, Chan, and Yu (2009) studied the use of both the “Conceptions of Assessment Inventory” (CoAI-IIIa) and another self-report measure, the “Practice of Assessment Inventory” (PrAI), with primary and secondary teachers from Hong Kong to both investigate the previously validated conception models with this new population and to investigate the relationship between their conceptions of assessment and their self-reported practices. While the model fit using the same assessment conceptions model used with populations in New Zealand and Queensland was insufficient, the differences in conceptions scores suggest why adequate model fit was not attained. One difference that may explain the need for a better fitting model was the very high ($r=.91$) correlation between the student accountability and student improvement conceptions. While a positive correlation is expected, this degree suggests that despite teachers endorsing this conception to varying degrees, teachers in Hong

Kong believe that holding students accountable is positively associated with students' improvement. As the authors attributed this difference to the cultural differences between Hong Kong and previously surveyed populations, they suggested culturally similar populations would most likely not experience similar model fit inadequacies.

Brown and Michaelides (2010) sought to validate this "Conceptions of Assessment Inventory" (CoAI-IIIa) with a non-English speaking population of teachers to investigate whether previously used models were an adequate fit for this population of teachers from Cyprus. The best fitting model for this population deviated from the previous models. While similar to prior research the differences are attributed to the different policies and practices that are found between Greece and Australia and New Zealand, this research suggests a set of steps for making sense of assessment conceptions with new populations of teachers. Brown and Michaelides (2010) suggest that while other English speaking populations should expect to encounter fewer problems of this variety, these statistical procedures can be followed if a good fitting model is used.

Summary of Teachers' Assessment Conceptions

The research with respect to assessment conceptions suggests that in measuring teacher assessment conceptions quantitatively as opposed to qualitatively, the underlying theoretical model is not compromised and there is more flexibility in how the information can be used. In addition, Gavin Brown's research suggests that this dissertation should attempt to use previously validated model specifications in light of his findings with respect to using heterogeneous English speaking teacher populations of primary and secondary school age students. If, however, adequate model fit is not sufficient in this

new teacher population, there are specific steps that can be taken with model validation to ensure accurate interpretation.

Teacher Assessment Literacy

The effort of understanding and quantifying teacher ability within the domain of classroom assessment has focused on teachers' understanding of and ability to differentiate what constitutes sound and unsound assessment practices (Stiggins, 1995). This notion of assessment literacy has been uniformly measured since 1990 using the "Standards for teacher competence in the Educational Assessment of Students" developed by the National Council on Measurement in Education (NCME), the American Federation of Teachers (AFT), and the National Education Association (NEA) (NCME, AFT, NEA, 1990). The seven standards published in 1990 are as follows:

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

(NCME, AFT, NEA, 1990)

The methodology for quantifying individual teacher assessment literacy with these seven standards has been with the use of objectively scored multiple-choice questions (Mertler & Campbell, 2005; Plake, Impara, & Fager, 1993). These measures included some number of questions with the intent to assess the individual standards. Plake, Impara, and Fager (2003) used a thirty-five item questionnaire, where each standard was assessed with five different multiple-choice questions. The scores were interpreted as a raw score out of the possible five per standard and overall out of the possible thirty-five. Mertler and Campbell's Assessment Literacy Inventory (ALI) (2005) differs from the previously developed measure in that five short teaching scenarios were used as the basis for seven multiple choice questions, one pertaining to each standard. The number of items per standard and total number of items were the same for both measures. Both measures were scored by summing the correct responses for a criterion referenced score.

Plake, Impara and Fager's study had two parts, the first of which was the previously discussed measure of assessment literacy (1993b), the second of which surveyed attitudes towards and comfort with various aspects of classroom assessment and assessment in general (1993c). Their survey was given to teachers ($n = 555$) in 45 different states, and overall suggested a lack of assessment literacy. The teachers in the study achieved an average of 66% correct ($m = 23.20$) of the 35 item measure. Mertler's (2005a; 2005b) use of the Assessment Literacy Inventory (ALI), and its precursor, the Classroom Assessment Literacy Inventory (CALI), suggests a similar level of achievement ($m = 22.98$). While achievement on the seven different assessment

standards varied in both studies, the findings suggest an incomplete mastery of assessment literacy at best.

While Plake, Impara and Fager (1993) use terms such as assessment beliefs to describe the second part of their measurement, it is clearly different from Brown's (2004) definition of assessment conceptions (as explained in detail earlier in this literature review). Plake, Impara and Fager (1993) measured how positively or negatively teachers viewed various aspects and practices of assessment, while Brown's conceptions focus on the function and use of various aspects of assessment. Establishing this difference is essential as we progress with our understanding of these phenomena, as Brown's (2004) conception model allows for the degree of endorsement of various functions that assessment might serve and will therefore afforded us the opportunity to quantify the degree to which individual teachers adopted different assessment conceptions. Plake, Impara, and Fager (1993) acknowledge the limitation of only discerning a positive or negative sentiment for various assessment practices. The specific decisions on which standards of assessment this study measured are explained in chapter three.

Assessment Literacy Measurement Critiques

In addressing the potential misconceptions that exist in teachers' assessment literacy in Canada by surveying pre-service teachers about their perceived confidence in different assessment domains, Deluca and Klinger (2010) illustrate how this research improved the current understanding of assessment literacy in this study population.

While their study demonstrates the continued need for assessment literacy understanding

in their population, their method for understanding assessment literacy is inadequate compared to the use of the Assessment Literacy Inventory as they simply surveyed teacher perceived confidence in various assessment related areas. Mertler (2005b, 2009) adequately demonstrated that equating confidence in assessment and assessment literacy is likely unwarranted. The other substantial critique involves the use of pre-service teachers as the entirety of the study population. While Mertler (2003b) uses pre-service teachers as a means of comparison of assessment literacy among different groups of teachers, Deluca and Klinger suggest that sampling pre-service teachers is a valid substitute for in-service teachers. These critiques are a model for this research as in-service teachers had their assessment literacy measured and used as a point of comparison with other assessment domains.

Individual Difference in Assessment Literacy

Plake, Impara, and Fager (1993) and Mertler (2005a) were trying to determine what factors influence teacher assessment literacy. Plake et al., 1993, investigated the relationship between assessment literacy and teachers' attitudes toward and comfort with different forms of assessment and assessment background. They found few significant links between the two parts of their survey. One interesting finding was that teachers with low comfort with standardized test reporting had lower literacy scores. This suggests that in this case higher assessment literacy coincides with a greater comfort with dealing with standardized tests. No other significant differences in overall assessment literacy between teachers with various notions of the perceived utility of and comfort

with assessment were found. This is of interest for this research as it looked at the relationship between assessment conceptions and assessment literacy. Assessment practices will focus more closely with the self-reported assessment practices that take place in the respondents' classroom.

While Plake and Impara (1997) did find a difference in assessment literacy achievement between teachers with various levels of experience, the higher scoring group of teachers with more classroom experience still demonstrated assessment literacy considered to denote incomplete mastery of the assessment standards. Mertler (2005a) found similarly that teachers with differing levels of experience scored differently on six of the seven standards of assessment literacy and that the teachers' overall achievement of assessment literacy suggests partial mastery. This dissertation collected demographic information about the years spent in the classroom, but used it as a descriptive statistic to describe teachers with high and low assessment literacy. Similar demographic differences in assessment literacy among groups of teachers have not been more generally explored and were not one of the foci of this research.

Sentiment among the teaching community is that assessment literacy is still relevant as it is both an area of deficit among teachers and one that may denote confidence in other instructional areas. This is suggested by teachers who request more assessment training are more likely to lack confidence in not only classroom assessment, but in other instructional domains as well (Popham, 2009). It is the suggestion that these relationships exist, but the lack of clarity as to how they specifically operate, that suggests the need for more research. Prior research suggests that this research focus on objectively scoring (Mertler, 2005a) items related to the "Standards for teacher

competence in the Educational Assessment of Students” developed by the National Council on Measurement in Education (NCME), the American Federation of Teachers (AFT), and the National Education Association (NEA) (NCME, AFT, NEA, 1990). The expectation based on prior work in this field is that while few statistically significant differences have been found due to a lack of research, differences may exist among various demographic variables similar to years of experience.

While these standards have remained the benchmark for what teachers strive to master within the domain of assessment in the classroom, there has been continued discussion about how best to facilitate this overall competency. The two predominant methods of assessment literacy remediation focus on either the undergraduate training of the teachers (Schafer, 1993; Stiggins, 1999) or in-service teacher training (Mertler, 2009; Plake & Impara, 1993a; Zhang, 1995). Both of these approaches have studied assessment literacy in isolation, with the hope of simply improving teachers’ assessment literacy. This dissertation did not attempt to remediate assessment literacy, but instead use all three assessment domains in question (conceptions, literacy, and practice) in order to understand their relationships. The ultimate implication of the research would be to use this information to effectively cater to in-service teachers.

Teacher Assessment Practices

The investigation of teacher classroom assessment practices has employed two distinct approaches since the publication of *the Standards for Teacher Competence in the Educational Assessment of Students* (1990). Teachers’ classroom assessment practices are like any observable phenomena: they can be investigated with either the teachers’

self-reported practices, or with independent observations of the assessment practices themselves. Both of these approaches purport to investigate the actual assessment practices used in the classroom to varying degrees of authenticity. As research study within this domain that used direct observation (Bachor & Anderson, 1994) suggests that no methodology would be without bias as the difference between observer bias and self-report inaccuracy is unknown. A corollary to direct self-report of the teacher's practices is to survey teachers' positive or negative beliefs about various assessment practices. These self-reported sentiments are elicited about very specific assessment practices, as opposed to the conception structures described earlier that encompass all assessment and the conception of the value of assessing in general. The rationale is that those teachers who support or ascribe positive beliefs toward a specific practice are more likely to engage in that same assessment in their classroom. While this study only used teachers' self-report to better understand actual classroom practices, it is reasonable to explore the literature surrounding both teachers' positive and negative sentiments about various assessment practices and the assessment practices they report to employ. This literature review does not focus on research that uses an observational methodology, as that is not within the purview of this dissertation.

Value Judgments about Practices

The contention of research that investigates classroom assessment practice through the teachers' sentiments about the assessments inquired upon is that teachers who hold positive sentiments are more likely to engage in those practices and vice versa. While research does not suggest definitively that this is or is not the case, it simply

increases the degree to which the research will in fact be measuring the positive and negative sentiments only. Cizek, Fitzgerald, and Rachor (1996) suggest that this survey method increases the likelihood that the survey results will only capture the ideal of what the teachers surveyed would like to practice, and not what assessment practices they employ.

Research using teacher positive and negative beliefs about specific forms of assessment has been conducted with the hope of either legitimizing or explaining their use in the classroom. For example, Ploegh, Tillema, and Segers (2009) used teacher questionnaires to determine if four predetermined criteria of quality assessment practices (authenticity, transparency, fairness, and generalizability) were implemented in peer assessment. The authors' conclusion that peer assessment is viewed favorably by those instructors who employed it is problematic as the study's population was that of teachers who use peer assessment. This creates the risk that the teachers who do not use peer assessment may view it positively or negatively. They might choose not to engage in this particular form of assessment due to lack of confidence or understanding. Either way, this research suggests the gap that might exist when measuring assessment practices in this way.

While the true relationship between the positive or negative endorsement of classroom assessment practices and their actual use in the classroom is unknown (Cizek, Fitzgerald, & Rachor, 1996), by asking teachers to positively and negatively rate an assessment practice, there is a clear implication that there is a correct and incorrect assessment practice to be chosen. In a study of the perceptions of various grading procedures and their links to various teaching styles (Bonner & Chen, 2009), the authors

concluded that surveying the teachers with a positive and negative scale may have become problematic as the teacher's responses were skewed in light of the value judgments placed on certain grading procedures. My contention is that a better means of using self-report to determine likely classroom assessment practices is through items purposefully devoid of positive and negative associations.

In addition to investigating positive and negative sentiments as an indication of whether or not a teacher will engage in an assessment practice, teachers' beliefs about what constitutes unethical assessment practice have been used to better understand what might exist in the classroom. The seventh standard within the *Standards for Teacher Competence in the Educational Assessment of Students* (1990) is devoted to teachers' ability to recognize unethical, illegal, or inappropriate assessment. Green, Johnson, Kim, and Pope (2007) surveyed both pre-service and in-service teachers who were enrolled as either graduate or undergraduate students. Respondents were asked to categorize 36 scenarios as either ethical or unethical assessment practice. Of the 36 items, there was significant high disagreement (40 – 70%) on more than half of the items. Scenarios dealing with grading practices were one of three sections with the highest degrees of disagreement. Specific scenarios such as, "A teacher weights homework heavily in determining report card grades" were deemed unethical and ethical by approximately half of the respondents. These disagreements suggest a high level of incongruity among teachers when it comes not only to what is unethical, but also what are inappropriate classroom assessment practices. Overall, this level of consensus is not surprising in light of the many different accounts of what classroom assessment looks like. Either way, the authors conclude this method of ethical or unethical endorsement is not adequate to

understand teacher assessment practice as some teachers had reported engaging in what they considered upon report to be unethical assessment practices.

This issue highlights the need for research in this domain. While the specific recommendations for practice exist, the value and conceptions that are held by those teachers who employ such practices are unknown. An investigation into the relationship between teachers' assessment conceptions, the assessment literacy they possess and the assessment practice they report to employ seems warranted.

Self-Report Classroom Assessment Practice

When research surveys use self-report to investigate classroom assessment practices, they have focused on two aspects of assessment use: the frequency with which various assessments are reported to be used in the classroom, and the role these practices play in their classroom. Limiting the investigation to only the role of assessment and the procedure of how assessment practices may suggest student achievement (Wyatt-Smith, Klenowski, & Gunn, 2010) becomes insufficient when trying to understand how these practices coalesce and suggest a larger practice repertoire.

While this dissertation separated conceptions about assessment and the self-reported assessment practices teachers engage in, some investigations into assessment practices have combined these inquiries within the survey used. In these studies, the authors survey the frequency or degree of importance of various assessment practices while surveying the perceptions that these practices suggest (Inbar-Lourie & Donitsa-Schmidt, 2009), or what would be appropriate assessment in various contexts (McNair et al., 2003). It is a concern of this dissertation that these studies implicitly indicate value

judgments by combining what this research will treat as two separate assessment phenomena.

Another approach was taken by Winterbottom and colleagues (2008), who used a thirty-item survey to understand teacher values and practices as they pertained to assessment. Each item on the questionnaire had two Likert scales, one in which the respondent indicated the level of use of a particular practice, and one to indicate the degree to which it aided specifically in enabling students to learn. Analysis of the relationship between the responses of these two scales yielded few practices that respondents reported using frequently but valued less strongly. While the issues of assessment practice and assessment value or conceptions will also be handled within this research, this research will survey them separately. The need for this distinction is highlighted by Winterbottom et al.'s research, as when measured together, respondents did not make a distinction between these two phenomena. It is also likely that when measured together, the issue of positive and negative associations with respect to the individual assessment practices again becomes an issue.

Another issue arises when assessment practices are only a small component of a larger measure (Martinez, Stecher & Borko, 2009). This not only yields less data than a discrete measure, but also implicitly devalues the assessment practices in relation to the other assessment phenomena.

A review of research into classroom assessment practice suggests two areas that should be addressed in future research. The first is in the design and statistical analysis of the data collected with regard to self-report classroom assessment practices. While Cizek, Fitzgerald, and Ranchor (1996) have the most comprehensive assessment practices

questionnaire, they stop their analysis at the descriptive statistics of how the various demographic groups responded. Their survey looks to investigate both the frequency of assessment practices and how these practices are used to generate meaningful statements of student achievement. However, the authors do not suggest that these patterns of response might suggest a more simplified understanding of classroom assessment practice. This is an area that this dissertation attempted to address by using respondent data to suggest practice factors that might help describe assessment practice in more than a single items response.

The second area that current research into classroom assessment practice is lacking is in the investigation of the source of the assessments used by teachers in the classroom. While not studied previously, this research will help to describe the frequency with which teachers use assessment designed by themselves and designed by others. This description will help to further suggest a set of assessment practice behaviors that can be used to describe the assessment practices that teachers report to occur in their classroom.

While classroom assessment practices have been studied indirectly using teachers' positive and negative sentiments as well as more directly in conjunction with other measures of teaching practice or pedagogy, they have not been measured without some positive or negative association with the practices and with an emphasis on the creation or genesis of the assessment. In addition, while prior research has only looked at the descriptive statistics of the frequency of specific assessment practices, this research used factor analysis to help create assessment practice profiles that may more clearly explain how several assessment practices are often employed together.

Suggested Connections between the Phenomena

There is a lack of research suggestive of clear relationships between the three assessment phenomena as defined in this research. In addition to describing the research sample, this proposed research will use the individual sample's values of assessment conceptions, assessment literacy, and assessment practices to more fully understand the relationship between them. The following discussion consists of the relationships between these phenomena as previously studied.

Teacher Conceptions and Reported Practice

Brown, Kennedy, Fok, Chan, and Yu (2009) sought to interpret the relationship between assessment conception and assessment practices with the use of both the Teachers' Conceptions of Assessment Inventory (TCoA-III A) and a self-report assessment practice measure, the Practice of Assessment Inventory (PrAI). The implication was that studying these two phenomena simultaneously would help inform teachers' actual assessment practices (Brown, 2004). The relationship between the assessment conceptions from the TCoA - III and the teachers' reported endorsement of assessment practices from the PrAI suggested a strong predictive validity between teachers' conceptions and the assessment practices that they endorse. The strongest predictions were found from the "assessment improves student learning" conception to the "assessment improves teaching" ($\beta=.73$) and from the "assessment is irrelevant" conception to the "irrelevant assessment practices" ($\beta=.71$). A critique is the similar wording of items from the two measures and the small amount of time between which the two survey administrations were given. While the survey conditions do not suggest that

teachers would have been disingenuous with their reported practices, it seems logical that teachers would respond similarly to items that have similar phrasing. The PrAI also asks the teachers for practices that they endorse, or would use, not what practices that they have or currently engage in the classroom. For these reasons, PrAI was not used in this study. This research is important in establishing a relationship between conceptions and reported practice. This dissertation used another measure of self-reported practice with dissimilar wording and no clear emphasis on what may or may not constitute appropriate classroom assessment.

Research Questions

The following research questions were addressed in this study:

1. What are the nature of current teachers' assessment conceptions, assessment practices and assessment literacy?

While individual teachers' assessment practices have used several methods of investigation, the phenomena are still devoid of a clear measure that tries to capture the assessment practices that go on in a teachers' classroom without clear bias or value being placed on certain assessment practices over another. This dearth of specific quantifiable information about the phenomena necessitates the question, "What is the nature of current teachers' assessment practices?" Similarly, an inquiry is warranted into the current levels of assessment conception adoption as it is an understudied area. A precursor would be to validate the Conceptions of Assessment Inventory (CoAI-IIIa) with this, a new population of teachers. A current measurement of teachers' classroom assessment literacy is also warranted in light of

repeated teacher scores suggesting partial or incomplete mastery of classroom assessment. Individually, each of these phenomena requires a more current measurement in the classroom.

2. What are the relationships among teacher conceptions of assessment, teacher assessment literacy, and assessment practices?

Getting an accurate reflection of these three assessment phenomena facilitates an investigation into the relationships that these three may possess. What are the relationships among teacher conceptions of assessment, teacher assessment literacy, and assessment practices? By using the practice profiles this research created, correlations between these profiles and various assessment conceptions and assessment literacy can be investigated. Existing research has not suggested the potential relationship between assessment literacy, assessment practice, and assessment conception adoption as defined in this research. As such, this research aimed to suggest both correlative relationships between aspects of assessment conception adoption and assessment literacy as well as predictive relationships between an individual teacher's assessment conceptions or assessment literacy, and the likelihood this teacher self-reports various assessment practices.

3. What individual differences in demographics and experience characterize teachers with different conceptions of assessment, teachers who self-report different assessment practices and teachers who hold different levels of assessment literacy?

There is a dearth of research to suggest which individual differences are more often found in teachers who possess high and low levels of assessment conception adoption, assessment literacy, and who express self-reported assessment practices. The question of, “What individual differences in demographics and experience characterize teachers with different conceptions of assessment, teachers who self-report different assessment practices and teachers who hold different levels of assessment literacy?” would be instrumental in tailoring assessment instruction for in-service teachers. Together, answers to these three research questions would allow teacher educators to discern the area of need within classroom assessment, how to most effectively understand how assessment practices relate to other assessment phenomena, and to whom various interventions would be most needed and appropriate.

CHAPTER 3

METHODS

This dissertation is designed to examine the assessment conceptions, assessment practices, and the assessment literacy of current classroom teachers. In addition, potential relationships between these three phenomena and various teacher individual differences were explored. My examination encompassed three research questions:

- 1. What are the nature of current teachers' assessment conceptions, assessment practices and assessment literacy?**
- 2. What are the relationships among teacher conceptions of assessment, teacher assessment literacy, and assessment practices?**
- 3. What individual differences in demographics and experience characterize teachers with different conceptions of assessment, teachers who self-report different assessment practices and teachers who hold different levels of assessment literacy?**

The measure given to participants consisted of four sections. The first section was a demographic survey. The second section was a measure of the teachers' self-reported classroom assessment practices. The third section measured the teachers' conceptions of the purpose of assessment. The last section was a measure of the teachers' literacy within the field of assessment.

Research Design

This study used a single administration cross – sectional survey design. Surveys were administered online or in person on printed surveys. Survey administration took place

during the 2015 calendar year. Participants took the survey in a single session. Participants were solicited through e-mail and Internet solicitation. School districts and school administrators were also solicited for teacher access to the survey.

Participants

Participants for this study were in-service teachers from elementary, middle and high schools in public, private, and parochial schools in the United States. A total number of seventy-six ($n=76$) teachers completed the survey. Three participants' survey results were removed due to incomplete survey completion. This sample's demographics represent a convenience sample of teachers to which this research had access. They are not reflective of a random sample of all teachers currently teaching in the United States.

Descriptive statistics for the demographic variables are presented in Tables 3.1 and 3.2.

Table 3.1

Frequency Distributions for Categorical Demographic Variables.

Variable		N	Percent
Gender	Female	56	73.70
	Male	20	26.30
Student Level	Elementary (k-6)	25	32.90
	Middle Grades (6-8)	13	17.10
	Secondary (9-12)	42	55.30
	Other	1	1.30
Content Area	English / LA	25	32.89
	Mathematics	16	21.05
	Soc. St. / History	20	26.31
	Science	25	32.89
	Foreign Language	4	5.26

Table 3.1, continued

	Special Education	15	19.74
	Physical Education	1	1.32
	Elementary Math	15	19.74
	Elementary LA	15	19.74
	Other	17	22.37
Education Level	BA / BS	22	28.90
	MA / MS	52	68.40
	Ed D	1	1.30
	PhD	1	1.30
School Type	Public	45	59.21
	Parochial	25	32.89
	Private	2	2.63
	Charter	2	2.63
	Other	1	1.32
School Location	Urban	16	21.05
	Suburban	46	60.53
	Rural	4	5.26
	Other	10	13.16
Assessment Course Taken	Yes	28	36.84
	No	47	61.84
State Certified	Yes	67	88.16
	No	9	11.84
Formal Grading Policy	Yes	34	44.74
	No	32	42.11
	Unsure	10	13.16

This sample of teachers is predominantly represented by secondary school educators (55.30%) teaching in public schools (59.21%). The other academic subject areas are not disproportionately represented by the general percent of the sample. The exceptions are foreign languages (5.26%) and physical education (1.32%) educators. While this sample represents an educated group of teachers with 71.05% of the teachers sampled having attained a master's degree or higher, their education did not in general include a course pertaining specifically to assessment, as 61.84% reported not having

taken an assessment course in their studies. The majority of the sample taught in a suburban (60.53%) setting, but were split with regards to whether or not their school had a formal grading policy for measuring their student's achievement.

Table 3.2:

Descriptive Statistics for Years of Teaching Variables

	Years in Current Grade Level Teaching	Years Total In Teaching
Mean	8.91	13.84
Median	7	11
Maximum	42	52
Standard Deviation	8.34	11.87
Skewness	1.92	1.55
Kurtosis	4.68	2.04
% with 10 Years Exp. or Less	64.00%	47.70%
% with 21 Years Exp. or Less	93.30%	84.20%

The two continuously measured demographic variables pertaining to years of teaching suggest a positively skewed, leptokurtic distribution with respect to years in the classroom. This sample's experience in the classroom can be further described with the cumulative percent of the sample. While 9.80% of the sample did have over thirty years of total experience in the classroom, 84.20% of the sample had twenty-one years or less of total time in the classroom with that percent rising to 93.30% when only the current

classroom was considered. This suggests a sample with less time in the classroom than the mean may initially suggest.

Procedure

Recruitment

A convenience sample of participants was solicited by e-mail and website and personal teacher access. Teachers were solicited personally on an individual basis through an introductory e-mail explaining the study aims and a link to the online survey. The personal solicitation is attached in Appendix A. The solicitation also requested that participants pass along access to the study to other teachers that they felt comfortable contacting. Several schools and school districts were solicited with the intent that study participation would be made available to the entire teacher population. School and school district officials were sent an IRB approved solicitation outlining the study, its aims, and what would be asked of the teachers. This is attached in Appendix B. While most schools and school districts declined to participate, two high schools within the Archdiocese of Philadelphia, one approved private school, and one private Catholic elementary school made the study available to their entire teacher population for voluntary participation.

Apparatus

This measure was administered in an online format through the website *SurveyMonkey* (https://www.surveymonkey.com/s/Snyder_SofAPC_L). The survey started with a consent page and then proceeded through the survey. Once

completed, the survey made a link available to be entered into a drawing anonymously.

Measures

This study used one survey titled the Survey of Assessment Practices, Conceptions, and Literacy (Appendix C) consisting of four sections. The first section was self-report measures of demographic information. The second was a questionnaire developed for this research of general in-class assessment practices. The third measure was the Conceptions of Assessment III (TCOE-III) Abridged Survey (Brown, 2006). The fourth measure was a selection of items from the Assessment Literacy Inventory (ALI) (Mertler & Campbell, 2005). Both the COE-III and the ALI have been validated with various teacher populations. These measures and their scoring procedures are explained below.

Demographics Questionnaire

A demographics questionnaire was included to help explore individual differences in teachers that report different assessment practices and conceptions, or demonstrate various degrees of assessment literacy. The section included items addressing gender, age of students taught, content of instruction, teacher education level, scholastic setting, and years spent teaching, as well as whether an assessment course was ever taken, whether state certification is held, and the nature of teachers' perception of their undergraduate preparation regarding teaching and assessment. These items were scored

and interpreted as categorical variables except for the items inquiring about the number of years spent teaching, which was coded continuously.

Assessment Practices

The teachers' current assessment practices were explored with a 32 item, self-report measure. All of the items were scored from one to six; six always denotes the highest level of frequency or degree of use of various assessment practices. The first fourteen items gauge the teachers' frequency of use of different assessment practices for either determining reported grades or more generally evaluating student understanding. The next set asks specifically about the frequency of use for low and high stakes classroom assessments. Items 17 – 24 survey the degree to which different practices are employed in assigning and combining grades. The final eight items explore the frequency of the source of the participants' high and low stakes assessments. While these 32 items were generated for this research study, they were influenced by similar measures of teacher assessment practice used in other contexts (Cizek, Fitzgerald, & Rachor, 1995; Suurtamm, Koch, & Arden, 2010). For the present study, it was crucial that these items not convey what is or is not conceived as appropriate assessment in any academic setting, but instead measure the frequency and degree of use of various assessment practices.

Scoring

The items within the questionnaire were analyzed to detect patterns of assessment practice within the items. Principal axis factoring exploratory factor analysis was

conducted on the teacher responses to arrive at a smaller number of factors that better represent overall teacher assessment practice. Each participant received a standardized score for each of the adopted eigenvectors. Both these factor scores as well as the individual item responses were used in the analysis.

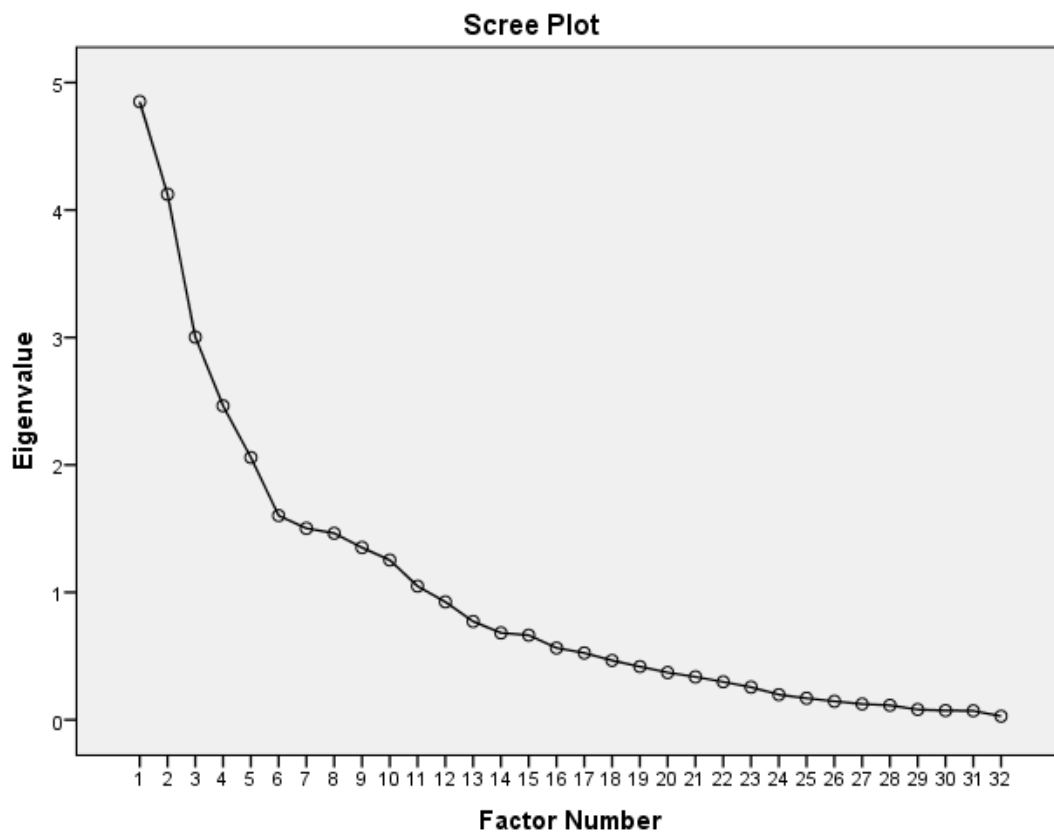
If there was missing data for four or more items within this measure list-wise deletion was used and all responses from that participant were disregarded. When respondents were missing three or less item, or 10% of the practice measure, they were given the overall sample mean for that item.

Factor Analysis Extraction

A principal axis exploratory factor analysis was used with the 32-item assessment practices measure to find a set of independent factors that better represent classroom teacher assessment practices. Once screening suggested a factorable matrix, principle axis factoring was used as it gives better estimates than principle components with small samples (Snook & Gorsuch, 1989). The factor analysis employed an oblique Promax rotation. Oblique rotation was used with the underlying factors as there is no prior research to suggest that the underlying practice factors would be unrelated and be perfectly uncorrelated. Oblique analysis was chosen as orthogonal rotation sets the factors to correlate at 0 (Yong & Pearce, 2013). Promax was used over other oblique rotation methods such as Oblimin as it yields that simplest factor structure. Ultimately both of these rotation methods produced very similar factors. The number of eigenvectors that were extracted was based on several pieces of information. The Kaiser-Guttman stopping rule (1960) was used first to limit extraction only to eigenvectors that exceed 1.

As this yielded eleven possible eigenvectors, this was deemed insufficient for factor extraction. The scree plot was then inspected along the expectation that the eigenvector extraction was to account for at least 50% of the total variance explained. The point of inflexion in the scree plot is just after the fifth factor. In addition, the five factors solution accounts for 51.56% of the variance explained. The determination was made to retain the five factors that naturally have an eigenvalue greater than 2. These guidelines were used in conjunction with an understanding of the items and their context to extract factors that produce a more coherent and simplified understanding of assessment practice.

Figure 3.1. Scree Plot of Principle Axis Factoring of Assessment Practices



The Kaiser-Meyer-Olkin Measure of sampling adequacy was used to suggest that the sample size allows the extraction with the number of variables used. Its value ranges from 0 to 1, with a recommended value of greater than .6. This study has a very small sample and had a Kaiser-Meyer-Olkin measure of sampling adequacy of .50. This is to be expected. While more participants would have been ideal, the ratio of participant to item between 2:1 and 5:1 puts this study in line with about 40% of published factor analyses (Costello & Osborne, 2005). Bartlett's test of sphericity ensures correlation among the variables by testing the null hypothesis that the correlation matrix is an identity matrix with all inter-variable correlations being 0. Bartlett's test of sphericity was significant ($\chi^2 (496) = 1398.35, p < .01$) confirming a pattern of relationships.

Once factor extraction was achieved, items are described as loading on only one factor. An item was described as loading on a factor if the loading was above .4. Items that did not load on any one factor were ignored in the description of the factors. There were no items that had factor loadings on two factors that exceeded .4. Each respondent then received a standardized score for each extracted factor. These standardized scores were used in later analyses. The extracted factors of assessment practice were analyzed and described in light of their variable membership. Factor loadings for the five extracted factors are summarized in Tables 3.3 and 3.4. They are separated into two tables for presentation clarity.

Table 3.3:

*Summary of First Three Factors from Exploratory Factor Analysis Results of Assessment**Practice (N =76)*

Item	Factor Loadings		
	External Source	Tests & Quizzes	Absolute Assessment
Source of High Stakes Assessment - District	.87	-.11	.21
Source of Low Stakes Assessment - District	.77	-.12	<.10
Source of High Stakes Assessment - Self	-.73	<.10	<.10
Source of Low Stakes Assessment - Self	-.61	<.10	.11
Source of High Stakes Assessment - Curriclm	.55	.15	<.10
Source of Low Stakes Assessment - Curriclm	.48	.14	<.10
Use Quizzes for Understanding	<.10	.86	<.10
Use Paper Pencil Test for Understanding	.12	.84	<.10
Use Quizzes for Grades	-.15	.83	<.10
Use Paper Pencil Tests for Grades	.16	.81	.15
Source of Low Stakes Assessment - Peer	.25	-.17	.70
Source of High Stakes Assessment - Peer	.19	-.15	.64
Degree Effort Influences Grade	<.10	-.11	-.61
Degree Individual Ability Influences Grade	<.10	-.11	-.58
Use In-Class Response for Grade	.27	-.21	-.50
Use Homework for Understanding	<.10	<.10	.47
Eigenvalues	4.85	4.12	3.00
% of variance	15.16	12.89	9.38

Note: Factor loadings over .40 appear in bold.

The first factor accounts for 15.16 % of the variance and suggests an underlying factor dealing with the use of externally available and possibly suggested assessments. Called “external source” for this study, the items loading on this factor focused on the source of the low and high stakes assessments the participants self-reported to use. These items suggest that those participants with high standardized scores for this factor acquire their assessments from the curriculum they use and the district in which they teach. It also suggests that they do not design the assessments themselves.

The second factor dealt exclusively with the frequency with which participants used paper and pencil tests and quizzes for both grades computation and assessing student understanding. While accounting for 12.89% of the variance it suggests a straightforward set of behaviors. This factor is suggestive of a clear assessment repertoire of frequent student testing and quizzes to understand and grade student achievement.

Factor three included items dealing with the source of assessment, what components influence grade combination, and the use of in-class discussion and homework as a means of assessing student understanding. These items together accounted for 9.38% of the variance and suggested a pattern of assessment practice that is more varied and complex than the other practice factors. In total, high standardized scores on this factor are suggestive of acquiring assessments from other teachers while also discounting the students’ perceived effort and individual ability when scoring those assessments. This factor is also representative of a use of homework for the purposes of assessing student understanding and a lack low usage of in-class responses as a means of grading. In this study, practice three will be called “absolute assessment”. This practice

factor is suggestive of a reliance on concrete and tangible products of assessment. High standardized scores are associated with the use of homework as a means of formative assessment, but not the use of a student's in-class response to generate a grade or the consideration of the student's individual ability and perceived effort to factor into the assessment score. All of these practices are congruent with the use of product oriented, absolute assessment, where score interpretation on assessments do not vary across students. In addition, this factor also addressed the source of the assessments used. This practice factor suggests a use of other teachers' assessments, while also indicating the aforementioned consideration of difficult to quantify aspects of assessment such as individual ability and effort.

This repertoire of assessment practices is representative of an underlying degree to which the educator adheres to assessment practices that avoid subjective and unstandardized methods of assessment scoring. As they also suggest the source of their assessment to be other teachers, there may be a collaborative nature to their student assessment. In practice, teachers with high scores use their peers' product oriented assessment to produce achievement oriented scores that do not suggest individual variation in ability or perceived effort. What relationship this practice factor may have with assessment literacy and assessment conceptions is described with research question two.

Table 3.4.

Summary of Factors Four and Five from Exploratory Factor Analysis Results of Assessment Practice (N =76)

Item	Factor Loadings	
	Written & Discussion	Projects
Use Written Work for Grade	.68	.26
Use Written Work for Understanding	.65	.14
Use In-Class Response for Understanding	.54	<.10
Use Projects for Understanding	<.10	.76
Use Projects for Grade	.16	.71
Eigenvalues	2.46	2.06
% of variance	7.70	6.43

Note: Factor loadings over .40 appear in bold.

Factors four and five are more straightforward with each suggesting the use of a particular assessment type for the purposes of student grades and understanding assessment. Factor 4 includes items asking about written work or varying kinds and in class responses. Factor 5 includes only 2 items about the use of projects to assess students' understanding and grade formation. Factor 4 and 5 were presented separately for ease of viewing.

Reliability

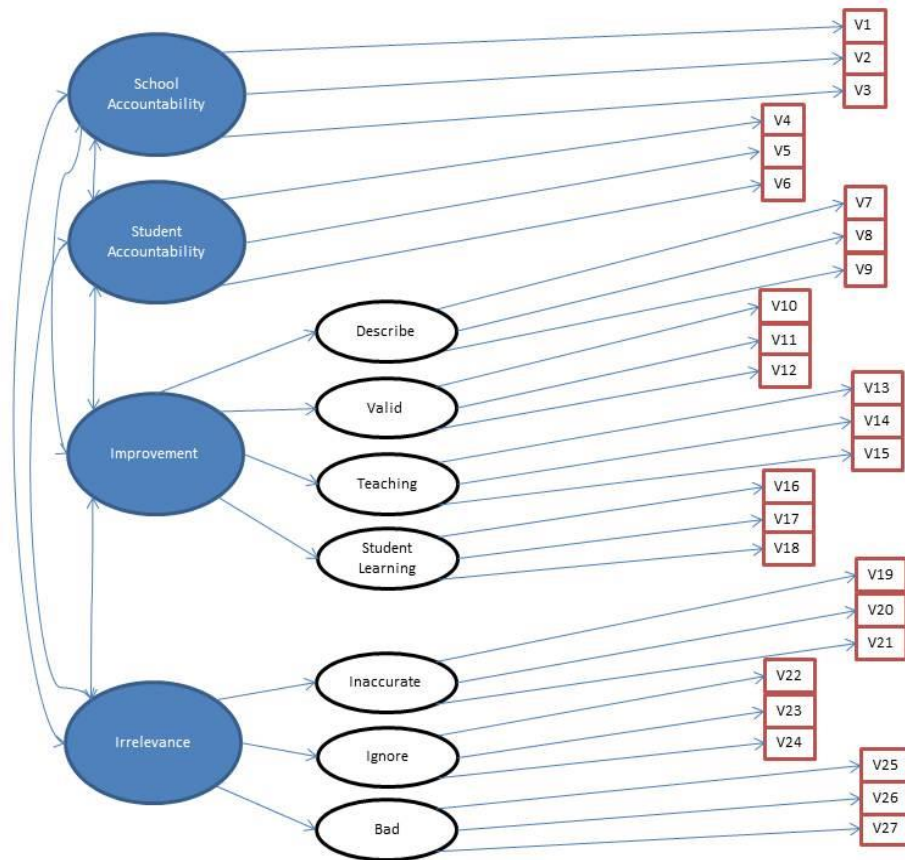
The teacher assessment practice measure has been designed for this study, and has no prior estimates of reliability. The data for these items within this study yielded a Cronbach's alpha (α) of .669.

Conceptions of Assessment

The Teachers' Conceptions of Assessment III – abridged survey (Brown, 2006) was used to discern the level of agreement with the four pre-validated conceptions of assessment: assessment is irrelevant; assessment aids student learning; assessment keeps schools accountable; or assessment keeps students accountable. This measure has been used with English and non-English speaking populations in Australia, New Zealand, and Hong Kong, and Cyprus (Brown, 2009, 2009, 2011, 2011), and significant deviation was not expected in the present population. The original survey had 50 items, but Brown (2006) has validated an abridged version of the same measure that reduces the number of items to 27. This measure and its underlying model have recently been tested to ensure that the same questionnaire can be used for both primary and secondary school teachers (Brown, 2011). The author gave permission to use the abridged COA-III for this study.

This model employed 27 primary variables, each represented by an item within the abridged COA-III. The conception factors of school accountability and student accountability each consist of the loadings of three primary variables. The factor of improvement consists of four second order factors, each with three variables. The factor of irrelevance has three second order factors that also has three variables each. This model is represented in Figure 2 below.

Figure 3.2: Adopted Conceptions of Assessment Model from Brown, G.T.L. 2011



Examples of items measuring the school accountability conception would be “Assessment provides information on how well schools are doing,” or “Assessment is an accurate indicator of a school’s quality.” Examples for the student accountability conception were items such as, “Assessment is placing students into categories,” or “Assessment determines if students have met qualification standards.” The assessment aids student improvement standard was measured with items such as, “assessment results provides feedback to students about their performance,” or “Assessment results can be depended on.” The final conception that assessment results are irrelevant used items such

as, “ Assessment forces teachers to teach in a way that is against their beliefs,” and “Teachers conduct assessments but make little use of the results.”

The model used by Brown (2011) is expected to adequately represent the latent assessment conception relationships within the present study population. Its use in that similarly heterogeneous teacher population yielded no better fitting model. This study sample is inadequately small to employ the same model fit procedures to suggest a better fitting model. Instead, a principal axis factoring exploratory factor analyses was used to understand how these latent variables were measured with this population. Any variations from the specifications of the original model would be exploratory and require future research to determine if a better fitting model should be used with respect to assessment conception scores. All scores for assessment conceptions were derived using the author’s specifications and guidelines.

Scoring

The responses from the items were summed and averaged across the latent assessment conceptions: assessment is irrelevant; assessment aids or improves student learning; assessment keeps schools accountable; or assessment keeps students accountable, giving each participant four unique scores. The unitary conception scores ranged from 1 to 6 for each conception. The author has supplied all original source material in reference to both validating the measure and scoring the individual conceptions.

Cases with fewer than 90% of the responses for the Teachers Conceptions of Assessment III were disregarded from all analysis. Values of responses missing at

random were calculated using expectation maximization (EM) using maximum likelihood procedures that use the original mean, standard deviation, and variable covariances (Brown, 2011).

Reliability and Validity

The Teachers' Conceptions of Assessment III – abridged (TCoA-IIIa) has been validated with multiple populations using the fit statistics of the model parameters to the collected data. The good fit of the proposed model to the data suggests that the four conceptions can efficiently and validly be assessed with this measure. Both populations in New Zealand ($\chi^2_{311} = 841.02$; RMSEA = .057; TLI = .87) and Queensland, Australia ($\chi^2_{933} = 3283.56$; RMSEA = .042; TLI = .81) have suggested a validly fitting model. Model fit statistics using more recent data of primary and secondary New Zealand teachers ($\chi^2_{622} = 1601.39$, $\chi^2/df = 2.58$, $p = .11$, RMSEA = .040) validates the use of one model for both groups of teachers (Brown, 2011).

Assessment Literacy

While this study measured respondents' assessment literacy, it did not measure assessment ability. Assessment literacy is specifically the engagement and ability within the discourse of assessment. It does not indicate whether a teacher uses or engages in appropriate assessment in the classroom; this could only be discerned by observing teachers in the classroom.

The Assessment Literacy Inventory (ALI) (Mertler & Campbell, 2005) was established in response to the *Standards for Teacher Competence in the Educational*

Assessment of Students (1990). This study used an adapted version of the complete measure with the author's permission. The measure uses five vignettes with seven questions for each vignette. The questions map explicitly to one of the seven standards suggested. This study employed three vignettes, with three questions for each vignette. The three standards measured are as follows:

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3. The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.

An example of an item measuring standard 2 is:

One of the middle school math teachers is redesigning her tests to make greater use of “story problems” as a way to check students’ math understanding. She consults with Mr. Valdez to see what, if any, concerns she should be aware of when constructing assessments of this type. Which statement is not an appropriate recommendation when designing story-based math tests?

- A. make sure that the reading level is grade appropriate
- B. avoid scenarios more familiar to certain groups over others
- C. check for clarity of sentence construction
- D. incorporate scenarios used during instruction**

This dissertation measured the first three standards for both a logistical rationale in designing a measure that would not induce respondent fatigue, but also a theoretical determination of the value of measuring standards four through seven for this research. Standards one through three focus on teachers’ ability to select, develop, administer and score assessments, as well as interpret the assessment’s results. In addition to there being a possible redundancy when attempting to measure standards four, five, six and seven as

discrete from the first three, the last three standards focus on theoretical tenants of appropriate assessment that may necessitate an alternate method of determination. As the last three standards focus on upholding validity within their results, facilitating assessment results communication, and recognizing ethical and unethical practices, they are decidedly different than the standards that focus on the standards that more directly relate to classroom assessment practices (Zhang, 1995).

Scoring

Each response was coded as correct, incorrect, or unanswered. Scores are displayed both as a percent correct and as a percent of answered questions with correct responses. Each teacher received four different scores; a score per standard (three) and for a total assessment literacy. Since the nine assessment literacy items were scored as either correct, incorrect, or unanswered, missing data were treated as an unanswered item.

Reliability and Validity

The individual items within the Assessment Literacy Inventory (ALI) all assess content specific to one of the standards described in the *Standards for Teacher Competence in the Educational Assessment of Students* (1990). Content validity is suggested through the individual item content.

With a population of pre-service teachers, the item analysis of the ALI suggests that this measure adequately assesses teachers' assessment literacy (Mertler, 2005). The Kuder-Richardson reliability ($r_{kr20} = .74$) was above the commonly accepted standard of .7. Item difficulty averaged .681, with a range of .212 to .992. With an average difficulty

for the sample above .5, an average item discrimination for the items of .313 suggests that the items are of high quality as items with high item difficulty (“easier” items) cannot mathematically discriminate high and low achievers. The author suggests that one area of concern for translating this to in-service teachers is the time elapsed between assessment training and taking the ALI for in-service teachers. With the median number of years’ experience for this sample of 11 years, the likelihood is that it has been at least that long since the participants received assessment training.

The number of items within each scale and the approximate time required to complete is summarized below in Table 3.5.

Table 3.5

Number of Items on Each Scale and Expected Time to Complete.

Scale	No. of items	Time to complete
Demographics questionnaire	11	3 minutes
Assessment Practices	32	6 minutes
Conceptions of Assessment	27	5 minutes
Assessment Literacy	9	9 minutes
Total	79	23 minutes

The intent of this study is to better understand these three core concepts within teacher assessment and begin to understand their relationships. It is very likely that a

model may exist to help understand how teachers' assessment conceptions, practices and literacy influence each other. It is also likely that these relationships may be predicated upon individual differences such as age of students or subject of instruction, or experience or level of teacher education. Either way, this study is meant to inform how these incredibly dynamic structures operate within different teaching environments.

CHAPTER 4

RESULTS

The dissertation results are described in order of the research questions proposed.

Question 1: What are teachers' current assessment conceptions, assessment practices and assessment literacy?

Assessment Conceptions

Normality

Data were screened for univariate normality through inspection of histograms of each of the four conceptions of assessment and calculation of skewness and kurtosis statistics. Skewness and kurtosis statistics were converted to z -scores by dividing each statistic by their respective standard errors. Values of the skewness and kurtosis z -scores greater than 1.96 are significant and would indicate a potential problem (Field, 2005). The conception that assessment helps improve student achievement (Improvement) was the only conception that violated this recommendation. The distribution was slightly leptokurtic with a kurtosis of 1.19 ($SE = .55$) and a skewness of .73 ($SE = .28$). A review of the histogram suggests sufficient normality to proceed with the analysis (Mertler, 2013)

Assessment Conceptions Scores

Teacher assessment conceptions were determined using the scores from the Conceptions of Assessment III Abridged (TCoA-IIIa) questionnaire. Descriptive statistics were generated for each of the four conceptions for the study sample. Pearson

correlations were also tested to determine inter conception relationships. They are displayed in Table 4.1.

Table 4.1.

Assessment Conception Scores & Pearson Correlations

Conception	Mean	Irrelevance	School Accountability	Student Accountability	Improvement
Irrelevance	2.99	-			
School Acc.	3.01	-.36**	-		
Student Acc.	3.75	NS	.36**	-	
Improvement	3.88	-.37**	.57**	.24*	-

* $p < .05$. *, $p < .01$ **

A repeated measures analysis of variance suggested a main effect for conception type, ($F(3,73) = 53.39, p < .01, \eta^2 = .69$). A test of Tukey's HSD was conducted to detect statistically significant differences between conception scores for the sample. The school accountability conception and the irrelevance conception were not statistically different. Nor were the student accountability and assessment improves student learning conceptions statistically different from each other. Both the school accountability and irrelevance conceptions were statistically lower than the student accountability and improvement conceptions. By themselves, this simply means that this sample of teachers held the conceptions that assessment can improve student achievement and assessment holds students accountable to a greater degree than assessment holds schools accountable and that assessment is irrelevant.

Prior research suggests the correlations that should exist if the same underlying relationships exist in this sample. The conception that assessment holds schools accountable performs as is predicted from prior research by suggesting a positive

correlation with both the conception that assessment holds students accountable, , $r = .36$, $p = .001$, and that it improves student learning, , $r = .57$, $p < .001$, while correlating negatively with the conception that assessment is irrelevant, , $r = -.36$, $p = .001$. The conception that assessment improves student learning is predicted to negatively correlate with the conception of assessment irrelevance as well, $r = -.37$, $p = .001$. The correlation between the conception that assessment holds students accountable and improves student learning, $r = .24$, $p = .036$ is also found in this study to confirm what prior research predicts. The correlation between holding students accountable and it being irrelevant is predicted to be a weak positive correlation and in this study is found to be statistically insignificantly different than zero. Other than this small deviation, the conception scores in this study correlated as predicted by prior research and suggest the same underlying conception relationships as found in previous populations (Brown, 2004). These correlations generally suggest that within this sample of teachers, the higher the degree of adoption for the conceptions of school accountability, the higher the degree that student accountability and the improvement of student learning are held and vice versa. They also suggest that the higher the degree that the conception that assessment is irrelevant is held, the lower the degree to which assessment holds school accountable and improves student learning are held. These correlations do not suggest a correlational relationship between the conceptions of irrelevance and that of assessment holding students accountable.

Assessment Practices

A detailed description of the factor analysis procedure and assessment practice factors extracted is found in chapter 3 of this dissertation. As these factors were rotated with an oblique rotation and allowed to be correlated, the assessment practice factor Pearson correlations are summarized in Table 4.2.

Table 4.2.

Assessment Practice Factor Pearson Correlations

	External Source	Tests & Quizzes	Absolute Assessment	Written	Projects
External Source		.15	-.30**	.04	.19
Tests & Quizzes			-.10	.08	.18
Absolute Assessment				.26*	-.09
Written					.25*
Projects					

* $p < .05$. *, $p < .01$ **

There are three significant correlations among the five practice factors obtained. Two of the three involve practice factor three, called absolute assessment in this study. The assessment practices of engaging in absolute grade consideration and using your peers as a source of this assessment was significantly negatively correlated with the practice factor suggesting their assessment's source was the school district and the curriculum materials, $r = -.30$, $p = .008$. This is consistent with the aspect of factor three suggesting the teachers' assessment source was other teachers. This correlation suggests

that the external source nature of factor one may not extend to peers as external sources of assessment. It also suggests that while not part of the absolute assessment practice factor, it is likely that these practices are associated with not using assessment provided from the district or curriculum while designing your own. The absolute assessment practice factor was significantly positively correlated with the practice factor of using written products, $r = .25$, $p = .026$, as well. Again, while not part of the practice factor, it may suggest that the practice of avoiding relative grading practices such as individual ability and effort does not come at the expense of avoiding more complex measures of achievement such as written essays and papers.

These factors suggest that this sample of teachers' self-reported assessment practices have discrete aspects of assessment practice. The practice factors deal with the methods of assessment, the source, or the nature of the assessment itself. Practice factor two, four and five indicate the methods of preferred assessment for this sample of teachers. Factor two indicates that teachers who use quizzes also use tests and use them for both grades and to assess general student understanding. Factors four and five are more narrow and suggest simply that teachers use written work or projects for both grading and student understanding. Practice one indicates the source of the assessments self-reported to be used. It suggests that teachers who receive their assessment from their curriculum or from their district are unlikely to create their own. Factor three indicates assessment practices that suggest a practice of generating assessments scores that indicate absolute achievement with respect to a set of educational objectives, while likely receiving those assessments from other teachers. These practice factors should not be interpreted as ability or competency with respect to these practices.

Assessment Literacy

Normality

Data were screened for univariate normality through inspection of histograms of each of the four conceptions of assessment and calculation of skewness and kurtosis statistics. Skewness and kurtosis statistics were converted to *z*-scores by dividing each statistic by their respective standard errors. Values of the skewness and kurtosis *z*-scores greater than 1.96 are significant and would indicate a potential problem (Field, 2005). None of the three assessment literacy standard scores or the cumulative assessment literacy score violated this screening. Despite this it is worth noting that with only three items per standard, there were only four possible scores for each of the individual assessment standard scores.

Literacy Scores

Teacher assessment literacy scores represent a percent correct out of the total number of items. For each of the three standards, that is three items each. The total assessment literacy score is percent correct out of the possible nine items. Unanswered items were considered incorrect responses. The three Standards of Teacher Competence in Educational Assessment of Students are as follows:

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3. The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.

(NCME, AFT, NEA, 1990)

The percent correct scores for the three standard scores and the total assessment literacy score are displayed in Table 4.3.

Table 4.3.

Assessment Literacy Scores in Percent of Items Correct

	Standard 1	Standard 2	Standard 3	Total Literacy Score
Mean	69.74	39.47	54.39	54.53
SD	23.20	31.13	30.23	18.78

A repeated measures analysis of variance suggested a main effect for literacy standard, ($F(2,74) = 31.20, p < .01, \eta^2 = .46$). A test of Tukey's HSD was conducted to detect statistically significant differences between standard scores. All three individual standard scores were statistically different suggesting a progression of difficulty. This sample considered standard one items to be the easiest items. Standard three items were more difficult, with the most difficult items and lowest percent correct pertaining to standard two. These results indicate that the sample answered 54% of the assessment items correctly. The three items dealing with the assessment development and appropriateness for instructional decisions were answered incorrectly the most often. In total these scores suggest an only partial mastery of the content within the nine items. It is important to reiterate that these items are multiple-choice items measuring the assessment standards listed above and not an indication of competency in assessment practice.

Question 2: What is the relationship between teacher conceptions of assessment, teacher assessment literacy, and assessment practices?

The relationship between the scores of assessment conceptions, assessment practices and assessment literacy was assessed through the use of Pearson correlation and stepwise multiple regressions. As causal relationships between these three constructs are not specifically suggested by previous research, all significant correlations between the five extracted practice factors and the assessment conceptions and assessment literacy scores were used to explain the variance in the practice factors. All Pearson correlations are displayed in Table 4.4.

Table 4.4.

Pearson Correlation between Assessment Practice Factors and Assessment Conceptions and Assessment Literacy

	School Account	Student Account	Improve	Irrelevant	Literacy Stand 1	Literacy Stand 2	Literacy Stand 3	Tot Lit Score
External Source	.37**	.15	.05	-.14	-.07	-.14	.03	-.09
Tests & Quizzes	.23*	.40**	.25*	-.27*	-.15	.09	-.10	-.06
Absolute Assessment	-.28*	-.14	-.18	.19	.23*	.34**	-.13	.21
Written	-.01	-.06	.01	.17	-.03	-.04	-.16	-.12
Projects	.08	.02	-.14	.03	-.07	-.07	.04	-.05

* $p < .05$. *, $p < .01$ **

While it is likely bi-directional, an understanding of true causality among the three domains is not within the scope of this research. Results of this dissertation need

not suggest causality in order to aid in the design of and implementation of interventions for any of the three phenomena in question. In light of the cumbersome nature of their true causality and the dearth of research, this dissertation focused on trying to develop a prediction model that might suggest what aspects of assessment conceptions and assessment literacy explain the variance in assessment practice. Upon exclusion of all non-significant and small significant correlations ($r < .2$), stepwise multiple regression was employed to create a regression equation to help predict the standardized value of the significant assessment practice factors. Collinearity diagnostics were obtained to provide tolerance and variance inflation factor (VIF) scores for the variables.

As factors four and five did not have any significant Pearson correlations with any of the assessment conceptions or assessment literacy scores, they were not used in further analyses. Practice factor one, source of teacher assessment is external, was significantly correlated, $r = .37$, $p = .001$, with the assessment conception that assessment helps to hold schools accountable. This conception explained 13.6% of the variance ($R = .37$, $F(1,74) = 11.65$, $p < .001$) in the practice factor. The suggestion is that the conception that assessment holds schools accountable significantly predicted if they acquired their assessment from their district or course curriculum. This is reasonably interpreted as the degree to which teachers hold the conception that assessment holds schools accountable predicts some of the variance in their likely use of district or curriculum provided assessments.

Practice factor two, the use of tests and quizzes as a means of assessment, significantly correlated with all four assessment conceptions. It was positively correlated with the conceptions of holding schools and students accountable and aiding in student

improvement, and negatively correlated with the conception that assessment is irrelevant. Multiple regression analysis was used to test if the four conceptions of assessment would significantly predict the self-reported use of tests and quizzes. This four conceptions explained 19.4% of the variance ($R = .44$, $F(4,71)=4.27$, $p=.004$). Only the conceptions that assessment holds students accountable ($\beta=.33$, $p=.007$) and that assessment is irrelevant ($\beta=-.25$, $p=.043$) significantly predicted the self-reported use of tests and quizzes in the classroom. This suggests that the variance in teachers' use of tests and quizzes as a means of assessing students is predicted by the degree to which they hold the conception that assessment holds students accountable and inversely by the degree to which they believe assessment is irrelevant.

Practice factor three, absolute assessment, was significantly correlated with assessment literacy standard one, $r = .23$, $p = .042$, assessment literacy standard two, $r = .34$, $p = .003$, and negatively correlated with the conception that assessment holds schools accountable, $r = -.28$, $p = .015$. Multiple regression analysis was used to test if assessment literacy and the school accountability conception would significantly predict the assessment disengagement. This three variables explained 24.6% of the variance ($R = .50$, $F(3,72)=7.83$, $p<.001$). Assessment literacy for standard two ($\beta=.35$, $p<.002$), and the school accountability conception ($\beta=-.33$, $p=.002$) significantly predicted the degree to which teachers self-reported practices that indicate a profile of absolute assessment. This suggests that self-reported use of those assessment practices associated with disregarding individual ability, effort, and in class responses as well as using other teachers' assessments is predicted by a competency in the most difficult assessment literacy standard and inversely with the conception that assessment holds schools

accountable. The relationship between the practice factor and the school accountability suggests that the almost 25% of the variance that teachers engage in absolute assessment practices is based on the degree to which the teacher holds the conception that assessment holds schools accountable and their assessment literacy. More specifically, it suggests that the variance in absolute assessment practice is explained by the teachers' fluency and mastery with the language and ideas of assessment development. This suggests that this practice is likely not due to a lack of understanding or fluency in assessment, and may indicate a thoughtful choice in the selection of peers' assessments.

Question 3: What are the individual characteristics that describe teachers with different conceptions of assessment, teachers who self-report different assessment practices, and teachers who hold different levels of assessment literacy?

To better understand the differences in the specific characteristics of the participating teachers with respect to any of the assessment practice factors, the assessment conceptions, and partial or total assessment literacy, an analysis of variance (ANOVA) or correlation was conducted with the eleven demographic items. Only significant ANOVAs are reported and discussed. The two demographic variables dealing with years of experience were explored using a Pearson correlation. All significant correlations between the two measures of experience and any of the four assessment conceptions, the five practice factors, and assessment literacy are reported. The results from the ANOVAs are reported in the order in which the items are presented in the survey. There were no significant differences in any of the measured assessment variables with respect to the gender of the participant, whether the participant taught

within the United States, the participant's school affiliation, the participant's school district location, or if the participant was currently certified to teach in the United States, so they were excluded.

Student Grade Level Instructed

A one-way between subjects analysis of variance (ANOVA) was conducted to compare the effect of student grade level instructed on the first assessment practice factor, assessment from an external source. The ANOVA suggests a mean difference in the participants' score of self-reporting the use of the district's or school's assessments and not their own on the basis of the student grade level instructed [$F(3,72) = 10.91, p < .001, \eta^2 = .31$]. Post hoc comparisons using the Tukey HSD test indicated that the mean standardized score of elementary school teachers ($M = .81, SD = .16$) was significantly greater than that of middle school teachers ($M = -.20, SD = 1.09$) and of secondary school teachers ($M = -.35, SD = .74$). However, middle and secondary school teachers did not significantly differ. What this suggests is that teachers who instruct in elementary schools self-report using assessments that originate from their school or school district while not creating their own more than teachers in a middle school or secondary school setting. This does not suggest if the teachers are required to use prescribed assessments or choose to. It only suggests that they report using externally generated assessments more frequently.

Content Specialization

Two one-way between subjects analysis of variance (ANOVA) were conducted to compare the effect of content specialization instructed on the first, external source, and fourth, written and response, assessment practice factor. There was a significant effect of the content instructed on practice factor one, the participants' score of self-reporting the use of the districts' or schools' assessments and not their own, [$F(6, 66) = 3.35, p = .006, \eta^2 = .23$]. Post hoc comparisons using the Tukey HSD test indicated that the mean standardized score of English and Language Arts teachers ($M = -.70, SD = .62$) was significantly less than that of Special Education ($M = .47, SD = .84$). There were no other statistically significant differences among practice factor one. There was also a significant effect of content instructed on practice factor four, the use of written and in-class response for assessment, [$F(6, 66) = 2.61, p = .025, \eta^2 = .19$]. The post hoc comparisons using the Tukey HSD test indicated that the mean standardized score of English and Language Arts teachers ($M = .61, SD = 1.07$) was significantly greater than that of Special Education ($M = -.71, SD = .84$). Again, there were no other statistically significant differences among the practice factor.

These comparisons suggest that special education teachers report using district and school generated assessments more often than English and language arts teacher, who in turn use written work and in-class discussions as a means of assessment more often than special educators.

Formal Grading Policy

One-way ANOVAs were used to detect significant effects for the existence of a formal grading policy with respect to the assessment literacy score for standard one, choosing assessment effectively, $[F(2, 73) = 3.92, p = .024, \eta^2 = .10]$, the assessment conception that assessment holds schools accountable $[F(2, 73) = 4.35, p = .016, \eta^2 = .11]$, the assessment conception that assessment aids in student improvement $[F(2, 73) = 5.47, p = .006, \eta^2 = .13]$, and the assessment practice factor four, use of written and in-class discussion $[F(2, 73) = 4.12, p = .007, \eta^2 = .13]$. The third option was if they were unsure if such a policy existed. Teachers who reported not having a formal grading policy ($M = .76, SD = .23$), had greater assessment literacy for standard one than those teachers who did ($M = .61, SD = .22$). Additionally, teachers who reported not having a formal grading policy in their school scored statistically significantly higher than those teachers who did in both the conception that assessment holds school accountable and that assessment aids student improvement. Lastly, the practice factor of the use of written and in-class discussion as assessment was reported more frequently by teachers who did have formal grading policy in their school.

These comparisons suggest that teachers in schools without a formal grading policy hold the conceptions that assessment holds schools accountable and aids in student improvement, that they may be more assessment literate with respect to the selection of assessments, and that they may use written products for assessment purposes more than teachers at school with a formal grading policy. These are interesting findings in light of the many externally mandated policies and assessments some teachers report. It must also be disclosed that after the data collection took place, it was determined that the item

could have multiple interpretations and should have had further clarification. The item does not delineate what the grading policy should refer to at a minimum. Some respondents may have responded they did not have a formal grading policy because they interpreted the item to refer to a completely prescribed grading policy for all individual assessments and their summative combination. The intent for the item was to capture which respondents had any policies within any aspect of grading and assessing that they had to adhere to.

Education Level & Assessment Course

The post-secondary education level and the self-report if they took an assessment course were analyzed with a t-test to determine the equality of means, as there were only two groups. The education levels of the participants were regrouped into Bachelor of Arts or Science (BA/BS) and a Masters degree or higher to simplify interpretation. Teachers who reported obtaining a masters degree or above, ($M = .16$, $SD = .91$) had higher standardized scores on the absolute assessment practice factor than did teachers holding a BA or BS ($M = -.40$, $SD = .80$), $t(74) = 2.60$, $p = .011$, Cohens $d = .65$. Teachers who reported having taken an assessment course ($M = -.31$, $SD = .94$) had a lower standardized score on the fifth practice factor, use of projects, than did those teachers who did not ($M = .16$, $SD = .83$), $t(73) = 2.17$, $p = .034$, Cohens $d = .52$. Together these analyses suggest that a teacher's amount of post-secondary education will influence their assessment practices, but not in the same impact that way that having a specific course in assessment does.

A potential issue was detected with the item inquiring if a standalone course in assessment was taken. It is possible that some respondents, while having indicated that they did not take a standalone course in assessment, received assessment training in each of their pedagogically oriented courses. As the intent of the item was to capture those respondents with assessment training, the item should have been phrased accordingly.

Years of Experience

The demographic variables dealing with years of experience were explored using a Pearson correlation. The only significant relationships were suggested with the assessment literacy scores. Both measures of years of experience correlated significantly and negatively with the overall assessment literacy score. This sample suggests that lower assessment literacy scores are associated with more years of experience in either the teacher's current classroom or overall.

Table 4.5.

Pearson Correlation between Assessment Literacy and Teaching Experience

	Assessment Literacy Score
Experience in Years in Grade Level	-.33**
Experience in Years Total	-.46**
$p < .01^{**}$	

This may indicate a disconnect between the measure of assessment literacy and the practical experience of teaching in the classroom. It may also indicate that assessment literacy is not reinforced once the teacher leaves his or her degree program.

This would result in lower scores for teachers the further they are removed from their academic training.

CHAPTER 5

DISCUSSION

Summary of Findings

This study sought to understand the relationships between the assessment conceptions held, the assessment practices self-reported to be used and the degree of assessment literacy for classroom teachers in the hope that a more thorough understanding will enable more specific research in the future. The ultimate goal is the better understanding of the relationships between these assessment phenomena and how this understanding can most effectively instruct and remediate various teacher groups within the context of classroom assessment.

The teachers sampled held the assessment conceptions that assessment holds students accountable and improves student learning to a greater degree than they held the conceptions that it holds schools accountable or is irrelevant. The assessment practices that emerged are suggestive of a pattern of assessment behaviors indicating what kind of assessment is used and where the teacher obtains it. The assessment literacy scores are suggestive of only partial mastery of the standards of teacher competence in educational assessment of students.

This dissertation gives insight into the relationship between the three assessment phenomena by suggesting predictive factors that contribute to the self-reported assessment practice factors. The assessment practice of obtaining classroom assessment from external sources has a predictive relationship with the assessment conception that assessment holds schools accountable. Ultimately, holding the conception that schools are held accountable by classroom assessment accounts for some of the variance in the

use of assessments provided by the district or the curriculum. The practice of using quizzes and tests to assess student understanding and create grades can be partially predicted by the degree to which they think that assessment holds students accountable and the inverse of the degree they hold the conception that assessment is ultimately irrelevant. The third predictive relationship involves the practice factor indicating a pattern of absolute assessment. This use of assessment techniques that ignore relative methods of grade designation such as student perceived effort and individual ability can be partially predicted by the degree the teacher possesses a proficiency in assessment literacy and the inverse of assessment holds school accountable. This third predictive relationship suggests that the variance in the degree of these absolute assessment procedures is influenced by the very specific knowledge of assessment terminology and procedure, as well as the degree to which the teacher holds that assessment does not function to hold schools accountable.

The third aim of this study was to further identify the teachers who held the various assessment conceptions, reported the various practices, and demonstrated competency in within assessment literacy. Only a few of the demographic variables collected were suggestive of difference among any of the assessment variables in question. A logical continuation of this research would be to further measure the assessment phenomena in question with larger samples of specific teacher demographics so more statistical difference can be explored. This would also be necessary for the construction of effective assessment remediation programs to be developed.

Assessment Conception Scores

This dissertation sample held the conceptions that assessment holds students accountable and improves student learning to a greater degree than they hold the conceptions that assessment holds schools accountable or is irrelevant. The nature of these scores functioned as would be expected with a convenience sample of teachers. This study's findings echo the previous studies of this measure in that participant school accountability and improvement conception scores were higher than school accountability and irrelevant conception scores (Brown, 2011; Brown, Lake, & Matters, 2011). The correlations found between the four conceptions also suggest the measure functioned as would be expected. Without the adequate study sample size, a true confirmation of the model fit for the four conception model was not possible. In light of the inability to confirm the model stipulated by Brown (2006), for this sample, an exploratory factor analysis was conducted to suggest potential areas for further research.

While not the purview of this study, a principal axis exploratory factor analysis was conducted to explore differences in how these items functioned with respect to this study sample. The same procedures were used in the determination in the rotation of the factor solution and the number of factors retained as was used in the factor analysis for self-reported assessment practices. The goal was to determine potential differences that may exist in the measurement of the four assessment conceptions with this population. Factor analysis of the twenty-seven item Conceptions of Assessment III (TCoE-III A) Abridged Survey was conducted with the identical protocol suggested a three factor solution. The Kaiser-Meyer-Olkin Measure of sampling adequacy was .631. This exceeded the minimum threshold of .6. Three of the items had loadings greater than .4

on more than one factor and were ignored. The remaining twenty-four items suggest a similar factor structure to the one used in the study, but with three factors representing assessment conceptions, not four.

The new three-factor structure combines the items pertaining to the conception that assessment keeps students and schools accountable with a few items previously associated with aiding student improvement. The new general accountability conception encompasses the idea that assessment holds students and schools accountable. The key difference with respect to this sample of teachers is the inclusion of items pertaining to the inherent reliability and validity that assessment possesses. It suggests that this new conception may indicate a system of beliefs that assessments hold students and schools accountable in large part due to their level of inherent accuracy and consistency the teacher understands assessment to possess. The conceptions of aiding student improvement and assessment being irrelevant are largely intact.

While in need of further confirmation of the correct model structure with potentially new samples, the nature of the teachers' conceptions of assessment are still intact and suggestive of a significant aspect to understanding teacher classroom assessment. There is clear evidence in prior research as well as this dissertation that teachers do not hold one conception of the purpose of assessment, but instead have varying degrees of several different conceptions. These conception values can be used in the understanding of the other assessment phenomena as well as potential remediation of assessment practices in the classroom.

Assessment Practice Factors

The determination of potential practice factors was required within this study as there was no measure of assessment practice that would yield scores of underlying practice factors. This study wanted assessment practices to be represented quantifiably, so a model of prediction might be suggested. Despite the need for more participants to better account for sample size inadequacy, the procedures used in factor extraction revealed several practice factors that made logical sense with respect to the likely practice profiles that exist within teachers. Practice factors two, four and five suggest discrete sets of assessment practice behaviors centering on the use of quizzes and tests, written products, and projects. These practice factors are helpful in the understanding of how these assessment practices are likely to group together.

One of the aims of this research was to explore assessment practices beyond the straightforward designation of what specific assessment practices are employed. Practice factors one and three do this by exploring the source of the assessments and what the teachers consider when scoring and interpreting assessment results. Practice factor one gives a clear indication that an underlying practice factor suggests the degree to which the educators obtain their assessments from outside sources. In this case the outside sources are the district and the curriculum. When this is paired with the practice of not creating their own assessments, this practice factor further suggests a practice of looking externally for the assessments used in class. Practice factor four is similarly enlightening as it suggests a profile of using other teachers' assessment as well as discounting relative aspects of grading. This factor indicates the degree to which the assessment score

ignores or includes individual ability and effort. For this reason, the factor is indicative of an absolute grading profile.

Future research could explore why a teacher reports a reliance on absolute assessment practices. In addition to the simple belief that classroom assessment results should be an indication of mastery of educational objectives and be consistently applied to all students, it is possible that the practice factor also represents a level of difficult assessment avoidance. The practices of trying to quantify and account for student perceived effort and individual ability is time consuming and subjective. This practice might be avoided by some teachers as a means of focusing their time and effort on other aspects of instruction.

Ultimately, in light of the negative correlation with the first practice factor and subsequent prediction models, it is my interpretation that it is more suggestive of a practice of assessment focused on the use of absolute methods of assessment.

Assessment Literacy

The Assessment Literacy Inventory (ALI) (Mertler & Campbell, 2005) is still being used as a metric to interpret the teachers' competency with the standards for teacher competence in educational assessment (Delosa & Morales, 2015). It is, however, reflective of a very specific domain of content. Of the practice factors that emerged, only one of the factors had a predictive relationship with the any degree of assessment literacy. While the items are adequately suggestive of a partial mastery of these items with current teachers, this partial mastery may indicate little else besides the fluency in assessment language and the application of its meaning. It is reasonable to suggest that the training

and remediation of assessment literacy may not assist in the alterations and changes to the teachers' classroom assessment practices. It may simply be an indication of how far removed the teacher is from her formal pedagogical instruction, as the years of experience suggests.

Assessment Phenomena Relationships

The predictive relationship between teachers' assessment conceptions and assessment literacy with their assessment practices was a primary focus of this research. The practice of using external assessments was partially predicted by the conception that schools are held accountable by assessment. With only 13% of the variance being explained, this may suggest that a relatively small amount of the variance in this practice is accounted for with this conception. With so much unexplained variance, future research should endeavor to determine what accounts for other aspects of the variance in this score.

Nearly 20% of the variance in the use of tests and quizzes was accounted for in the conceptions that students are held accountable by assessment and then inversely by the conception that assessment is irrelevant. This makes reasonable sense as teachers would not engage in the practice of giving tests and quizzes as a means of both grading and obtaining student understanding if they did not hold the conception that this practice was holding the student accountable and was not irrelevant. Ideally, a relationship between the variance in use of any specific assessment practice should be explained by the conception that assessment improves student learning. Future research should aim to suggest the conceptions responsible for other specific assessment practices.

Nearly 25% of the variance in the practice of absolute assessment is accounted for by one assessment literacy standard and inversely the conception that assessment holds school accountable. To have this much variance in a practice factor, which is suggestive of using other teacher's assessments, explained by the score of three items dealing with assessment development is interesting. This may indicate that these teachers know the procedure of assessment development, but choose not to engage in it. The variance is also explained inversely by the conception of school accountability. Perhaps the combination of these teachers' conception that assessment infrequently holds schools accountable with their proficiency in assessment literacy has made them want to find absolute measures of achievement and use others assessment. This relationship is complex and in need of further clarification.

Demographics Variable Frequency

Summatively, there is not a clear indication that a particular demographic of teachers is different with respect to the three assessment phenomena addressed in this dissertation. The statistical differences among the demographic variables were predominantly within the domain of self-reported assessment practice. This dissertation suggests that elementary schoolteachers are different with respect to the first assessment practice factor, external sources for assessment. A more robust sample of teachers would likely suggest more of these differences and significantly contribute to the ways these differences can guide remediation. Similarly, teachers engaged in special education were only significantly different from English and language arts teachers. This again seems an

example were a more robust sample would help further clarify the differences in the ways these groups of teachers engage in the domains of assessment.

This dissertation also suggests that the implementation of a formal grading policy has implications with respect to all three of the assessment domains investigated. Future research should attempt to understand more fully the specific parameters of these policies, as well as the ramifications of non-compliance. With respect to the amount of post-secondary education and if participants recalled taking a course in assessment, again these differences manifested themselves in the teachers' assessment practice. This finding is reflected of more recent research suggesting that some assessment practices are best predicted by the occurrence of particular teacher assessment training (Koloi-Keaikitse, 2016).

Lastly, this dissertation is helpful in what demographic variables may be excluded from future research as well. This would help to lesson the item load on the participants. Ultimately, this study demonstrated the need for future research that would not only adequately describe the various assessment phenomena with demographic frequencies, but also describe different demographic groups with levels of assessment phenomena.

Intended Audience

The intended audience for this research could be nearly all those affected by student assessment. A primary audience would be school administrators who wish to both understand and support their teachers better in the domain of student assessment. While the individual practices, assessment conceptions, and assessment literacy may vary with other teacher populations, the likelihood of influence between them makes this

research very important for those in a position to monitor and support classroom teachers. Teachers and instructors of pre-service teachers would also benefit from this research in the context of self-analysis and reflection. This research may prompt a thoughtful review of a teachers assessment practices and conceptions that help put the focus of their assessment on the students the assessments are meant to measure. In addition to the teachers themselves, it is also the intent that parents may find this research impactful with respect to understanding their own student's assessment results. This research may facilitate a more meaningful dialogue between the parent and teacher, by giving the parent a more knowledgeable base from which to inquire about their own child's assessment.

Limitations and Future Research

The main limitation of this study is with respect to the sample size and sampling procedure of the study population. The sampling constraints of this research limit my ability to suggest the true nature of the practice factors that emerged, and thus the likely predictive relationship between the three assessment phenomena. In addition to an increased sample size, this study is limited to anecdotal observations about the types of teachers who possess the varying levels of assessment phenomena. This would not affect the ability to suggest the relationships between these phenomena, but to suggest statistically significant differences among demographic groups. This limitation is particularly important with the use of these results in the assisting of current in- service teachers with their current classroom assessment needs. Despite the desire for a random sample of teachers, it is unlikely that teachers who do not find this aspect of teaching

interesting would be unlikely to complete the measure. This limitation of a convenient or biased sample of teachers is likely to persist.

Another limitation of this research was the respondent fatigue with respect to the time required in taking the online survey. As stated previously, the measure for assessment literacy is likely to have been used differently. Ultimately, it seems prudent to divide the assessment measures into multiple administrations.

Future research requires first an adequate confirmation of the model of assessment conceptions and of the assessment practice factors found in this study. With confirmation, the prediction models suggesting what phenomena may account for the variance in assessment practice can be used effectively to challenge and improve current classroom assessment.

REFERENCES

- Agee, D. (1991). Double-barrelled assessment: Teachers & students as partners. *Adult Learning, 2*(7), 7.
- Anderson, L. W. (2003). *Classroom assessment :Enhancing the quality of teacher decision making*. Mahwah, N.J.: L. Erlbaum Associates.
- Bachor, D. G., & Anderson, J. O. (1994). Elementary teachers' assessment practices as observed in the province of british columbia. *Assessment in Education: Principles, Policy & Practice, 1*(1), 63.
- Baird, J. (2010). Beliefs and practice in teacher assessment. *Assessment in Education: Principles, Policy & Practice, 17*(1), 1-5.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice, 17*(2), 215-232.
- Brew, C., Riley, P., & Walta, C. (2009). Education students and their teachers: Comparing views on participative assessment practices. *Assessment & Evaluation in Higher Education, 34*(6), 641-657.
- Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, N.J.: Pearson Merrill Prentice Hall.
- Brown, G. L., & Hirschfeld, G. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment In Education: Principles, Policy & Practice, 15*(1), 3-17.
- Brown, G. L., Kennedy, K. J., Fok, P., Jacqueline Kin Sang, C., & Yu, W. (2009). Assessment for student improvement: understanding Hong Kong teachers'

- conceptions and practices of assessment. *Assessment In Education: Principles, Policy & Practice*, 16(3), 347-363.
- Brown, G. L., Lake, R., & Matters, G. (2009). Assessment Policy and Practice Effects on New Zealand and Queensland Teachers' Conceptions of Teaching. *Journal Of Education For Teaching: International Research And Pedagogy*, 35(1), 61-75.
- Brown, G. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching & Teacher Education*, 27(1), 210-220.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological Reports*, 99(1), 166-170.
- Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, 3
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive–informal assessment practices: New zealand secondary students' conceptions of assessment. *Learning & Instruction*, 19(2), 97-111.
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching & Teacher Education*, 27(1), 210-220.
- Brown, G., & Michaelides, M. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal Of*

Psychology Of Education - EJPE (Springer Science & Business Media B.V.), 26(3), 319-337.

Brown, G.T.L. (2002). Teachers Conceptions of Assessment (Doctoral Dissertation). The University of Auckland, Auckland, NZ.

Buhagiar, M. A., & Murphy, R. (2008). Teachers' assessments of students' learning of mathematics. *Assessment in Education: Principles, Policy & Practice*, 15(2), 169-182.

BULLOUGH, J., ROBERT V., CLARK, D. C., & PATTERSON, R. S. (2003). Getting in step: Accountability, accreditation and the standardization of teacher education in the united states. *Journal of Education for Teaching*, 29(1), 35.

Butler, S. M., & McMunn, N. D. (2006). *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning* Jossey-Bass, An Imprint of Wiley.

Chow, A. (2010). Improving Learning through Student-involved Assessment. *International Journal of Learning*, 16(12), 87-101.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159.

Coffey, J., Sato, M., & Thiebault, M. (2005). Classroom Assessment: Up Close and Personal. *Teacher Development*, 9(2), 169-184.

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation*, 101-9.

- Davis, D. S., & Neitzel, C. (2011). A Self-Regulated Learning Perspective on Middle Grades Classroom Assessment. *Journal of Educational Research*, 104(3), 202-215.
- Dekker, T., & Feijs, E. (2005). Scaling up strategies for change: change in formative assessment practices. *Assessment in Education: Principles, Policy & Practice*, 12(3), 237-254.
- Delandshere, G. (1994). The assessment of teachers in the united states. *Assessment in Education: Principles, Policy & Practice*, 1(1), 95.
- Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record*, 104(7), 1461-1484.
- Delosa, J. j., & Morales, K. k. (2015). Assessment Literacy of Pre-Service Teachers and the NCBTS Domain. *IAMURE International Journal Of Education*, 131-12.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment In Education: Principles, Policy & Practice*, 17(4), 419-438.
- Draper, S. W. (2009). Catalytic assessment: Understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology*, 40(2), 285-293.
- Field, A. (2005). *Discovering statistics using SPSS, 2nd edition*. London England: Sage.
- Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics*, 18(3), 402-423.
- Gardner, J. (2010). *Developing teacher assessment*. Maidenhead; New York: Open University Press.
- Goodenough, W. H. (1990). Evolution of the Human Capacity for Beliefs. *American Anthropologist*, 93, 597-612.

- Gorsuch, R.L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, 25, 33-39.
- Green, S. K., Johnson, R. L., Kim, D., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching & Teacher Education*, 23(7), 999-1011.
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365-381.
- Hines, S. K. (2009). Investigating faculty development program assessment practices: What's being done and how can it be improved? *Journal of Faculty Development*, 23(3), 5-19.
- Impara, J. C., & Plake, B. S. (1995). Comparing counselors', school administrators', and teachers' knowledge in student assessment. *Measurement & Evaluation in Counseling & Development (American Counseling Association)*, 28(2), 78.
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Teachers' Assessment Background and Attitudes Toward Testing. *Theory Into Practice*, 32(2), 113.
- Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of english as a foreign language. *Assessment in Education: Principles, Policy & Practice*, 16(2), 185-204.
- James, M., & Pedder, D. (2006). Beyond method: assessment and learning practices and values. *Curriculum Journal*, 17(2), 109-138.

- Jett, D. L., & Schafer, W. D. (1993). Ready or not, teachers K-12 move to center stage in the assessment arena: Implications for state.. *Measurement & Evaluation in Counseling & Development (American Counseling Association)*, 26(1), 69.
- Keith, T. Z. (2005). *Multiple regression and beyond*. Boston, MA: Allyn & Bacon.
- Kennedy, M. M., & American Association of Colleges for Teacher Education. (2010). *Teacher assessment and the quest for teacher quality :A handbook* (1st ed.). San Francisco: Jossey-Bass.
- Kim, J.-O., & Mueller, C.W. (1978). *Introduction to factor analysis: What it is and how to do it*. Newbury Park: Sage.
- Kim, J.-O., & Mueller, C.W. (1978). *Factor analysis: Statistical methods and practical issues*. Newbury Park: Sage.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling, 3rd edition*. New York: The Guilford Press.
- Koloi-Keaikitse, S. (2016). Assessment training: a precondition for teachers' competencies and use of classroom assessment practices. *International Journal Of Training & Development*, 20(2), 107-123.
- Lawrance, H. I. (2010). One mathematics teacher's thoughts on assessment. *Mathematics Teacher*, 104(1), 12-13.
- Marton, F. (1981). Phenomenography – Describing conceptions of the world around us. *Instructional Science*, 10, 177–200.
- McDonald, B. (2007). Self-assessment for understanding. *Journal of Education*, 188(1), 25-40.

- McNair, S., Bhargava, A., Adams, L., Edgerton, S., & Kypros, B. (2003). Teachers speak out on assessment practices. *Early Childhood Education Journal*, 31(1), 23.
- Mertler, C. A. (2005a). Secondary Teachers' Assessment Literacy: Does Classroom Experience Make a Difference?. *American Secondary Education*, 33(2), 76-92.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-113.
- Mertler, C. A. (2010). Teachers' perceptions of the influence of no child left behind on classroom practices. *Current Issues in Education*, 13(3), 1-34.
- Mertler, C. A. (2010). Teachers' perceptions of the influence of no child left behind on classroom practices. *Current Issues in Education*, 13(3), 1-34.
- Mertler, C. A. (2010). Teachers' perceptions of the influence of no child left behind on classroom practices. *Current Issues in Education*, 13(3), 1-34.
- Mertler, C. A., & Campbell, C. (2005 b). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the "assessment literacy inventory"* Online Submission.
- Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urda, T. (2000). *Patterns of adaptive learning survey (PALS)*. University of Michigan.
- Miller, M.D., Linn, R. L. & Gronlund, N.E. (2012). *Measurement and Assessment in Teaching* (11th Ed.). Upper Saddle River, NJ: Pearson Education, Inc.

- Nan, H., Braden, J., White, J., & Elliott, S. (2006). Effect of an Internet-Based Professional Development Program on Teachers' Assessment Literacy for All Students. *Teacher Education & Special Education*, 29(4), 244-260.
- Nartgün, Z. (2009). Student views on the assessment practices of instructors during instruction. *Educational Sciences: Theory & Practice*, 9(4), 1807-1818.
- Nartgün, Z. (2009). Student Views on the Assessment Practices of Instructors during Instruction. *Educational Sciences: Theory & Practice*, 9(4), 1807-1818.
- National Council on Measurement in Education, American Federation of Teachers, & National Education Association (NCME, AFT, NEA). (1990). *Standards for teacher competence in educational assessment of students*
- Nolen, S. B., Horn, I. S., Ward, C. J., & Childers, S. A. (2011). Novice teacher learning and motivation across contexts: Assessment tools as boundary objects. *Cognition & Instruction*, 29(1), 88-122.
- O'Sullivan, R. G., & Johnson, R. L. (1993). *Using performance assessments to measure teachers' competence in classroom assessment*
- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication Analysis in Exploratory Factor Analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation*, 17(14/15), 1-8.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62, 307-332.
- Pellegrino, J. W., Chudowsky, N., Glaser, R., & National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

- Perrin, G. (2005). Teachers, testers, and the research enterprise--a slow meeting of minds. *ELT Journal: English Language Teachers Journal*, 59(2), 144-150.
- Peterman, F. (2005). *Designing performance assessment systems for urban teacher preparation*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Philippou, G., & Christou, C. (1997). Cypriot and Greek Primary Teachers' Conceptions about Mathematical Assessment. *Educational Research & Evaluation*, 3(2), 140.
- Plake, B. S., & Impara, J. C. (1993a). Teacher Assessment Literacy: Development of Training Modules. Report on an NCME-Based Kellogg Foundation Grant.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993b). Assessment competencies of teachers: a national survey. *Educational Measurement: Issues & Practice*, 12(4), 10-12, 39
- Plake, B.S. & Impara, J.C., (1997) Teacher Assessment Literacy: What do Teachers Know about Assessment?, In Phye, Gary D. (Ed), *Handbook of classroom assessment : learning, achievement, and adjustment* (pp. 53-68), San Diego : Academic Press.
- Ploegh, K., Tillema, H., & Segers, M. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation*, 35(2/3), 102-109.
- Popham, W. (2009). Assessment Literacy for Teachers: Faddish or Fundamental?. *Theory Into Practice*, 48(1), 4-11.
- Popham, W. J. (2003). *What every teacher should know about educational assessment*. Boston, Mass.: Pearson Education.
- Popham, W. J., & California. (1971). *Designing teacher evaluation systems;a series of suggestions for establishing teacher assessment procedures as required by the stull*

bill (AB 293), 1971 california legislature. Los Angeles: Instructional Objectives Exchange.

Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly*, 42(4), 203 - 220.

Preacher, K.J., & MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13-43

Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching & Teacher Education*, 27(2), 472-482.

Riggan, M., & Oláh, L. N. (2011). Locating interim assessments within teachers' assessment practice. *Educational Assessment*, 16(1), 1-14.

Ruiz-Primo, M., & Furtak, E. (2006). Informal Formative Assessment and Scientific Inquiry: Exploring Teachers' Practices and Student Learning. *Educational Assessment*, 11(3/4), 205-235.

Sammons, P., Day, C., Kington, A., Gu, Q., Stobart, G., & Smees, R. (2007). Exploring variations in teachers' work, lives and their effects on pupils: key findings and implications from a longitudinal mixed-method study. *British Educational Research Journal*, 33(5), 681-701.

Schafer, W. D. (1993). Assessment Literacy for Teachers. *Theory Into Practice*, 32(2), 118.

Schönrock-Adema, J., Heijne-Penninga, M., van Hell, E. A., & Cohen-Schotanus, J. (2009). Necessary steps in factor analysis: Enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Medical Teacher*, 31(6), 226-232. doi:10.1080/01421590802516756

- Schwager, M. T., & Carlson, J. S. (1994). Building assessment cultures. *Education & Urban Society*, 26(4), 390.
- Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment?. *Studies In Educational Evaluation*, 37(1), 49-54.
- Shulman, L. S. (1987). Assessment for teaching: An initiative for the profession. *Phi Delta Kappan*, 69(1), 38-44.
- Snook, S.C., & Gorsuch, R.L. (1989).Principal component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, 106, 148-154.
- Soled, S. W. (1995). *Assessment, testing, and evaluation in teacher education*. Norwood, N.J.: Ablex Pub. Corp.
- Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan*, 69(5), 363-368.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77, 238-245.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues & Practice*, 18(1), 23-27.
- Stones, E. (1994). Assessment of a complex skill: Improving teacher education. *Assessment in Education: Principles, Policy & Practice*, 1(2), 235.
- Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice*, 17(4), 399-417.

- Sylvia Yee, F. T., Pamela Pui, W. L., Alice Wai, K. C., & Ping, M. W. (2010). A case study of teacher learning in an assessment for learning project in hong kong. *Professional Development in Education*, 36(4), 621-636.
- Tabachnick, B.G., & Fidell, L.S. (2000). *Using multivariate statistics*(4thEd.). New York: Harper-Collins
- Taylor, C. S., & Nolen, S. B. (2008). *Classroom assessment :Supporting teaching and learning in real classrooms* (2nd ed.). Upper Saddle River, N.J.: Pearson/Merrill/Prentice Hall.
- Tileston, D. W. (2004). *What every teacher should know about student assessment*. Thousand Oaks, Calif.; London: Corwin Press.
- Wang, J., Kao, H., & Lin, S. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching & Teacher Education*, 26(3), 522-529.
- Webb, M., & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 16(2), 165-184.
- Winterbottom, M., Brindley, S., Taber, K., Fisher, L., Finney, J., & Riga, F. (2008). Conceptions of assessment: trainee teachers' practice and values. *Curriculum Journal*, 19(3), 193-213.
- Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives*, 18(19), 1-37.
- Ysseldyke, J. E. (1987). Do tests help in teaching? *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 28(1), 21-24.

- Ysseldyke, J. E., & Algozzine, R. (2006). *Effective assessment for students with special needs :A practical guide for every teacher*. Thousand Oaks, Calif.: Corwin Press.
- Zhang, Z., & Burry-Stock, J. (1995). *A multivariate analysis of teachers' perceived assessment competency as a function of measurement*

APPENDIX A
PERSONAL SOLICITATION LETTER

Dear K-12 Educator,

I would like to ask for your participation in my current research of classroom teacher assessment phenomena by completing an online survey of classroom assessment. I have a great appreciation for your time and as a teacher myself, I would not ask you to participate lightly. This research is of great significance to me both personally and professionally, and responses from educators such as you will help to facilitate a meaningful discourse and ultimately lead to a greater understanding of how to cater to our needs as educators.

The survey link below will take you to the online survey and should take approximately 20 minutes to complete.

https://www.surveymonkey.com/s/Snyder_SofAPC_L

The purpose of this study is to investigate the specific assessment phenomena of teacher held assessment conceptions, assessment literacy and self-reported assessment practices. Information about these phenomena individually and with respect to each other would help to inform our understanding of classroom assessment and ultimately aid in its remediation and development.

Your responses will remain confidential and will be gathered anonymously.

Please make sure to follow the link after the survey to enter an e-mail address for a chance to win one of five (5) \$25 Amazon gift cards.

With Great Thanks,

Mark Snyder

mrsnyder@temple.edu

(215) 680-2770

Ritter Hall 464

1301 Cecil B. Moore Ave.

Philadelphia, PA 19122

APPENDIX B

SCHOOL PARTICIPATION LETTER

Study Title: Investigating the Link between Current Classroom Teachers' Conceptions, Literacy, and Practices of Assessment

Student Investigator: Mark R. Snyder, M.Ed., Doctoral (PhD) Student, Temple University

Primary Investigator, Advisor: Julie L. Booth, Ph.D., Assistant Professor of Educational Psychology, Temple University

Purpose of Study:

Since the publication of *A Nation at Risk: The imperative for Educational Reform* in 1983, academic assessment went through a fundamental change within the educational system. In light of the report's assertion that many students were being failed by their educational system and recommendation that states submit report cards demonstrating achievement levels and progress, academic assessment began a shift from a tool to determine educational progress of specific students to its use as a barometer by which educational systems could be judged (Miller, Linn & Gronlund, 2012). This shift in assessment's role in the classroom created an assessment culture where both greater import and greater scrutiny were placed on all forms of assessment in the classroom.

The necessity for research into the classroom assessment domain is predicated upon several presuppositions. While there is debate about the method and form of assessment that should be used with various students (Alkharusi, 2008), there is consensus that the assessment practices teachers employ do have an impact on their student's achievement. (Brown & Hirschfeld, 2008; Segars & Tillema, 2011) Given the impact on student achievement, the individual assessment choices teachers make within their classroom and the underlying reasons these choices are made are worth study. It is also clear that in addition to assessment's role in ultimate student achievement, the overall negative student views of the assessment these students encounter in the classroom (Nartgun, 2009) and the many teachers' low initial self-reported knowledge of and self-efficacy within the domain of classroom assessment (Dekker & Feijs, 2005; Nan et al., 2006) suggest the need for continued research.

As there is no indication that enacting teacher change in the domain of assessment differs significantly from other domains, (Schwager & Carlson, 1994) research should inform the current values, attitudes and practices of the teachers in question, prior to suggesting the method or form of intervention. Schwager and Carlson conclude that there are two components to school change, the scholastic environment in which the teacher instructs, which consists of the degree of support and innovation, and the attitudes and beliefs of the teachers themselves. This research will attempt to further clarify the later.

Understanding this culture of assessment that has been cultivated in the thirty years since *A Nation at Risk*'s publication is impossible without a clear understanding of the teachers within this culture. This dissertation will focus on three distinct phenomena within the assessment culture: the conceptions teachers hold about the assessments that take place in their classroom and school, the assessment practices these teachers choose to employ, and the degree of proficiency with assessment (or, assessment literacy) that teachers possess. This dissertation is an attempt to not only understand what these three phenomena look like in teachers with varying content proficiency and age of students instructed, but to investigate the link between the phenomena as well.

This dissertation will use teacher survey data to suggest a set of self-reported practice profiles that should help describe the specific assessment practices that often are employed together. Further, it will identify correlations between these practice profiles and teachers' assessment conceptions and assessment literacy. Ultimately, the research is designed to suggest predictive relationships between the assessment conceptions and literacy teachers might possess and the kinds of assessment practices that they report to employ.

Given the import that assessment has in the classroom, if the educational community was aware of what teachers' perceptions, practices, and literacy were with respect to assessment we could engage teachers more effectively. As these phenomena do not operate in isolation, our understanding of them should not be similarly limited. Using the relationships between these phenomena and the knowledge of what demographic characteristics are likely with respect to them in both high and low degree would enable teacher educators to tailor both curriculum and professional development. Gathering this information is the first step by which appropriate professional development and interventions can be designed and implemented in the assessment domain.

The following research questions will be addressed in this study:

1. What are the nature of current teachers' assessment conceptions, assessment practices and assessment literacy?
2. What are the relationships among teacher conceptions of assessment, teacher assessment literacy, and assessment practices?
3. What individual differences in demographics and experience characterize teachers with different conceptions of assessment, teachers that self-report different assessment practices and teachers who hold different levels of assessment literacy?

Procedure: Data collection will include the administration of a one-time questionnaire taking approximately 20 minutes. The questionnaire will include questions about teachers' 1) demographic backgrounds 2) conceptions of assessment (e.g., assessment is irrelevant, assessment aids student learning, assessment keeps schools accountable, etc.), 3) assessment literacy (i.e., recognizing and evaluating sound assessment practices), and 4) current reported

assessment practices (i.e., self-report of the frequency and degree to which various assessment practices are utilized in the teacher's classroom).

The questionnaire will be administered either in person or online via the website SurveyMonkey.

https://www.surveymonkey.com/s/Snyder_SofAPC_L

Request for Participation: I am requesting to administer the questionnaire to any K-12 teachers. This could be done during pre-existing K-12 professional development, during in-service meetings at the schools in your district, or on the teachers own time after the survey link is made available to them.

No in-class observation of teachers is required. All participants will read a confidentiality statement per Temple University IRB guidelines. This study will not assess or analyze individual teacher, school, or school district performance or ability. If desired by the school, I am happy to provide a brief talk/discussion on classroom assessment topics as a 'thank you' for their cooperation.

Contact information: Mark Snyder

mrsnyder@temple.edu

(215) 680-2770

APPENDIX C
SNYDER SURVEY OF ASSESSMENT PRACTICES,
CONCEPTIONS, AND LITERACY

GENERAL DIRECTIONS: Thank you for taking part in my research study. The questions here are for background information only. All of your responses will be kept in strict confidence. Under no circumstances will you be identified nor will your responses be shared with anyone. If you have any questions, please contact Mark Snyder at Temple University at 215-680-2770

Your responses are very important to me. We thank you in advance for your cooperation.

Demographic / Background Info

1. What is your gender?
 - ☐ Female
 - ☐ Male
2. Do you currently teach in the United States?
 - ☐ Yes
 - ☐ No
3. Which of the following is the **most appropriate** description of the level of student you teach? (Select all that apply)
 - ☐ Elementary (K – 6)
 - ☐ Middle (grades 6 – 8)
 - ☐ Secondary (grades 9 – 12)
 - ☐ Other
4. If you have specific content specialization, which of the following most appropriately identifies your content area? (Select all that apply)
 - ☐ English / Language Arts
 - ☐ Mathematics
 - ☐ Social Studies / History
 - ☐ Science
 - ☐ Foreign Language
 - ☐ Special Education
 - ☐ Physical Education
 - ☐ Elementary Math
 - ☐ Elementary Language Arts
 - ☐ Other
5. Which **best** describes the educational level you have attained?

<input type="checkbox"/> B.A. or B.S.	<input type="checkbox"/> Ed.D
<input type="checkbox"/> M.A. or M.S.	<input type="checkbox"/> Ph.D

6. Which **best** describes your current school affiliation?

- ☐ Public
- ☐ Parochial
- ☐ Charter
- ☐ Other

7. Which term best describes the location of your school district?

- ☐ Urban
- ☐ Suburban
- ☐ Rural
- ☐ Other

8. **Including the current year**, how many years of experience do you have?

_____ Years in Current Grade Level

_____ Years in Teaching in Total

9. To the best of your knowledge, did you take a **standalone** course in classroom assessment as part of your undergraduate or graduate training?

- ☐ No
- ☐ Yes

10. To the best of your knowledge, do you currently hold state certification in your current teaching discipline?

- ☐ No
- ☐ Yes

11. Does your district have a formal policy for determining and/or assigning grades?

- ☐ No
- ☐ Yes
- ☐ Not sure

Assessment Literacy: Please circle the response (A – D) that answers each question.

Scenario #1

Mr. Okawa, a fifth-grade teacher, is planning his instruction for the next grading period, aware of the fact that his students will be taking the statewide achievement test near the end of the grading period.

1. Mr. Okawa's mathematics unit for this grading period will focus on multi-step problem-solving. He wants to assess his students' problem-solving abilities at the end of the unit to determine if any reinstruction will be necessary prior to the statewide test. Which of the following assessment strategies would be the most appropriate choice?
 - A. He should choose the assessment included in the teacher's manual from the textbook he uses.
 - B. He should choose an assessment which is consistent with the content and skills he taught.
 - C. He should choose a different standardized assessment that provides a score on similar skills.
 - D. He should choose an assessment which covers single-step problem-solving skills.

2. Mr. Okawa decides to develop his own assessment in order to determine if any reinstruction will be necessary. He also wants to use his assessment as a means of anticipating how his students will perform on the statewide assessment. In order for him to accurately approximate his students' performance, which of the following would be the most appropriate type of assessment for him to develop?
 - A. a performance assessment
 - B. a multiple-choice test
 - C. a portfolio assessment
 - D. an essay test

3. Juan, another student in Mr. Okawa's class, receives a scaled score of 196 on the reading comprehension portion of the statewide assessment. The cut score is 200; therefore, Juan does not pass this subtest. However, the subtest has a standard error of measurement equal to 6. Which of the following is the best decision for Mr. Okawa to make regarding instruction appropriate to meet Juan's needs?
 - A. Juan has clearly not achieved the minimum level of reading comprehension and should receive remedial reading instruction.
 - B. Mr. Okawa knows that Juan could have scored higher, so the results of the test should be ignored.
 - C. Juan may likely have achieved the minimum level of reading comprehension and nothing different or additional should be done.
 - D. Mr. Okawa knows that Juan should have scored much lower, so the results of the test should be ignored.

Reported Assessment Practices

Assessment Practices	Never					A lot
	1	2	3	4	5	6

How often do you use the following to get a sense of students' understanding:

1. Paper-and-pencil tests	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Quizzes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. In class assignments	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Homework performance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Projects / Presentations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Responses of students in class	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Written work (essays, papers, etc)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How often do you use the following to help determine a report card grade:

8. Paper-and-pencil tests	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Quizzes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. In class assignments	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Homework performance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Projects / Presentations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Responses of students in class	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Written work (essays, papers, etc)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please tick one box for each

Assessment Practices	Never	Once per Month	Once per 2 Weeks	Once per Week	3 / Week	Once a Day+
	(1)	(2)	(3)	(4)	(5)	(6)
15. About how often do you give <u>minor</u> assessments?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. About how often do you give <u>major</u> assessments?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Assessment Practices	Not at All					A Lot
	1	2	3	4	5	6

To what degree do you consider the following when assigning and/or adjusting grades on assignments, tests, etc:

17. Percentage or number correct (e.g. 100-94 = A, 90-93 = A-)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. The difficulty of the test, assignment, etc.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. How the class as a whole performed on the assignment, test, etc	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. The individual students' ability level	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Effort (i.e., how hard the student tried on the test, assignment, etc)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

To what degree do you consider the following when combining grades from assignments, quizzes, tests, etc., into a final grade:

22. Average all of the marks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Throw out some of a student's low scores and average the rest of the marks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Weigh major assignments more heavily, then average the marks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Assessment Practices		Never 1	2	3	4	5	A Lot 6
What is the frequency of the following in terms of the source of your <u>minor</u> assignments?							
25.	I develop my own minor assignments.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26.	I use assignments created by another teacher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27.	I use assignments provide with curricular materials.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28.	I use assignments developed by the district.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
What is the frequency of the following in terms of the source of your <u>major</u> assignments?							
29.	I develop my own major assignments.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30.	I use assignments created by another teacher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31.	I use assignments provide with curricular materials.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32.	I use assignments developed by the district.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Assessment Literacy: Please circle the response (A – D) that answers each question.

Scenario #2

Ms. Green is an eighth-grade American History teacher. She has just finished teaching a unit on the Industrial Revolution and wishes to make decisions about her students regarding their higher-order thinking skills. Ms. Green has decided to give her students a single assessment in the form of an end-of-unit multiple-choice test. She anticipates that most of her students will perform well on the test

4. Based on her goal, what can you conclude about her decision to administer a multiple-choice test?
 - A. This is an appropriate choice for a unit assessment.
 - B. The test scores may not be valid for this purpose.
 - C. The test scores may not be reliable for this purpose.
 - D. A true-false test would be more appropriate.

5. To determine the quality of her multiple-choice test, Ms. Green should conduct an item analysis and examine all of the following **except**
 - A. item difficulty values.
 - B. item discrimination values.
 - C. reliability coefficients.
 - D. validity coefficients.

6. Some of Ms. Green's students do not score well on the multiple-choice test. She decides that the next time she teaches this unit, she will begin by administering a pretest to check for students' prerequisite knowledge. She will then adjust her instruction based on the pretest results. What type of information is Ms. Green using?
 - A. norm-referenced information
 - B. criterion-referenced information
 - C. both norm- and criterion-referenced information
 - D. neither norm- nor criterion-referenced information

Conceptions of Assessment

- This survey asks about your conceptions and understandings about ASSESSMENT, whatever that term means to you. Please answer the questions using YOUR OWN understanding of classroom 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. 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 1	2	3	4	5	Strongly Agree 6
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"/>	<input type="checkbox"/>
2. Assessment places students into categories	<input type="checkbox"/>	<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"/>	<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"/>	<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"/>	<input type="checkbox"/>
6. Assessment results are trustworthy	<input type="checkbox"/>	<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 against their beliefs	<input type="checkbox"/>	<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"/>	<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"/>	<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"/>	<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"/>	<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"/>	<input type="checkbox"/>
13. Assessment feeds back to students their learning needs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Assessment information modifies ongoing teaching of students	<input type="checkbox"/>	<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 1	2	3	4	5	Strongly Agree 6
15. Assessment results are consistent	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>
17. Assessment results are filed & ignored	<input type="checkbox"/>	<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"/>	<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"/>	<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"/>	<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"/>	<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"/>	<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"/>	<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"/>	<input type="checkbox"/>
25. Assessment interferes with teaching	<input type="checkbox"/>	<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"/>	<input type="checkbox"/>
27. Assessment is an imprecise process	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Assessment Literacy: Please circle the response (A – D) that answers each question.

Scenario #3

Mr. Valdez is an English teacher in the newly built middle school. Experienced in issues of classroom assessment, Mr. Valdez is often asked to respond to the district's questions concerning best practices for evaluating student learning.

7. Ms. Franklin, also an English teacher, asks what type of assessment is best for evaluating her 6th graders' writing skills. Which of the following methods is likely to provide the *best* response to her question?
 - A. selected response methods
 - B. true/false statements
 - C. completion items
 - D. essay prompts

8. One of the middle school math teachers is redesigning her tests to make greater use of "story problems" as a way to check students' math understanding. She consults with Mr. Valdez to see what, if any, concerns she should be aware of when constructing assessments of this type. Which statement is *not* an appropriate recommendation when designing story-based math tests?
 - A. make sure that the reading level is grade appropriate
 - B. avoid scenarios more familiar to certain groups over others
 - C. check for clarity of sentence construction
 - D. incorporate scenarios used during instruction

9. At the end of each class period, Mr. Valdez does a quick "check in" with his students to get an impression of their understanding. In this example, the primary purpose for conducting formative assessment is to
 - A. identify cumulative knowledge.
 - B. determine content for the final exam.
 - C. plan classroom instruction.
 - D. evaluate curriculum appropriateness.