

University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

5-2013

Examining alignment : national and local assessments and the common core state standards in mathematics.

Ryan M. Higgins
University of Louisville

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

Recommended Citation

Higgins, Ryan M., "Examining alignment : national and local assessments and the common core state standards in mathematics." (2013). *Electronic Theses and Dissertations*. Paper 617.
<https://doi.org/10.18297/etd/617>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

EXAMINING ALIGNMENT: NATIONAL AND LOCAL ASSESSMENTS
AND THE COMMON CORE STATE STANDARDS IN MATHEMATICS

By

Ryan M. Higgins
B.S., Indiana University Southeast, 2004
M.Ed., Xavier University, 2008

A Dissertation Submitted to the Faculty of the
College of Education and Human Development of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

May 2013

Copyright 2013 by Ryan M. Higgins

All rights reserved

EXAMINING ALIGNMENT: NATIONAL AND LOCAL ASSESSMENTS
AND THE COMMON CORE STATE STANDARDS IN MATHEMATICS

By

Ryan M. Higgins
B.S., Indiana University Southeast, 2004
M.Ed., Xavier University, 2008

A Dissertation Approved on

March 4, 2013

By the following Dissertation Committee:

Elizabeth Todd Brown

Dissertation Director

William S. Bush

Samuel C. Stringfield

Melissa Evans-Andris

Dena H. Dossett

DEDICATION

This dissertation is dedicated to my husband

Chris Higgins

without who I would not have believed in myself.

&

To my parents and mother-in-law

Gary and Suzanne Hansen & Rhonda Higgins

who helped me through every trial and tribulation.

ACKNOWLEDGMENTS

I would like to thank my family, friends, and all those who advised me throughout this process. Without their love, support, and encouragement I could not have succeeded in conquering my dissertation.

I first and foremost want to thank my advisor and dissertation chair, Dr. E. Todd Brown, for sticking with me through the good, bad, and the ugly. Despite my bouncy ball-type start, she helped me finish strong and with confidence. I am honored to be Todd's only doctoral child and thankful that she was willing to be my Ph.D. Mom. The support and encouragement Todd supplied went far beyond the call of duty and her willingness to point out my awkward writing was invaluable. Thank you for never giving up on me and never allowing me to give up on myself.

I would also like to thank my dissertation committee for each doing their part to contribute in making my dissertation a success. Specifically, I'd like to thank Dr. Bush for encouraging me to look closely at the mathematical practices. Without the time he spent with me in discussion and the encouragement to take on the challenge, I would not have found such a strong future research passion. I would also like to specifically thank Dr. Dossett. I appreciate her prompt responses to the million questions I have asked throughout the time we've worked together. Also, Dr. Evans-Andris and Dr. Stringfield for their kind words and extra direction in qualitative research and APA intricacies.

There are several people who have contributed in either my dissertation research or through additional support. The following people contributed in my success in one

manner or another: Kristen Harbour, Sarah Bush, Amy English Hunter, Kathy Cash, my mom, Jennifer Barnes, Jonna and Kate Frakes, Barbara Fondren, Jill Jacobi-Vessels and Melissa Shirley.

Finally, I'd like to thank my family, immediate and extended: Suzanne and Gary Hansen, Rhonda Higgins, Kristina and Adam Harbeson, and Steve and Rita Higgins. I have not forgotten my husband and children, Chris, Paige and Paiton. They deserve special thanks for all of their love, support, encouragement, and forgiving. I could not have made it without the immeasurable amount of sacrifice Chris has given for my success. I can only hope to give back as much as has been given to me by all of my family and friends.

ABSTRACT

Examining Alignment: National and Local Assessments and the Common Core State Standards in Mathematics

Ryan M. Higgins

March 4, 2013

In support of the national movement to improve mathematics instruction and assessment, states and districts are looking for the best tools to measure student progress toward proficiency. There is a national dialogue about how to use 8th-grade measurements like ACT EXPLORE and NAEP as predictors of student success and school accountability. This dissertation shares research that examined alignment of district and national assessments to the 8th-grade Common Core State Standards in Mathematics (CCSSM). Three research questions were examined in this study to determine the extent of content validity and mathematical practice representation in three 8th-grade assessments compared to the CCSSM. The three assessments were: the ACT EXPLORE, the 2009 NAEP released mathematics items, and the Jefferson County Public School (JCPS) district interim assessments.

The study utilized a mixed methods research design to answer the three research questions. To determine the extent of content validity in application to the 8th-grade CCSSM content standards, a quantitative principal component analysis to determine domains represented by assessment items was performed on the JCPS interim assessments and the ACT EXPLORE. Qualitative alignment data were gathered from the

three assessments utilizing the Webb Alignment Tool (WAT), which also resulted in quantified data. To analyze the representation of the mathematical practices in the three assessments, the researcher used a method similar to that identified in the WAT to determine which practices each assessment item included. Thus, the mathematical practices were analyzed utilizing both qualitative and quantitative data and also included a meeting with content experts to discuss results of the data gathered.

Findings of the study indicated that the JCPS interim assessments were considered well aligned to the 8th-grade CCSSM content standards. However, the ACT EXPLORE and 2009 NAEP released items did not align with the 8th-grade CCSSM. Further analyses of the two national exams revealed a higher level of alignment with the 7th-grade CCSSM content but warrant further analysis. Due to a lack of methodological support for identifying the mathematical practices, no results were reported regarding the level of representation of the practices in each of the assessments.

TABLE OF CONTENTS

	PAGE
DEDICATION.....	iii
ACKKNOWLEDGMENTS.....	iv
ABSTRACT.....	vi
LIST OF TABLES.....	xiii
CHAPTER I: INTRODUCTION.....	1
Background.....	1
This Specific Study.....	11
Research Questions.....	11
Study Significance.....	11
Delimitations.....	12
Assumptions.....	13
Definitions.....	13
Organization of the Dissertation.....	14
Summary.....	14
CHAPTER II: LITERATURE REVIEW.....	15
Introduction.....	15
Literature Search.....	15
National Assessment of Educational Progress.....	16
History of NAEP.....	16

NAEP Reliability and Study Relevance.....	19
Mathematical Strand Areas and Study Relevance.....	20
NAEP Research.....	22
ACT EXPLORE.....	31
History of ACT EXPLORE.....	31
ACT EXPLORE Reliability and Study Relevance.....	32
Mathematical Strand Areas and Study Relevance.....	33
Relevant ACT Research.....	34
Benchmark Exams/Interim Assessments.....	37
Assessment Models.....	37
Use and Relevance in JCPS.....	41
Eighth Grade Mathematics.....	45
State Mathematical Standards at Eighth Grade.....	49
Comparisons.....	49
Common Core State Standards.....	51
History.....	51
CCSS Mathematical Practices.....	54
Assessing the Practices.....	57
Importance of Research Exploring the Alignment of the Assessments to CCSSM.....	60
Content Validity.....	64
Content Validity and Standards Alignment.....	65

Evidence of Content Validity in an Assessment.....	66
Standards Alignment.....	67
Webb’s Alignment Tool (WAT).....	68
Survey of Enacted Curriculum (SEC).....	70
Achieve.....	71
The Council for Basic Education (CBE).....	72
Alignment Research Applicable to this Study.....	73
Cognitive Complexity.....	75
History and Relevant Research of Cognitive Complexity Models.....	75
Theoretical Constructs of Cognitive Complexity.....	81
Theoretical Framework of this Study.....	85
Selected Frameworks for Alignment.....	85
Conceptual Framework.....	87
Conclusions.....	89
Chapter II Summary.....	89
Research Questions.....	89
CHAPTER III: METHODOLOGY.....	91
Design of the Study.....	91
Restatement of Purpose and Research Questions.....	91
Study Design.....	92
Population and Sample.....	95
Population.....	95
Sample and Sampling Procedures.....	95

Instrumentation.....	96
Description of Instrument.....	96
Instrument Reliability.....	100
Data Collection Procedures.....	101
Alignment Procedures.....	101
Data Collection.....	102
Data Analysis.....	103
Methods Used to Analyze Data.....	104
Research Question One.....	104
Content Coverage.....	104
Research Question Two.....	105
Research Question Three.....	105
Reliability.....	106
Limitations.....	106
CHAPTER IV: RESULTS.....	108
Introduction.....	108
Primary Analysis.....	110
Principal Component Analysis.....	110
Secondary Analysis.....	114
Webb Alignment Tool (WAT).....	114
Summary of WAT Findings.....	124
Tertiary Analysis.....	125
CCSS Mathematical Practices Alignment.....	125

Summary of Mathematical Practices Findings.....	130
Conclusion.....	132
CHAPTER V: DISCUSSION.....	134
Introduction.....	134
Summary of the Study.....	134
Restatement of the Problem Statement.....	134
Restatement of the Purpose and Research Question.....	135
Findings and Relation to the Literature.....	136
Research Question 1.....	136
Research Question 2.....	146
Research Question 3.....	154
Conclusions and Implications.....	156
Limitations of this Study.....	160
Recommendations for Future Research.....	165
Summary.....	167
REFERENCES.....	170
CURRICULUM VITAE.....	195

LIST OF TABLES

TABLE	PAGE
1. Webb’s Depth of Knowledge for Mathematics.....	87
2. Number of Items and Item Type.....	96
3. Alignment Levels Using the Four Criteria.....	98
4. 8 th -grade CCSSM Categorical Concurrence.....	117
5. 7 th -grade CCSSM Categorical Concurrence of ACT EXPLORE and 2009 NAEP.....	118
6. 8 th -grade CCSSM DOK Consistency.....	120
7. 7 th -grade DOK Consistency for ACT EXPLORE and 2009 NAEP.....	121
8. 8 th -grade CCSSM Range of Knowledge Consistency.....	122
9. 7 th -grade CCSSM Range of Knowledge Consistency.....	123
10. Summary of Alignment for each Assessment.....	124
11. Summary of 7 th -grade Alignment for ACT EXPLORE and 2009 NAEP.....	125
12. Coding Results for the Assessment Item in Figure 1 for Part B Only.....	129
13. DOK Representations.....	144

CHAPTER I

INTRODUCTION

The intent of No Child Left Behind legislation (2002) is to achieve a 100% proficiency level of all public school students in mathematics and language arts. This aim created the need for accountability targets to monitor state progress called Adequate Yearly Progress (AYP) goals. State response to these goals have been mixed and are not all positive, Cronin, Dahlin, Adkins, and Kingsbury (2007) researched state responses to these requirements finding some states had lowered proficiency scores on their assessments to increase the number of students passing. An additional finding by Cronin et al. (2007) was that some states had lowered the rigor of their standards that resulted in lower level of rigor on assessments allowing a greater number of students to reach proficiency.

In response to these issues, the Council of Chief State School Officers (CCSSO) along with the National Governor's Association worked together to write and publish the Common Core State Standards (CCSS) for mathematics and language arts. The Common Core State Standards in Mathematics (CCSSM), a set of high expectations and level of rigor were created with the goal of increasing the mathematics readiness of students for college and careers.

Background

Since the release of the CCSS in 2010, there are currently no national exams specifically designed to measure achievement of students on the CCSS. States are

choosing their own methods of measuring student achievement on the CCSS as well as the types of progress monitoring instruments that will be used. Thus, the use of current national and district assessments are being considered as possible accountability measures and student proficiency predictors. As states begin their transition to the CCSS, it is vital to consider the content validity of such national and district assessments as they apply to the CCSS.

National Assessment of Educational Progress. The National Assessment for Educational Progress (NAEP) has been in existence for more than 40 years. The reliability of NAEP data and NAEP scores is considered to be relatively high. This is because of the longevity of test usage, the number of students in the sample, the range of demographics represented and the fact that the exams are highly monitored. There are many empirical studies that have used NAEP data and common themes emerge. One theme was varied types of comparison studies. Uzell, Simon, Horwitz, Amelga, Lewis, and Casserly (2011) studied mathematics achievement of Trial Urban District Assessment (TUDA) participants. The goal of the Uzell et al. (2011) study was to determine whether achievement gaps were decreasing within the TUDA participants. The outcome showed that NAEP data is usable by each state or TUDA participant to determine gap trends. However, the data will only be reliable if the district is using its own growth data from the NAEP assessment for longitudinal comparisons. Data compared to that of other TUDA districts or states cannot be used reliably (Uzell et al., 2011).

An example of those using the NAEP data is that of the district being studied in this research. The Jefferson County Public School district is involved with National

Assessment for Educational Progress (NAEP) as a Trial Urban District Assessment (TUDA) participant. The intent of TUDA is to determine the level of “feasibility of using NAEP to report on the performance of students at the district level” (NAEP, 2012). Thus, as a TUDA participant the study district can use NAEP data to monitor proficiency in mathematics for the students in their district and compare the district level of proficiency with that of the state and the nation.

One Lubienski (2006) study compared gap trend data of different ethnicities and how teacher practices affect changes in gap trends. The results indicate that there were a number of positive effects between teacher practices and Black and Hispanic achievement. For example, teacher knowledge of the NCTM *Standards* showed a “significant positive predictor of fourth-grade mathematics achievement” (Lubienski, 2006, p. 20). Bandeira de Mello, Blankenship, McLaughlin, and Rahman (2009) studied state proficiency scores by using an equating method to mapping the state scores to the NAEP scores. The overall finding of the study determined that the equating method utilized by Bandeira de Mello et al. (2009) allowed for the identification of states with higher standards but not a large number of students reaching proficiency.

Another theme found in the literature was validity studies about using NAEP data. A study completed by Daro, Stancavage, Ortega, DeStefano, and Linn (2007) compared the NAEP fourth- and eighth-grade mathematics assessments to the published NAEP frameworks. The study showed the assessments represented the framework by covering content and skill but some strands were over represented than others.

A study by Kiplinger and Linn (1994) explored the score validity of state to NAEP achievement. Their main focus was on the reliability of using test linking to

compare results from the NAEP assessments. The idea of linking tests is that results from one test could be compared to results of another test. It was determined that, if linking of tests occurs, states must ensure the content of the tests are similar (Kiplinger and Linn, 1994). If the content of the tests is not determined to be similar, the equating of the tests does not provide valid evidence.

ACT EXPLORE. The ACT EXPLORE is a part of the ACT Educational Planning and Assessment System (ACT EPAS). The system consists of three exams given at specific times to monitor student progress over time. The ACT EXPLORE is given during eighth-grade as students are preparing for high school. The ACT PLAN is given during tenth-grade as students are considering preparation for college. The ACT is given during eleventh- and twelfth-grade as students begin the journey into college or career. Similar to the NAEP, ACT reaches across the nation and has a large sample size, thus contributing to the level of reliability of the scores obtained from the exams. Additionally, the ACT EXPLORE *Technical Manual* (ACT, Inc., 2011), a reliability coefficient of 0.82 is expressed for the exam.

There is a limited amount of empirical research available on the ACT exams other than the studies conducted by ACT. One study completed in 2005 by ACT involved a state that reported 56% of the state's students reaching proficiency on the mathematics state exam. However, ACT (2005) exam results reported far fewer students in the same state showing the probability of college success. There are two possible reasons for this discrepancy. First, the state's standards may have been lacking in college and career readiness standards. Also, consistent with the Uzell et al. (2011) study, the state may have

lowered their proficiency score for the exam or lowered the rigor in their standards to show a higher number of students showing proficiency level scores.

Despite a thorough search of the literature using multiple sources, no studies on the validity of the EXPLORE exam were revealed. The only information available was in the EXPLORE technical manual reporting the method for designing the test and ensuring content validity through coding.

Beginning in 2007, the Kentucky Department of Education (KDE) mandated the use of ACT EXPLORE for all public schools. The KDE website (2012a) gives reasoning for the mandate of ACT EXPLORE as a “way to determine possible inappropriate practices.” Additionally, the study district has indicated a move toward using the ACT EXPLORE as an accountability measure in the district.

Interim assessments. Interim assessments are often used synonymously as benchmark assessments and are defined by Perie, Marion, Gong and Wurtzel (2007) as those used by teachers, but also used for gathering data. Operational definitions were identified to distinguish the difference in timing and purpose of each type of assessment referred to within this study. For instance, the operational definitions for diagnostic assessments are assessments given before the beginning of a unit to find problem areas in previously learned material. Summative assessments are defined as tests given at the end of a unit, class or year as a culmination event. Also, formative assessments are assessments used during instruction and outcomes should be addressed immediately. Interim assessment data can be synthesized and disaggregated for use by administrators and district staff. A study by Davis, Caros, Grossen, and Carnine (2002) intended to determine whether a certain interim assessment could predict achievement on large-scale

assessments including the SAT9. Davis et al. (2002) concluded that, if the reliabilities of each exam are high and internal consistency is high, then, yes, the interim assessments could be used as predictors.

The Philadelphia school district has used interim assessments since 2004. A study completed by Christman et al. (2009) explored interim assessment scores along with teacher survey data and found that professional leadership and school support climate were better predictors of achievement on the state standardized exam than the interim assessments. Another study from the Philadelphia school district conducted by Bulkley, Christman, Goertz, and Lawrence (2010) looked at interim assessments to determine whether format of the pacing for the interim assessments was working to increase student performance on achievement tests. Similar to this study, the interim assessments were intended to provide feedback to teachers. In the Bulkley et al. (2010) study teachers were instructed to use one week after the exams for re-teaching in areas where students showed weaknesses. The research showed that the timing of the feedback given to teachers from the interim assessment data was important to the impact on student achievement on the larger standardized assessments. For interim assessments to make an impact on student achievement, teachers must get immediate feedback and then immediately apply the data to the re-teaching week.

The study district has developed a set of interim assessments, which have been used for less than five years. The assessments are updated to reflect the state's Core Content for Assessment. The study district has provided information on the purpose of interim assessment use in the district. The indicated uses are as follows:

- Interim assessments are intended to provide feedback to the teacher on students.

- Interim assessments are intended to keep teachers focused on key concepts and to provide specific pacing.
- Interim assessments give students a chance to see items similar to that on standardized exams.
- Interim assessments give information to administrators et al. on progress of students. (Dossett, personal communication, September 28, 2011)

Eighth-grade mathematics. An ACT study published in 2005 determined that college and career readiness is most affected by the level of achievement obtained by the end of eighth-grade. An additional study of eighth-grade mathematics reported the need for stronger and more rigorous mathematics requirements for eighth-grade since high school teachers reported spending almost 25% of their time re-teaching what should have been learned by the end of eighth-grade (ACT, 2007b). A study conducted prior to the publication of CCSS by Porter, Polikoff, and Smithson (2009) of all state eighth-grade mathematics standards showed only 20% of the standards were cohesive across all states.

Eighth-grade is the focus of this study with research that shows the impact of eighth-grade on future achievement and the importance of eighth grade as a pivotal year before high school. There is also evidence that mathematics standards are not cohesive across states but the intention of the CCSS is to help address this issue.

Common Core State Standards in Mathematics (CCSSM). The CCSS were developed to help to ensure college and career readiness. This purpose is supported by several studies. The American Federation of Teachers (1998) completed a comparison study of exams between Japan, France, and U.S. state exams with commercial exams available in the United States. They determined that the U.S. exams lacked rigor and

addressed extremely broad and not very deep curricula (AFT, 1998). ACT (2007a,b,c) has shown through multiple studies that students in the United States lack preparedness for high school let alone college and career readiness. In one ACT (2007) study it identified the lack of agreement between secondary teachers and higher education teachers on specific standards necessary to be prepared for college as a hindrance for preparing students for college and career.

CCSS have only been adopted by states for two years and there are a few studies already. Carmichael, Wilson, Finn, Jr., Winkler, and Palmeieri (2009), prior to the CCSS draft release, compared the CCSSM standards to the NAEP assessment for mathematics. The study utilized a rubric to assess the NAEP; however, the rubric development and description was not discussed in their methodology. Content experts gave NAEP a grade of “C” for testing the CCSSM standards citing lack of rigor and lack of clear language in assessment items. ACT also completed a validity study of the EXPLORE exam in 2010 reporting a 100% alignment with the CCSSM content standards.

Specific to the Kentucky, there are two studies that compare the state’s mathematics standards for assessment and the CCSSM. The first, a Carmichael, Martino, Porter-Magee, and Wilson (2010) study used the same rubric as the one developed for the previous study of NAEP completed in 2009. The state was given a grade of “D” with the researchers citing a lack of organization, clarity and rigor when the state was compared to the CCSSM standards. However, the Carmichael et al. (2010) study was completed prior to the state modifying their program of study to address the national standards. Contradictory to the Carmichael et al. (2010) study; the KDE published its own “Crosswalks” (state, 2012) comparing the CCSSM to the KDE’s Program of Study. For

each of the CCSSM standards, the content experts for the state gave each standard a 0 - 3 depending upon the level of representation already existing in the state's current standards, or program of study. The state cited 16 excellent matches, 12 good matches, and five weak matches (state, 2012).

An additional component to the CCSSM content standards is the standards for mathematical practices. Designed around the research and development of the NCTM *Principles and Standards for Mathematics* (2000) and the National Research Council's (2001) mathematical proficiencies, the standards for mathematical practices were designed to increase student proficiency and conceptual understanding. McCallum (2012) describes the practices as "the various way in which proficient practitioners of mathematics carry out their work" (p. 7). Although several authors agree that the inclusion of the practices within assessment items is imperative (Kepner & Huinker, 2012; McCallum, 2012), there is little support for methods of doing so.

Alignment. While the KDE compared standards to standards, they were looking for content coverage and content matches. Some studies, such as this study, are performed comparing standards and assessments. This alignment checks for validity of an assessment to measure the standards for which the test is designed. Several tools have been designed to complete this type of alignment. There are two specific tools, supported by the CCSSO, which stand out in the literature with far more published studies using the tool.

The Webb Alignment Tool (WAT) is considered a qualitative and quantitative combination. Qualitative document analysis is performed which results in quantitative data from the coding process. Webb (Webb, 1997a, 1999; Webb, Alt, Ely, & Vesperman,

2005) performed several studies using the WAT and has revised it for: clarity, the number of coders most appropriate for the highest level of reliability and, to develop a general rubric to assist in determining successful alignment.

The Survey of Enacted Curriculum (SEC) designed by Porter, differs from the WAT in a few ways. First, the SEC is more heavily focused on classroom instruction versus standardized assessment alignment (Blank, Porter, & Smithson, 2001). Second, to use the SEC, teachers must enter classroom data, such as lesson plans and classroom assessments, into an online tool. After the data is entered, outside coders complete the alignment and send the report back to the teacher. Thus, removing any personal alignment process from the classroom teacher, or researcher. Finally the WAT uses four cognitive complexity levels and the SEC model uses five levels.

Cognitive complexity has also been referred to in the literature as cognitive demand and depth-of-knowledge (DOK) and is generally defined as the level of demand that is placed on a student to answer a test item or to learn a standard. Bloom is likely the most well known for his *Taxonomy of Educational Objectives* (1956) which applies specifically to the level of questioning used by teachers. The taxonomy contains a hierarchy of levels of cognitive demand identified by the verbs used in the classroom. Although Bloom's model has been a foundation for others, research has shown and is supported by Hess (2006), that the use of verbs alone on assessment items is difficult because the verbs can be used in multiple contexts. This study uses Webb's tool, which includes four levels of cognitive complexity, defined in the tool as depth-of-knowledge.

This Specific Study

Research Questions

The purpose of this study was to align the Common Core State Standards in mathematics (CCSSM) to the available test items of NAEP, ACT EXPLORE, and the interim assessments utilized by the Jefferson County Public School District (JCPS). By completing the alignment, the research reported the extent to which each exam represents the CCSS eighth grade mathematics content standards. The research questions explored in this study are as follows:

- 1) To what extent do the (1) eighth-grade mathematics released items in the 2009 NAEP assessment, (2) ACT EXPLORE, and (3) items from the Jefferson County Public School District interim assessments align with the eighth-grade CCSSM?
- 2) To what extent are the CCSS Mathematical Practices represented in each strand of the eighth-grade 2009 NAEP mathematics assessment released items, ACT EXPLORE, and the Jefferson County Public School District interim assessments?
- 3) Which of the tests shows the most content validity and which shows the most mathematical practices represented compared to CCSSM?

Study Significance

The study state has adopted the CCSSM and has developed standardized assessments that assess their students on their proficiency on the CCSSM. Thus, when accountability measures are put into place, these measures must report valid and reliable information regarding student achievement. In order to be certain that the use of the assessments as accountability measures or progress monitoring tools is justifiable,

alignment must be strong between the assessments and the standards. The results of this study will provide beneficial information regarding the usability of the NAEP, ACT EXPLORE, and the district's interim assessments as accountability measures or progress monitoring tools.

Additionally, according to the literature, the mathematical practices are not being included into assessment at this point in time. It is highly recommended (Confrey & Krupa, 2010; Kepner & Huinker, 2012; Krupa, 2011; McCallum, 2012) that this take place for teachers and administrators to include the CCSS Mathematical Practices in instruction and intervention practices. However, methods for ensuring inclusion of the practices in assessment items; for designing assessment items including the practices; and, a procedure for aligning the practices with the content in the items have not yet been designed.

Delimitations

A few delimitations for this study exist. First, grade eight is the only grade level data to be used in this study. The data available at this grade are vast due to the availability of NAEP and ACT EXPLORE data. Neither NAEP nor ACT assess at every grade level; however, both are available for eighth-grade assessment. In this study only mathematics content is assessed whereas multiple content areas focusing on CCSS content are tested in the proficiency assessments.

Finally, the assessments, which are to be analyzed in this study, are not the only assessments that could be utilized. Other international assessments which are given periodically could be used, such as: Civics, Arts, Economics, Foreign Language, Geography, Reading, Science, Technology, U.S. History, World History, and Writing.

However, JCPS participates in the NAEP mathematics and language exams and ACT EXPLORE consistently and uses the EXPLORE as an accountability measure. If alternate assessments were analyzed the data may be beneficial to others who use the information from those assessments.

Assumptions

This study relies upon the most recent test items available. Thus, the assumption is that the assessments used for alignment in this study are the latest version of the assessments available. This study also assumes that the national assessments are not intentionally aligned to the CCSS content in eighth-grade mathematics prior to the study.

Definition of Terms

Alignment: The comparison of standards to assessment items. The representations of the items on an assessment to the standards have been listed by the state.

Cognitive complexity: The level of thought required by a student to answer a question on an assessment.

High-Stakes assessment: A standardized assessment that is utilized as a measure of accountability for students, teachers, administrators, states or any combination therein.

Interim assessment: A test utilized at a specified point in instruction as both a diagnostic or formative tool and as a method for gathering student and classroom data on a district level.

Low-Stakes assessment: An assessment with little or no link to accountability measures for any person, school, or state.

Reliability: The level to which an assessment can be trusted to represent correct achievement data in application to the CCSSM.

Validity: Represented by an assessment reflecting the standards of the CCSSM.

Organization of the Dissertation

The following chapters of the proposed dissertation are as follows: Chapter II will describe the literature in relation to the three assessments utilized in this study (NAEP, EXPLORE, and the interim assessments), history of the Common Core State Standards in Mathematics, and a review of alignment models; Chapter III will describe the study design, sample, and the methods for aligning the CCSSM and the assessment items as designed by Webb (1999); and Chapters IV and V will provide the results of the data analyses and the conclusions, discussion, limitations and recommendations respectively.

Summary

The literature cited in this study supports the need for further preparation of eighth-grade students in mathematics and in utilizing a set of national standards to support fairness and true measures of mathematics proficiency and growth across states. Furthermore, it supports a true determination of NCLB requirements for progress by schools and meeting of Adequate Yearly Progress. In addition, support is shown for aligning standards to assessments, and multiple models are supported.

The goal of this study was to determine the extent of alignment among the items on the three assessments and to determine which of the assessments most validly represents the CCSS mathematics content for eighth grade. Thus, it provides valuable information regarding the usability of the assessments as accountability measures.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The purpose of this study was to determine the level of alignment between the mathematics content portions of two national and one local interim assessment with the CCSSM content for eighth-grade. After completing the alignment, this study will determine the content validity of each assessment when compared to the CCSSM. The result of the study will allow administrators to choose assessments, which are well aligned in content and in mathematical practices, to use as accountability measures.

This chapter provides a review of the literature as it relates to the alignment of assessments and content validity. First, each of the three assessments studied will be discussed and will include information about their history, purpose, and past research utilizing the assessment. Second, literature regarding the CCSSM and eighth-grade standards is presented followed by the literature supporting the frameworks to be utilized in this study.

Literature Search

The following databases, available through the University of Louisville library, were utilized for developing this review of pertinent research: EBSCO Academic Search Premier, Education Resources Information Center (ERIC), ProQuest Digital Dissertations, ProQuest Research Library and Wilson Web. Sources were also retrieved from the following publications and organizations: American Educational Research

Association (AERA); Council of Chief State Standards Officers (CCSSO); National Council for Teachers of Mathematics (NCTM); the National Center for Education Statistics (NCES); ACT; and Achieve, Inc. Finally, search for additional literature was conducted using other internet based search engines, such as Google Scholar. Key search terms for conducting the research review were as follows: cognitive complexity; depth of knowledge; Bloom’s Taxonomy; National Assessment for Educational Progress; state to NAEP links; No Child Left Behind; Common Core State Standards (CCSS); state test comparison; content analysis; content validation; cognitive complexity; benchmark assessment; interim assessment; state to state assessment comparison; standards to test comparison; ACT EXPLORE; history of assessment; history of NAEP; history of ACT; NAEP reliability; ACT reliability; content validity and construct validity.

The literature used in this review included books, technical manuals, peer-reviewed manuscripts from research journals, conference proceedings, technical reports and technical or historical information from websites specifically linked to NAEP or ACT.

National Assessment of Educational Progress (NAEP)

History of NAEP

The National Assessment for Educational Progress, commonly referred to as “The Nation’s Report Card,” is used to track longitudinal student progress in multiple subject areas in 4th, 8th, and 12th grades across the country. NAEP is a congressionally mandated assessment guided by the National Assessment Governing Board (NAGB). Tests have been conducted every two or three years since 1969 and have been used as a “trusted and reliable” data resource for policy makers, administrators and teachers (IES,

2010, p. 3). Currently, NAEP is given to fourth-, eighth-, and twelfth-graders in a variety of subjects including: mathematics, reading, science and writing (IES, 2010, p. 5).

Several other assessments are evaluated periodically in addition to the main areas, which are tested consistently every two years. The other tests given are in the arts, civics, economics, geography and US History (IES, 2009).

NAEP is a referential data source for judging America's progress in education in the tested subject areas. Since the sample of students tested is so large, it is unnecessary for all students to be tested on each item (Johnson, 1992); therefore, it is a more reasonable way to gauge progress while having a high level of integrity for research use. NAGB currently prohibits state administrators and federal officers from making any judgments about student performance in a state based upon its NAEP proficiency scores. That is to say, state educational policy makers and school officials cannot determine pass or fail rates according to AYP requirements on proficiency or increase in proficiency for the state dependent upon any portion of NAEP score data. Many articles have been written both for and against the use of NAEP data as an AYP indicator. Buckendahl et al. (2009) state, "Although there were no legislative mandates to officially use NAEP as a tool in NCLB's accountability system, there have been calls to formally include it in future policies. Thus, NAEP's intended purpose has potentially expanded from description and evaluation to include, at least for some stakeholders, accountability." (p. 4)

The NAGB gathers input from committees (content experts, teachers, researchers, and content associations like the National Council of Teachers of Mathematics (NCTM)) to develop frameworks for the NAEP. The committees work together to determine what

should be assessed at each grade level; they do not determine what or how mathematics should be taught (NAGB, 2011, p. 2).

“Since 1990, the Main NAEP mathematics assessment has been guided by, although not determined by, a framework based on the National Council of Teachers of Mathematics *Curriculum and Evaluation Standards for School Mathematics*.” (Lubienski, 2006, p. 5) NCTM states, “NCTM has participated in the development of the mathematics framework for the national NAEP several times since it was first administered in 1990. The 1990 NAEP framework reflected NCTM's Standards document, *Curriculum and Evaluation Standards for School Mathematics*.” (NCTM, 2004) The National Council of Teachers of Mathematics framework outlines mathematical teaching and learning guidelines in their *Curriculum Focal Points for Pre-kindergarten through Grade 8 Mathematics* (NCTM, 2006). The NCTM Focal Points are an overarching guide to the skills and content strands at each grade level. These define expectations of mathematical competence for students up to grade eight. Most recently, in preparation for the NAEP given in 2011, the National Assessment Governing Board (NAGB) (2010), contracted with Achieve, Inc. to assess the quality and rigor of the NAEP (NAGB, p. 3). Achieve, Inc.'s findings indicated a need for a change of the scope of the twelfth-grade assessment (NAGB, 2010, p. 3).

The National Center for Education Statistics “is responsible for developing test questions, administering the assessment, scoring student responses, conducting analyses of the data, and reporting the results” (IES, 2010, p. 2). The framework set by NAGB for NAEP is another set of guiding principles in mathematics strand content and teaching practices that guide state assessment and standard development. However, according to

NAGB (2003), the objectives in NAEP are based upon the *Focal Points* in the principles and standards developed by the NCTM framework. Aside from congressional policy mandating the administration of the NAEP (IES, 2010, p.2), the National Assessment Governing Board (NAGB), determines the content areas to be assessed, sets proficiency rates and controls the way data is disseminated and reported (NAGB, 2011). Thus, it is imperative that state benchmarks are focused on both sets of frameworks to be successful on the NAEP assessment.

NAEP Reliability and Study Relevance

NAEP started in the late 60's and is the longest standing nationally given assessment available, earning an automatic credibility for the assessment and its results. NAEP is also the most widely studied assessment due to its nationwide reach. It has been compared with a variety of other assessments, other NAEP assessments, and across states. Given that the assessment is highly monitored when it is given, it is one of the most reliable in standardization. NAEP is frequently used for comparisons of state achievement data since each state utilizes its own set of standards and develops its own assessments; thus, state assessments are not appropriate for comparing the achievement levels of students across states.

Reliability of NAEP state-to-state results is pertinent to this study. Wells, Baldwin, Hambleton, Sirecei, Karatonis, and Jirka (2009) assessed the equity of NAEP results from state-to-state comparisons. Wells et al. (2009) found, despite rising concern at the time regarding the reliability of NAEP scores when comparing state proficiency levels, the scores were well founded. Original concerns regarding curriculum differences and standards requirements among states may cause some states to be at a disadvantage;

thus, it was determined that comparisons of achievement between the states would be invalid. Wells et al. (2009) used NAEP 2005 reading and mathematics results that were purposely chosen from: 1) two states with a highly aligned curriculum with the NAEP assessment; 2) two states with a curriculum not highly aligned; 3) and one state somewhere in the middle with their curriculum. The study showed that the level of curriculum alignment did not seem to affect the reliability of the scores released by NAEP for each state when achievement level proportions were compared (p. 406). Thus, according to Wells et al. (2009), it is not necessary to scale the NAEP scores separately.

Mathematical Strand Areas and Study Relevance

The NAEP mathematical framework is overseen and updated by the NAGB. In this study, data analyses and alignment will be completed on 2009-released NAEP mathematics items. 472,100 eighth-grade students completed the 2009 NAEP assessment. The significant number of students in the sample contributed to the reliability of the NAEP assessment. However, information regarding the 2011 NAEP will be provided periodically due to its significance in regards to the achievement reported. Additionally, while Kentucky adopted the CCSS in 2010, it began using the standards in the 2011-2012 academic year; thus, neither the 2009 or the 2011 assessments results were affected by the changes to the standards and curriculum for the study district. For the purpose of this study, we will refer to the NAEP 2009 framework since the data from 2009 is the most recent available data for research use. In the NAEP 2009 mathematics framework are areas regarding content, cognitive complexity, and item type. At the 8th-grade level, the NAEP framework includes the following content areas and descriptions (NAGB, 2008, p. 5):

- Number Properties and Operations (including computation and understanding of number concepts)
- Measurement (including use of instruments, application of processes, and concepts of area and volume)
- Geometry (including spatial reasoning and applying geometric properties)
- Data Analysis, Statistics, and Probability (including graphical displays and statistics)
- Algebra (including representations and relationship)

It is important to understand that the framework does not describe what should be taught in mathematics, but only describes what is to be assessed.

Each assessment is comprised of “three item types: multiple choice, short constructed response, and extended constructed response” (NAGB, 2008, p. 51). The item types are defined as follows:

- Multiple choice - students select the best answer from a choice of answers given
- Short constructed response - items that require students to give a numerical answer, picture, or brief explanation and can be scored as right or wrong
- Extended constructed response - items that require numerical responses in addition to a description or picture. These items have several parts and are scored using multiple levels of points.

One goal that is evident in the NAEP mathematics framework is to create a balance among content items and complexity based upon what should be expected for each grade level. This ideal balance is constantly reviewed according to the changing generations of students.

NAEP Research

State-to-NAEP comparisons. The design of the NAEP allows for states to compare progress in mathematics achievement across each state and certain size school districts. In a study conducted by Uzzell, Simon, Horwitz, Amelga, Lewis, and Casserly (2011) who analyzed student achievement in mathematics and reading between 2007 and 2010, researchers sought to determine whether urban schools were improving and whether achievement gaps were closing. During their study, the authors used data provided from the NAEP to gain information about a specific demographic of schools, specifically those qualifying as Trial Urban District Assessment (TUDA) schools. The findings from the study are not the focus here; however, the focus on conducting comparisons across the nation by state or district was important. This strategy has been a popular form of determining growth among states despite the caution levied by the U.S. Department of Education. Although states can look at their own longitudinal growth scales and look more closely at specific districts, it is unrealistic to compare NAEP scores from state-to-state. As noted by Uzzell et al. (2011), “the nation does not have an assessment system that allows us to measure progress relative to the same standards across all school districts in the country” (p. 1). However, with the advent of a set of CCSS, an upcoming assessment looming, and the majority of the states committed to adopting the CCSS, the NAEP may become a tool for addressing educational change and determining educational progress more than ever before.

State-to-NAEP comparisons are present in studies attempting to show causality for discrepancies in state trends compared to NAEP trends (Sanford & Fabrizio, 1999; Wei, Shen, Lukoff, Ho & Haertel, 2006). A study to compare state assessments in North

Carolina to the NAEP was conducted (Sanford & Fabrizio, 1999) due to observing a high level of inconsistencies in the test scores regarding students at basic and proficient levels of mathematics and reading. Using a three-level framework for assessing the two assessments, the authors examined technical aspects, content and complexity similar to the current study. By comparing each of the three areas noted, Sanford and Fabrizio (1999) further guided reform of the North Carolina testing program to be more congruent with the NAEP (p. 13). Aligning to the NAEP allows state policy leaders to understand student achievement for the state as compared to student achievement in the nation; thus giving a stronger picture of student growth or decline. This comparison also allows for state leaders to determine stability of student learning compared to national data.

Another reason for state-to-NAEP comparisons relates to the description of gap trends (i.e. race comparisons) (Lubienski, 2006; Erickson, Ho, Holtzman, Jaciw, Lukoff, Shen, Wei & Haertel, 2007). The Erickson et al. (2007) study researched standardized assessments measuring similar constructs. The intent was to determine if it is possible to compare scores from one assessment to the other. This would allow them to show that teachers are “coaching students to obtain a high score on one particular test instead of helping student to gain more generalizable knowledge or understanding about the subject matter” (p. 2). The authors compared a state high-stakes assessment to the NAEP assessment to see if it was obvious that teachers were teaching specifically to the state standards for testing purposes rather than for learning sake. Using a metric-free system, the authors compared six states and their data for multiple years (the academic years were different for each state) in order to show discrepancies in the White-Black and White-Hispanic achievement gaps. They succeeded in comparing the gaps between the state

tests and the NAEP for each of the six states, except that there were limitations present in the study through possible changes in state content standards (Erickson et al., 2007, p. 38).

Another purpose for comparing state and NAEP data, in addition to achievement gap comparisons, is the level of instructional practice changes made in order to close the gap (Lubienski, 2006; Wenglinsky, 2004). Both studies used 2000 NAEP mathematics data to review the affects of mathematics reform on the NAEP statistics particularly in reference to the NCTM curriculum standards that were developed in 1989. Wenglinsky's study was initiated by the debate about achievement gains in mathematics following the enacting of the NCLB (2002) reforms. Wenglinsky's (2004) concluded that the achievement gaps between black-white and Latino-white can be decreased through certain series of instructional practices. Lubienski (2006) claimed Wenglinsky's (2004) conclusions were unsupported. Through more thorough analyses and methodological changes from the original study, Lubienski supported the findings that Wenglinsky's conclusions did not properly use cross-sectional data and identified incorrect causal relationships amongst the data. Finally, comparison of proficiency standards between state and NAEP is an increasing trend. Again, with the intended goal of determining whether states are reaching a certain level of rigor in comparison to the NAEP, studies have been completed in order to compare scoring standards (Bandeira de Mello, Blankenship, McLaughlin, & Rahman, 2009). Equating is a method by which researchers and administrators attempt to link scores of one assessment to the scores of another by using an "equipercentile procedure" (Erickson, 1997, p. 145). Those researchers using this procedure to link state assessment results to the NAEP national

results believed they are able to compare results on their statewide assessment to that of the national assessments. Although equating, or test-score linking, was noted as an issue in making NAEP and state comparisons in the Bandeira de Mello et al. (2009) study, the researchers were able to equate the two sets of scores. They determined that certain states having higher standards, as measured via the NAEP scale, were not performing at higher proficiency levels as would be assumed (p. vi). Through mapping analyses, states can “compare the stringency of its criteria for proficiency with that of other states” and to compare proficiency levels of rigor of standards with NAEP (Bandeira de Mello et al., 2009, p. viii).

A frequent reason for comparing NAEP to state assessment scores is in mapping proficiency standards or achievement levels (Bandeira de Mello, Blankenship, McLaughlin, & Rahman, 2009; Linn, Graue, & Sanders, 1990). By mapping state standards onto the NAEP scales, a level of achievement for proficiency in one state can be compared with that of another; thus, “the mapping procedure offers an approximate way to assess the relative rigor of the states’ adequate yearly progress standards and the equivalent score representing the state’s proficiency standards can be compared to indicate the relative rigor of those standards” (Bandeira de Mello et al., 2009, p. v).

In the past, NAEP has been used to review and revise standards using content comparison analyses (Shapley & Brite, 2008; Timms, Schneider, Lee & Rolfhus, 2007). These studies were conducted in an effort to more closely correlate state standardized assessment with the expectations and outcomes of the NAEP. However, the Oklahoma study conducted by Shapley and Brite (2008) focused mainly upon methodology “comparing one set of assessment standards with another” (p. 1). While the study did not

analyze specific assessment items for either test, it did compare what content strands were fully addressed, partially addressed, or were not addressed in the Oklahoma state assessment. Shapley and Brite (2008) allowed administrators and policy makers to determine what to do with the results, as the sole reason for the study was only to give percentages of comparison in 4th-, 8th-, and 12th-grade assessments. Similar results were accounted for in the TIMSS et al.' (2007) study comparing New Mexico's standards with the NAEP. Again, the authors did not make specific recommendations regarding what the state policy makers should do regarding the alignment study; however, they did provide data regarding the reasons they documented partial fulfillments versus full fulfillment giving further information regarding possible changes in test items and content.

Before NCLB (2002) and the use of adequate yearly progress (AYP) reports, Linn, Graue, and Sanders (1990) used research to address comparisons of state and national data. They recognized the difficulty of norming across tests due to the different integrity types of the test. The different integrity types meant that each test observed held different comparisons with regard to the rigor of the questions as well as differences in the comparisons of match of test to the curriculum. The result of the study by Linn et al. (1990) indicated that accountability pressures caused broad ranges of achievement interpretations when states attempted to compare their achievement assessment results to that of the NAEP.

Therefore when states indicated achievement scores were improving and claiming their scores were higher than the national norm, in fact, they were not. States generalized the scores on their assessments against those of the NAEP scores. Linn et al. (1990)

reported that test norms for the states were changing and becoming loose in their proficiency requirements compared to that of NAEP; thus, the claims by the states of rising achievement compared to the national assessments were invalid.

NAEP validity studies. The research reported in this literature review describes the NAEP studies completed in relation to the use of NAEP and the validity or reliability in comparing the scores of NAEP. However, in this study, content validity is a central focus. For JCPS to determine whether the information from the tests is appropriate, they must be assured that the data are supported by their current standards. In 2009, the district was not utilizing the CCSSM. However, Carmichael, Wilson, Finn, Jr., Winkler, and Palmeieri (2009) reviewed multiple assessments including that of the study state and NAEP in comparison to the CCSS. Their 2009 study indicated that the NAEP aligned with the CCSS with an overall grade of “C” for their mathematics assessment. The Carmichael et al. (2009) study, however, compared the NAEP to a draft of the CCSS. The CCSS standards were in draft form until 2010.

A 2007 study completed by Daro, Stancavage, Ortega, DeStefano, and Linn explored the validity of the NAEP fourth- and eighth-grade mathematics assessments. Their goal was finding answers to multiple research questions. Of particular importance to this study were those questions focused on the following: (1) Did the NAEP framework offer appropriate content and skill coverage when compared with state and international assessments? and, (2) Did the item pool and design of the assessment reflect the NAEP mathematics framework (p. i)? Using a panel of expert reviewers made up of mathematicians and mathematics educators, the two questions were analyzed and determined to be “sufficiently robust” to provide validity to the conclusions about

national and state progress since 1990, as well as having “a reasonable framework” (Daro et al., 2007, p. iii). Additionally, in regards to the second question, “the item pool broadly aligns with the framework with some important exceptions” (Daro et al., 2007, p. iii). These exceptions came from the lack of focus in application to some of the strands represented by the NAEP framework; multiple suggestions were given to address the focus issue for specific strands (Daro et al., 2007).

Daro et al. (2007) determined that a relatively small percentage of the NAEP mathematics items were considered flawed (3% - 7%). However, a considerably higher number of items were considered marginal in content validity (23% - 30%) (Daro et al., 2007). The framework question applies generally to this study in its relevance to the alignment of NAEP to the CCSS. Similar to the Daro (2007) study, this study will utilize expert reviewers to determine the level of strand and standard alignment between the NAEP and CCSSM. Thus, validity support is provided for the use of the NAEP as a monitoring measure for JCPS.

Linn (1998) argued that interpretations of NAEP achievement levels should be approached cautiously. He suggested that the study evidence presents “a strong case for making more modest inferences and interpretations of achievement levels from NAEP than have frequently been made” (p. 23). Applicable to this study, Linn (1998) studied the “discrepancies between the assessments and the content standards” (p. 25). The focus of the study was on the achievement levels obtained by students and the validity of those scores. One of the specific components studied regarded the content standards set forth in the assessment frameworks compared to the content of the test. Linn (1998) determined that the content on the tests reflected that of the framework on a very broad level. He

encouraged using caution in accepting the NAEP scores based upon multiple results by defining a lack of coherence in the scale levels, item exemplars, and the content support.

Sireci, Robin, Meara, Rogers and Swaminathan (2000) studied the content validity of the 1996 NAEP science framework for eighth grade. The authors' intent was to determine if the science scores from NAEP "can be linked to the science content and skill domains the test is designed to measure" (Sireci, Robin, Meara, Rogers, & Swaminathan, 2000, p. 74). Using 10 science content experts, 60 items were analyzed for content and cognitive domain in a similar fashion to this study. A result of the study indicated that content experts have more difficulty in determining and agreeing upon cognitive domain than they do on the content standard.

Validity studies on NAEP extend further into score validity and how scores from NAEP should be viewed. In comparison, content validity studies are much fewer than those of score validation. The following section provides further information on score validity as it is frequently pointed out as an "issue" with NAEP results.

NAEP score validity issues. Hoping that linking their state assessments to the NAEP would help them gain more data on the level of their students' progress; states have long been interested in equating tests with one another (Bandeira de Mello, Blankenship, McLaughlin, & Rahman, 2009; Vockley, 2009; Timms, Schneider, Lee & Rolffhus, 2007; Kiplinger & Linn, 1994; Linn, Graue, & Sanders, 1990). Thus, equating tests is to treat them as if they are the same test and are interchangeable. Kiplinger and Linn (1994) conducted a study whereby state standardized assessment data between state and NAEP was equated and translated into NAEP scale terms. Data was collected for two different testing years allowing for stronger, more reliant evidence. The study raised

questions regarding the use of equating procedures on tests since the data provided information showing the instability of such methods (Kiplinger & Linn, 1994, p. 14). During their development, the intended use of the assessments was not to be compared to one another. If one were to compare the assessments, it is suggested that, “it would seem wise to assure, at a minimum, that the tests share a common content framework” (Kiplinger & Linn, 1994, p. 14). In equating, the assessments must have similar content matter but not necessarily similar difficulty levels. The Arenson (1999) findings indicated that any time equating is used between state assessments and NAEP, validity of the method is questioned due to the lack of certainty of similar content on the assessments.

Robinson (1997) contended that using state NAEP data in order to compare states against one another does not necessarily provide a valid picture to researchers and data analysts. The study showed demographics, such as: number of parents at home, parent education levels, community types, and state poverty rates, explain almost 90 percent of variation in NAEP test scores (p. 20). Thus, when viewing NAEP scores with the intent of comparing state information, demographic factors in each state must be considered when determining how comparable states are to one another.

Ideas for explaining test score differences in state assessments versus NAEP comparisons were presented in a paper presented by Wei, Shen, Lukoff, Ho and Haertel (2006). The authors described three possibilities for the discrepancies as being population differences, sampling, administration of the assessments, and content of the assessments (Wei, Shen, Lukoff, Ho & Haertel, 2006). By comparing content of state and NAEP assessments, researchers can ensure the reliability of the score and growth trend comparisons of the two types of assessments (Wei et al., 2006). Using the scaled to

content strands NAEP data, Wei et al. (2006) were able to determine that sampling was not an issue; however, since they were not able to disaggregate data for NAEP to item formats and strand areas, the researchers were unable to provide a causal explanation for the differences between state trend data and NAEP trend data (2006, p. 21).

Although NAEP has been used to compare norm scores and equate state standardized assessment proficiency levels to those at national levels, NAEP has been considered a “low-stakes” assessment as compared with states’ “high-stakes” assessments (Greene, Winters, & Forster, 2003, p. 3). Student perception of each test differs. For instance, Greene et al. (2003) noted that the Iowa Test of Basic Skills (ITBS) given to students in Chicago was considered high-stakes at some grade levels but not at others. In the case of students, a NAEP test may not be considered high-stakes since it was not correlated to grades, consequences to their school, or even a reward party.

It is clear that NAEP was not intended for individual student data collection or school data collection. Thus, the use of state data comparison to that of NAEP is difficult because a researcher cannot disaggregate data into a more specific set of comparisons. This shortcoming keeps researchers from obtaining “precise correlations” of data (Greene, Winters, & Forster, 2003, p. 5).

ACT EXPLORE

History of ACT EXPLORE

The ACT EXPLORE is intended to assess the academic progress of 8th graders and is used as either a part of ACT’s Educational Planning and Assessment System (EPAS) or as a lone assessment (ACT, 1995, p. 1). The EPAS system includes a set of tests given at specific points in a student’s education. The system provides teaches and

administrators a longitudinal view of student progress at several stages. In the first stage during eighth grade, the ACT EXPLORE test is given because students are preparing to enter high school. In the second stage, the ACT PLAN is given in 10th grade near the time students are making decisions about entering college. Finally, the ACT is given in 11th and 12th grade as students prepare for college or work after high school.

A goal of the EXPLORE program is in assisting teachers to develop instructional practices that address student needs (ACT, 1995, p. 2). According to research conducted by ACT (2006a, 2006b, 2006c, 2009), the use of the ACT EPAS system in some context (i.e. using the whole system or using only one test) shows benefits to students in achievement, college and career readiness, as well as enrollment and retention in college. One intent of EXPLORE, aside from determining college and career readiness, is to (1) inform teachers regarding instructional changes that may be necessary to help students, (2) supply students with information regarding possible courses they should consider, and (3) inform parents and administrators of the progress of the children tested (ACT, 2007, p. 1).

ACT EXPLORE Reliability and Study Relevance

Multiple states have adopted the Common Core for State Standards in an effort to ensure student are prepared for college and career after leaving high school. In 2010, ACT claimed a “leading role in the development of the Common Core State Standards” (ACT, 2010a, p. 1) not only by contributing research but also in contributing College Readiness Standards as a resource for the creation of CCSS. Since state adoptions of the CCSS, ACT has developed a classification system for identifying test items to the standards, clusters and domains of the Common Core State Standards to best estimate

student performance on the Common Core “in advance of state implementation efforts” (ACT, 2010a, p. 1). Thus, this study presented here is supported by ACT’s efforts in giving states a predictive measure by which they can compare students’ readiness for the CCSS to the CCSS standards.

Since the CCSS assessments have not yet been completed, the use of ACT comparisons to actual assessment items is a limitation. ACT is limited to comparisons of preparedness in only standard areas identified by CCSS and the interpretation of possible assessment items by ACT. Also, in their 2010 study, ACT revealed the limitation of states not yet using CCSS standards and, therefore, students were being taught using the various state standard systems (ACT, 2010b). A limitation of the ACT (2010b) study was due to the fact that ACT only had access to students who had taken the ACT and therefore there were fewer items that compare to the CCSS.

ACT has developed item reliability and validity for their assessments. Items are constructed and reviewed for educational and psychometric soundness, and then items are edited for fairness and bias. Items are then submitted as “tryout” items and placed into the booklets for students to attempt. Finally, response rates are recorded and assessed across multiple levels of demographic information to determine reliability and validity (ACT, 2007). ACT has established an internal reliability rate of 0.82 for nationally representative groups on the EXPLORE 8th-grade mathematics assessment with standard error of measurement at 1.71 (ACT, 2007).

Mathematical Strand Areas and Study Relevance

In an effort to ensure correlation with state curricula for the 8th-grade level for EXPLORE, ACT administers a national curriculum survey to assess state standards in the

areas of mathematics, reading, and science every three or four years (ACT, 2009, p. 1). Thus, ACT consistently monitors state policy changes and standards reform in order to give the most up-to-date information to educators.

The mathematics test consists of a multiple-choice test with 30 items given over a 30-minute time frame. ACT EXPLORE is divided into four content areas named and described with brief definitions below (ACT, 2007, p. 8):

- Pre-Algebra (operations with whole numbers, integers, decimals and fractions)
- Elementary Algebra (evaluation of algebraic expressions with substitutions, functions, one variable equations, and graphing)
- Geometry (different measurement systems, plane and solid figures, angles, parallelism, properties of triangles, and properties of circles)
- Statistics and Probability (data collection, representation and interpretation of data sets, use of and interpretations of graphs and charts of all types).

Relevant ACT Research

ACT test comparisons. Use of EXPLORE and the ACT PLAN assessments showed increases in average ACT scores when schools utilized both programs (ACT, Inc., 2009). Although this seems a logical correlation, it links to this study in considering the role of ACT in developing the CCSSM. Linking ACT and CCSSM frameworks lends validity to the use of ACT EPAS assessments as accountability measures. Additionally, for states adopting and using CCSSM standards immediately after its release, it is plausible that the students' ACT EXPLORE scores would be higher.

ACT and CCSSM share the common goal of aligning with student preparedness for college and career. In a 2005 case study, one state's mathematics proficiency

assessment had a passing rate of 56%, and ACT found that these same students did not meet proficiency on the ACT EXPLORE. (ACT, 2005, p. 4) This same result was reported for 10th graders in the state taking both the state assessment and the ACT Plan assessment with 56% of students passing the state assessment, but not indicating probable college success on the ACT Plan (ACT, 2005, p. 5). The ACT assessment can identify the areas in which students fall short in mathematics. Although it is an older study, ACT (2005) indicated they have repeatedly uncovered insufficiencies in college readiness standards and state standards in no less than 37 states (p. 2).

In 2011, ACT implemented a comparison study of college and career readiness of international and US students' performance using common content on the Program for International Student Assessment (PISA), ACT PLAN, and CCSS. In order to determine whether the US has internationally competitive standards for college and career readiness, ACT (2011) compared PISA and ACT PLAN since the sample populations of tenth graders were similar and both sets of standards aligned well (p. 6). The study involved only reading and mathematics since they are common content areas shared by PISA, PLAN, and CCSS. Although the methodology for determining the alignment of the content strands among the PISA, PLAN, and CCSS was not shared, ACT results indicated that the "performance standards of US college and career readiness in reading and mathematics are internationally competitive, and further validate the appropriateness of ACT's definition of college and career readiness as the right goal for US education (ACT, 2011, p. 7).

By determining that the content strands are aligned for PISA, Plan, CCSS, and ACT, (2011) the study also explored the strengths and weaknesses of US tenth-grade

students with other countries in the international data set. No documentation of equating scores or aligning proficiencies among the differing score sets was available; only the use of plausible values and composite scores for each set of tests were compared against other testing years to determine reliability of the data sets. Since it is reported that both the CCSS and ACT are internationally competitive, this study will be beneficial in checking the extent to which the ACT EXPLORE and CCSS mathematics standards are aligned for eighth-grade students.

ACT validity studies. In the *Technical Manual* for the ACT EXPLORE (ACT, Inc., 2011), a section for content validity is provided. The manual provides basic information regarding the establishment of content validity. The EXPLORE was based upon the ACT standards for college and career readiness which were determined through a careful analysis of multiple states' standards, textbooks, and secondary and postsecondary educator opinions. The standards were gathered from the 49 states that have published standards and were compared and disseminated for middle and junior high. ACT, Inc. (2011) states,

These three sources of information were analyzed to define a scope and sequence for each of the areas measured by EXPLORE. These detailed test content specifications have been developed to ensure that the test content is representative of current middle school and junior high curricula. (p. 47)

An extensive search of the ACT, Inc. website did not reveal any studies specifically focused on validity of any of the assessments they publish. The search spanned the 12 years of research and policy reports ACT, Inc. has published and allowed public access. Additionally, a search of the University of Louisville library system and the use of Google Scholar with multiple search queries revealed no validity information

in regards to ACT EXPLORE. The study state has had validity studies performed on the mathematics ACT but not the EXPLORE.

Benchmark Exams/Interim Assessments

Assessment Models

Formative assessment is a method teachers use to address student needs based on assessment results. Teachers are able to make instructional adjustments and educational decisions that aid achievement and growth in learning. Perie, Marion, Gong, and Wurtzel (2007) identified interim assessment as being different from formative assessments. They define interim assessments as “being given at the classroom level to provide information for the teacher, but unlike true formative assessments, the results of interim assessments can be meaningfully aggregated and reported at a broader level” (p. 1). Formative assessments take place during instruction where teachers can make immediate instructional changes, and interim assessments take place between formative and summative assessment. The current proposed study would use benchmark assessments as both formative and interim assessments. The benchmark assessments studied here are used in JCPS as an effort to gauge student progress on academic goal areas throughout the year. They are also used to identify strengths and weaknesses for teachers to address and strengthen instructional strategies as needed. The assessments are to provide feedback to teachers; in addition, the benchmarks are used to supply data to the district regarding content, which needs more support in curriculum or professional development. This concept fits into the definition of interim assessments provided by Perie et al. (2007), which will be the term used in this study to refer to the district’s benchmarks.

Despite rising popularity in interim assessments over the past 10 years, prior to the development of the policy work developed by Perie et al. (2007), studies on interim assessments were devoted to assessing the development and use of benchmark assessments. The term “benchmark assessment” was used in a similar format to that of the term “interim assessment” used by Perie et al. (2007). In a study designed to support the development of a useful benchmark assessment system, Davis, Caros, Grossen, and Carnine (2002) used benchmark assessments to link student performance with accountability measures on a larger state scale in order to determine the effectiveness of schools and teachers. In this way, the benchmarks are similar to interim assessments identified in the later 2007 policy work completed by Perie et al. (2007). Davis et al. (2002) conducted a study to determine the usefulness of a certain benchmark assessment to predict student performance on a high-stakes standardized assessment, the SAT9 and a high school exit assessment. Given that the reliabilities of the items were acceptable determined by a Cronbach’s alpha coefficient, the benchmark could be used to predict student outcomes on the final standardized assessments at the end of the year. The study also contained a parallel assessment used in fourth, sixth, and eighth grades to determine measures of internal consistency, which had results falling within acceptable limits as well.

The Philadelphia school district has used interim assessments as a form of accountability and instructional guidance for change since 2004 (Christman, Neild, Bulkley, Blanc, Liu, Mitchell, & Travers, 2009). During the 2009 study to analyze the impact of the benchmark exams, both mathematics and reading were assessed every six weeks. Christman et al. (2009) identified the primary goal of tracking student progress

was to increase student learning and to analyze whether the benchmark assessments used in Philadelphia were successful in doing this. By using student achievement scores on standardized assessments and teacher surveys from 2006 and 2007, Christman et al. (2009) determined that instructional leadership and professional climate were strongly linked to the higher student achievement scores. The leadership and climate variables were stronger predictors than a satisfaction with the benchmark exams variable (Christman et al., 2009). Although this study will not utilize teacher surveys, it is important to consider teacher affects on student achievement scores, especially with interim assessments since the teachers give these assessments in a non-standardized manner. Thus, this practice could affect the outcome and reliability of the use of the interim assessments when using data for an accountability measure.

In the Philadelphia study, qualitative analyses, through interviews with teachers and administrators as well as observations of meetings, revealed that teachers and administrators used the benchmark data to identify student gaps (Christman et al., 2009, p. 53). Thorough discussions and descriptions by the authors portrayed the importance of the meetings by the school leaders to ensure consistent meeting and actual analyses of benchmark data (p. 54). The Philadelphia study provided evidence that the use of interim assessments was warranted with the proper supports in place. It did not provide evidence that the assessments affected teacher's formative assessment decision making in instructional discourse (Christman et al., 2009).

Interim assessments offer schools a chance to predict overall outcomes on specific goals of end-of-year examinations as well as a predictor of the school or district outcome on adequate yearly progress (AYP) goals. In an effort to determine the usefulness of such

examinations in Philadelphia, a study conducted by Bulkley, Christman, Goertz, and Lawrence (2010) determined the context and expectations for such examinations as well as what support systems were in place to support the use of them. Bulkley et al. (2010) determined, from previous studies in 2005 and 2007 utilizing interviews and data from benchmark assessments conducted every six weeks beginning in 2004 for third through eighth grade, the assessments provided “feedback to teachers about their students’ success in mastering concepts and skills covered in the core curriculum during each five-week instructional period” (p. 193). The time frames were broken down into six-week increments in the following manner: five week instructional period, test day, and one week for revisiting goal areas referred to as a re-teaching week.

The other goal of using the benchmark assessments in this specific study was to set a pacing schedule, or a curriculum map (Bulkley et al., 2010, p. 193). In the study, teachers were given reports of student scores indicating not only scores for passing or not passing, but also an overall look at how students in a group did as a whole on certain concepts (p. 194). By incorporating a “re-teaching week” into the cycle, teachers were able to use the benchmark assessment as a formative assessment as they were able to review and re-teach concepts which students struggled on during the assessment. Although the study indicated that the use of the examinations was to “inform classroom instruction,” after the reports were given, district leaders indicated in their interviews that the examinations were used in an interim assessment manner at district level for predicting success rates on the state high-stakes examination (Bulkley et al., 2010, p. 200). They found that the system to inform instructional discourse and for informing school officials of success, or lack there-of, worked well for Philadelphia. The system

worked well because systems that were accessible to teachers were put into place. Additionally, professional development was given, and specific time was set-aside for teachers to address student needs according to the given data (Bulkley et al., 2010, p. 203).

In an effort to study the use of interim assessments to inform teacher practice, Goertz, Olah, and Riggan (2009) first reviewed multiple cycles of instructional improvement and found a commonality among the cycles in the use of “collection and interpretation of information about student learning.” (p. 9). Assessments were conducted to give teachers, administrators, policy makers, and other governmental officials information about student achievement. Goertz et al. (2009) posited that the timeframe for the use of the data gathered from the assessments was of greater concern. There was no true and distinct agreement among those who have written about formative and interim assessments regarding when the collected data should be used to address instructional change or gaps in knowledge shown by the tests (Goertz et al., 2009). Therefore, if an interim assessment is given, the data analysis must take place immediately afterward if the results are to affect teaching.

Use and Relevance in JCPS

The Classroom Assessment System and Community Access Dashboard for Education, CASCADE, used by the district began in the 2003 - 2004 academic year. The benchmarking assessments are given as a method to guide teachers in focusing on key focus points that are represented on state tests. The assessments are given approximately every six weeks throughout the school year as a response to “Streamline to Proficiency” (STP) (Dossett, personal communication September 28, 2011). The system

began exclusively in the mathematics content area but has since expanded into other academic areas and now includes an error analysis to help “teachers have an understanding of why students may have selected certain wrong answers” (Dossett, personal communication, September 28, 2011). The CASCADE assessment is a criterion-reference test such that content validity was established by alignment with the state’s Core Content for Assessment and reliability was established through gathering student data to ensure stability among administrations of the examination and items (Local District’s Accountability, Research and Planning, 2010).

Based upon the information given by Dossett (personal communication, September 28, 2011), this benchmarking technique tended to fall toward the previously defined interim assessments. Dossett explained that the CASCADE system has several purposes, including the following: 1) giving teachers immediate feedback after scanning assessments; 2) focusing classroom instruction based upon the key concepts to be assessed for the specific six week period; 3) giving administrators and educators predictive information for future achievement since the assessments are intended to test the same content which state developed standardized assessments test; and 4) giving students a chance to complete items which are similar in format to the high-stakes state assessment.

States are utilizing benchmarking systems to improve scores on state high-stakes assessments widely across the country (Christman et al., 2009; Clune & White, 2008; Goertz, Olah, & Riggan, 2009). Thus, the CASCADE system, developed by a team of district professionals is not a unique concept. Since the advent of NCLB in 2001, states have had increased pressure to produce strong scores and raise proficiency percentages

among students with the requirement that all students must reach proficiency in mathematics and language arts by 2014. In addition, with the adoption of Common Core State Standards in Mathematics throughout the nation, states are looking for strategies to prepare students for the new standards as well as ways to predict how students will perform on upcoming CCSSM assessments. With the CASCADE, a picture of content areas that are strong or weak for students in the district becomes clearer. Thus, when compared to a national assessment such as NAEP, administrators can view the content that requires further support. Since CASCADE is criterion-referenced and NAEP is both criterion-and norm-referenced, both examinations can be aligned with regard to content and cognitive complexity and then compared to determine possible predictability for students to do well on subsequent national assessments.

Mathematics Framework. District professionals originally developed the CASCADE mathematics framework. As curricula have transformed and evolved in the past years, “items have been added from various curriculum item banks and, recently an item bank was purchased by the district” to ensure a stronger correlation to the CCSS (Dossett, personal communication, September 28, 2011). The assessments were based upon a similar proficiency system as the standardized state assessment but were simplified from the broad topics noted on the state standards to better reflect the curriculum taught during the assessment period. To further standardize the benchmark assessments, district staff initially sought teacher feedback to address issues through questions and answers in order to develop a more valid benchmark assessment in relation to the state examination and the district curriculum. The JCPS Accountability, Research,

and Planning (2010) document reflects a 70% correlation between the CASCADE assessment system and the state standardized assessments.

Beginning in the 2011-2012 school year, the Kentucky Department of Education began altering the state standardized achievement test to reflect the changes in the CCSSM. Moreover, the state focused first on CCSSM in eighth grade in the spring of 2012. According to Dossett (personal communication, July 20, 2012), the test information will be released in mid-October of 2012. The benchmark assessment was also revised to reflect the changes on the state examination. Thus, based upon the information from the benchmark examination collected from the district, the following topics comprise the mathematics framework titles and are assessed in the following order:

- 1) Algebraic Thinking
- 2) Irrational Numbers
- 3) Radicals and Exponents
- 4) Linear Models and Bivariate Data
- 5) Geometry and Measurement

Each benchmark assessment contains from 13 to 19 multiple-choice questions and includes one constructed-response question. Each test is distributed to teachers with an answer key and recommendations about scoring the constructed-response questions with a specified number of points for each answer. The final page of the answer key contains suggestions for teachers with regard to possible student weaknesses depending upon the types of errors made by students.

The state's frameworks for the standardized assessment, of which the district's interim examinations are based, are identified with topic lists such as the following (Kentucky Department of Education (KDE), 2012a):

- 1) Number System (rational numbers and their estimations)
- 2) Expressions and Equations (radicals and integer exponents, proportional relationships, linear equations, solve and analyze linear equations and systems)
- 3) Functions (define, compare, analyze and model)
- 4) Geometry (congruence and similarity, Pythagorean theorem, real-world problems with volume)
- 5) Statistics and Probability (patterns of bivariate data)

Eighth Grade Mathematics

Eighth grade is a popular assessment grade level. NAEP, ACT EXPLORE and many state assessments target eighth grade as a grade level to assess student knowledge. Many social and physical factors affect student learning in the middle grades: social, emotional, and physical growth. Thus, concerns arise with regard to learning during these middle grades and to what extent eighth graders are prepared for high school. Declining or stagnate mathematics achievement scores across states have been revealed through a variety of reports from The International Mathematics and Science Study (TIMSS), ACT, and NAEP data (Schmidt, 2002). Countering these findings, a study conducted by the Center on Education Policy (CEP) (Chudowsky & Chudowsky, 2011) reported an increase in mathematics achievement by eighth graders in most states through academic year 2008-2009 (p. 3).

ACT supports the importance of focusing on eighth-grade achievement because students are not proficient in mathematics and language arts. They found a lack of college and career readiness beyond high school regardless of high school curriculum (ACT, 2008). In 2008, ACT reported, “our research shows that, under current conditions the level of academic achievement, students attain by eighth grade has a larger impact on their college and career readiness by the time they graduate from high school than anything that happens academically in high school ” (p.6). ACT (2007b) indicated a lack of high school readiness by eighth-grade students in mathematics where as much as 24% of time is needed in high school to re-teach pre-requisites for the new material in their classes (p. 20). These findings indicate the need for stronger and more rigorous mathematics in the eighth grade to align with the requirements of high school. This vertical alignment (ACT, Inc., 2007b) provides for a lower percentage rate of students needing to take remedial courses after graduating high school and gives teachers more time to teach rigorous and new material in high school classrooms.

Content strand focus. Developed over a two-year time frame and involving 49 states as well as four other countries which use preK-8 curriculum standards, NCTM developed its *Curriculum Focal Points* to guide administrators and educators in the mathematics strands and content that are needed at each grade level (NCTM, 2012a). Prior to the development of the Focal Points, NCTM developed a set of *Principles and Standards* in mathematics, which began with the *Assessment Standards for School Mathematics* in 1995, was revised in 1997, and released in 2000 as *The Principles and Standards for School Mathematics* (NCTM, 2012b).

These standards included two strands. The first strand focused on mathematic content, and the second strand focused on the processes for solving problems. The content strands were: number and operations, algebra, geometry, measurement, and data analysis/probability. The processes were: problem solving, reasoning and proof, connections, communication, and representation. The content strands were important in the alignment to assessments and CCSSM goals. The process strand was important in the identification of cognitive complexity.

Video data of classroom mathematics lessons from the TIMSS 1995 and 1999 studies (Jacobs, Hieber, Givvin, Hollingsworth, Garnier, & Wearne, 2006) were used to compare and assess the level to which teachers' lessons addressed the standards. After data were coded, Jacobs et al. (2006) found that teaching practices were not consistent with NCTM Standards (28). Teacher surveys indicated that participants were knowledgeable in the NCTM Standards and most confirmed their attendance at NCTM workshops or prior NCTM conferences. However, despite the claimed knowledge, the video analyses indicated that teachers did not in fact utilize either the NCTM content or process standards.

Strands Assessed pre- and post-CCSSM. The ACT (2007b) study regarding rigor of high school curriculum in addressing the needs for college and career, supports the need for a stronger set of standards. They found that students had not gained the knowledge necessary for being successful beyond high school. This finding was based on the "high percentage" of students requiring remedial mathematics courses out of high school (ACT, Inc., 2007b, p. 10). The study was not designed to prove or disprove the need for a set of national standards. However, there is an implication for such need since

the ACT is a voluntary common assessment taken by high school students across the nation. Although the CCSS is not designated as a core curriculum, it brings to the forefront whether the individual state standards actually prepare students for college (ACT, Inc., 2007b).

At the time of this literature review, information on CCSSM in various search engines and websites did not reveal recent research. A search of AERA journals revealed one article related to the comparison of pre- and post- assessment using CCSSM and previous mathematics and language arts strands. Porter, McMaken, Hwang and Yang (2011) compared the newly developed CCSSM with the NCTM requirements for mathematics. They completed a comprehensive content alignment between the CCSSM and NCTM standards using the Surveys of Enacted Curriculum (SEC) approach (p. 104). This approach compared mathematics content and cognitive demand by dividing them into 16 content areas. Using a formula for alignment index developed by Porter (2002) and deemed reliable through multiple studies (Porter, Polikoff, Zeidner, & Smithson, 2008; Porter, 2002; Polikoff, Porter, & Smithson, 2009; and, Porter, Kirst, Osthoff, Smithson, & Schneider, 1993), the mathematics standards revealed an average of 20% alignment between state standards and CCSSM standards in eighth-grade mathematics (Porter et al., 2011, p. 111). The greatest difference in content standards for individual states compared to the CCSSM across all grades involved the number of standards using procedures. There was a 38% emphasis in the eight- grade CCSSM mathematics on use of procedures (p. 112).

Another distinction made between the Porter et al. (2011) comparison was the mathematics alignment of CCSSM and NAEP. The research revealed only a 21%

alignment between content on NAEP and the new CCSSM (Polikoff et al., 2011). For this study, the NAEP comparison is important because it involves content alignment between NAEP and CCSSM. Thus, as a gauge of growth in achievement, one limitation of using NAEP involves the possible low content alignment, which could affect the level of validity in the measure since JCPS has begun using the CCSSM.

According to the study completed by Porter et al. (2011), the major topic areas, or strands, represented by CCSSM, along with the approximated percentage represented are as follows: number sense and operations (55%), measurement (17%), basic algebra (13%), geometry (7%), and data displays and statistics (6%) (p. 108). In this study, the NAEP, EXPLORE and the JCPS interim assessments will be reviewed to determine the percent at which each assessment is similar or different to the strands noted by Porter et al. (2011).

State Mathematical Standards at Eighth Grade

Comparisons

Concerns about the need for enhanced mathematics achievement and ways to strengthen guidance for teachers and students in order to improve school programs, Reys, Reys, Lapan, Holliday, and Wasman (2003) studied the impact of standards-based curricula on student learning (p. 75). The study provided information on eighth-grade mathematics achievement in three school districts in Missouri using a standards-based textbook with NCTM standards at the core of the curriculum compared to that of school districts utilizing other curriculum materials (p. 77). The sample had used either the standards-based curriculum or the alternate consistently for both sixth- and seventh-grade prior to the study. The research revealed only a few differences in achievement

scores, particularly in the areas of number, geometric sequence, spatial sense, and discrete mathematics (Reys, Reys, Lapan, Holliday, & Wasman, 2003, p. 87). Despite obtaining only a few statistically significant differences in the achievement scores, the differences were rather large in favor of the standards-based curricula schools (Reys et al., 2003, p. 88).

The National Science Foundation (NSF) invested more than \$90 million in mathematics curriculum development. Tarr, Reys, Reys, Chavez, Shih, and Osterlind (2008) studied NSF-funded textbook value and textbook teacher usage in mathematics classrooms versus non-NSF funded textbooks. The main conclusions in the study were that student achievement was significantly higher in classrooms where standards-based learning environments were established and consistently utilized the NSF-funded curriculum. The NSF-funded curricula embodies the NCTM standards and when teachers utilized the textbooks correctly the achievement results for students in those particular mathematics classes were much higher than those not in classrooms where teachers did not consistently utilize NSF-funded curricula (Tarr, Reys, Reys, Chavez, Shih, & Osterlind, 2008).

Chen, Reys, and Reys (2007) studied the mathematics strand of area and volume in mathematics textbooks across first grade through eighth grade in order to determine whether a strong alignment across states and countries existed for the same strand. They found significant differences in mathematics content, grade-level placement for each standard in the strand, and cognitive level of learning required across states and countries.

The information explored by each of these studies is important to this study because they identify discrepancies among standards and assessments on a national level. This finding is an overarching issue that guided the production of the CCSSM.

Common Core State Standards

History

In 1998, a study comparing the state and commercial assessments in the United States with assessments used in France and Japan cited issues in the US with regard to the rigor of the mathematics expected in the United States compared to other countries (American Federation of Teachers, 1998, p. 6). In an attempt to determine why the US showed favorable achievement in the early grades but a steady decline in achievement proficiency as the grades increased, when compared to students across the world, the American Federation of Teachers, (AFT, 1998) identified one issue. It reported that the “extremely broad and not very deep curriculum” as defined by TIMSS (p. 4) was a primary cause. The early grades had strong basic skills; however the content introduced as the grade levels increase was not as rigorous as that in other countries. As noted in the AFT study in eighth grade, the low level of content and level mastery in comparison with other countries was apparent.

Using surveys and other methods of data collection, researchers found that the demands of a 100% proficiency rate for all students in the core subjects by NCLB produced the outcome relayed “what gets tested is what is taught” (Stecher & Barron, 2001; Pedulla, Abrams, Madaus, Russell, Ramos, & Miao, 2003; AFT, 1998, p. 6, Wiley & Yoon, 2005; Resnick & Resnick, 1991). Since a more rigorous test alone will not raise achievement in the United States, clearly the standards must also become more rigorous.

A major limitation becomes clear. In using NAEP and ACT scores as a comparison tool, especially across states, the limitation is the differences in state standards. By developing a set of standards that are national, further data can be gained from the assessments. NAEP is intended to provide useful information to policy makers, school administrators, et al. with an interest in the educational progress of students in the United States. For some, this information may be useful in comparing rigor in education across the states, ensuring what is required in each state and fair. However, as was previously stated, NAEP is intended to measure what children should know by a particular age or grade based upon a framework set by the NAGB, not what each state's individual standards require of their students. Some educators have argued that, in an effort to meet Adequate Early Progress requirements, individual state standards and proficiency requirements have been lowered to assist the state in meeting the NCLB obligations (Cronin, Dahlin Adkins, Kingsbury, 2007; Schmidt, Houang & Shakrani, 2009, p. 5). Thus, the advent of a set of national standards has been introduced to alleviate the fairness of required learning standards (Schmidt et al., 2009, p. 5). In relation to a national set of standards, several states have opted to adopt one or the other of two offered assessments, which are currently being created to measure achievement of students as it relates to the new set of standards. This practice makes the achievement requirements more consistent across states.

In a report that examined national and international standards as well as the Common Core State Standards (although at the time they had only been a public draft), Carmichael, Wilson, Finn, Jr., Winkler, and Palmieri (2009) utilized a grading system to determine if PISA, TIMSS, NAEP, and the CCSS addressed a common set of standards.

This common set of standards, which were the grading benchmarks, were based upon a rubric designed by the authors of the study. The authors coded, and then developed, the rubric together in order to ensure the same content would be assessed, and the grade given would mean the same among all of them (Carmichael, Wilson, Finn, Winkler, & Palmieri, 2009). The mathematics strands assessed for the eighth grade were as follows:

- Arithmetic
- Ratios
- Geometry and measurement
- Problem solving (Carmichael et al., 2009, pp. 11-12)

In addressing the above content strands and developing a comprehensive grading rubric, Carmichael et al. (2009) determined grades for each of the CCSS and NAEP assessments. They (2009) included a review of strengths and weaknesses for each type of assessment. At that time, assessing the public draft of the CCSSM, the reviewers gave the CCSSM a grade of B. Strengths of the document were clarity and specificity, and weaknesses were, at that time, content and rigor (Carmichael et al., 2009, p. 21). Specifically, one issue cited problems in omitting important content according to their strands.

The authors also analyzed the NAEP. In this case the authors separated the grade levels, which was not done in the CCSSM review. This strategy reveals an inconsistency in the review system; however, the authors gave NAEP an overall grade versus a grade for each grade level, which is consistent with the technique utilized for the CCSSM review. In the eighth-grade assessment, “significant flaws” in content due to missing items in the authored specified strands for mathematics were cited (Carmichael et al., 2009, p. 23). Also cited as issues were (1) “unclear language” (Carmichael et al., 2009, p.

25) utilized in test items, and (2) a lack of clarity for teachers regarding the priority of certain strand areas. Overall, the authors believed that the NAEP should have received a C for the mathematics assessment.

Since the mathematics frameworks appeared similar on a broad scale, consideration should be given regarding what the individual standards for each area may include. This is important for each state due to the inclusion of mathematics content from the NAEP, ACT, and NCTM content or principles and standards frameworks in the CCSSM includes mathematics content from frameworks for NAEP, ACT, and NCTM. Both NAEP and ACT are norm-referenced examinations that provide administrators a picture of what can be obtained in performance with regard to the new CCSSM assessments. This study includes assessment items from NAEP and ACT. Therefore, in both studies, released items from the assessments will be analyzed alongside the current interim assessment utilized by JCPS. This analysis provides a frame of reference for which assessment is most strongly representative of the CCSSM.

CCSS Mathematical Practices

In addition to content standards, the CCSSM also include the standards for Mathematical Practices. The CCSSM practices have evolved from two previous publications developed from research and previous sets of practice standards for mathematics.

The National Council of Teachers of Mathematics (NCTM) identified Process Standards within its publication of the *Principles and Standards for School Mathematics* (NCTM, 2000). The publication was developed with the intent of guiding mathematics teachers in developing curricula that enable students to achieve certain levels of

conceptual understanding. Each of the named standards is applied to prekindergarten through grade 12. The five standards listed by NCTM (2000) are as follows:

- 1) Problem Solving
- 2) Reasoning and Proof
- 3) Communication
- 4) Connections
- 5) Representation

An additional publication released by the National Research Council (NRC) in 2001 described five strands of mathematical proficiency for students to learn mathematics. The following is a list of the proficiencies listed by the NRC (2001):

- 1) Productive Disposition
- 2) Procedural Fluency
- 3) Conceptual Understanding
- 4) Strategic Competence
- 5) Adaptive Reasoning

With both the NCTM (2000) and NRC (2001) standards for mathematics proficiency established, the creation of the CCSS Mathematical Practices included both sets of conditions. From these established frameworks, the CCSS Mathematical Practices evolved into eight themes that “provide the conditions under which students learn mathematics with deep conceptual understanding” (Hull, Miles, & Balka, 2012, p. 50).

The eight Mathematical Practices are listed below:

- SMP.1 Make sense of problems and persevere in solving them.
- SMP.2 Reason abstractly and quantitatively.

SMP.3 Construct viable arguments and critique the reasoning of others.

SMP.4 Model with mathematics.

SMP.5 Use appropriate tools strategically.

SMP.6 Attend to precision.

SMP.7 Look for and make use of structure.

SMP.8 Look for and express regularity in repeated reasoning. (CCSSO, 2012c, p. 53)

The eight practices as they are listed, appear concise and easily understood; however, Hull, Miles, and Balka (2012) relay that the “practices and supporting research are challenging to understand in regard to what they mean for students and teachers” (p. 50). Also supporting the challenge of interpretation of the practices and where teachers are to include them, Rasmussen et al. (2011) state,

The CCSSM itself provides little insight into how to integrate the mathematical practices with mathematical content, particularly across the many components of the mathematics education system that are implicated in this recommendation. As such, much research still needs to be done to address questions that arise. (p. 211)

Additionally, Rasmussen et al. (2011) set forth a call for those designing assessments, educational materials, and curriculum guides to take the practices into consideration by connecting them to mathematical content and instruction.

McCallum (2012), a co-writer of the CCSSM, explains that the practices are designed to describe the “various ways in which proficient practitioners carry out their work” (p. 7). McCallum (2012) describes the CCSSM content standards and practices explaining that they are intertwined. The practices are highlighted in certain content standards, at times, explicitly. For example, SMP.4: Modeling, is identified within

multiple high school standards (McCallum, 2012). Additionally, SMP.5: Using Tools Strategically, is identified in a seventh-grade content standard involving the solving of multi-step problems (McCallum, 2012). McCallum (2012) also describes multiple examples where the CCSS Mathematical Practices are used concurrently while students are working to solve problems. The practices and content were written to be used together and “were not intended as free-floating proficiencies, unconnected to content, nor are they uniformly applied over all the work that students do” (McCallum, 2012, p. 7). McCallum’s (2012) paper on the CCSSM reiterates the continuing theme of the literature; the practices are not understood the same way for every problem, in every situation, or in every grade.

Assessing the Practices

Confrey and Krupa (2010) prepared a report summarizing a conference organized by the Center for the Study of Mathematics Curriculum (CSMC), summarizing the discussions regarding the CCSSM. Confrey was cited as discussing several strategies for implementing the CCSS:

1. Phasing the implementation of the CCSS in a planned, purposeful, and coordinated way;
2. Articulating the expanding the underlying learning trajectories in the CCSS to guide assessment;
3. Re-visioning the relationship among the CCSS, curriculum materials, and classroom assessment to drive engagement, customization, and read;

4. Appropriating the 15% of a state's standards that do not have to comply with the CCSS and use it to define and deploy a broader college and career STEM agenda; and
5. Using longitudinal data systems to decipher and study curricular effectiveness as understood by the curriculum development and research community. (Confrey & Krupa, 2010, p. 5).

Of particular interest to this study, is the third strategy. During the conference, Burkhardt delivered a discussion of the third strategy. Confrey and Krupa (2010) summarize the plenary by saying, "his message to the group focused on the need to ensure that the CCSS are assessed in ways that promote, rather than inhibit, deep learning" (p. 5). Confrey and Krupa (2010) lend support to Burkhardt when supplying several recommendations at the end of their report of the conference. The authors recommend teachers design with the practices in mind, helping students to:

"Internalize the eight mathematical practices as the means to learn and understand the content standards. The practices sustain mathematics as the content evolves. As such, they make what student learn enduring and they ensure that students will continue to be prepared to learn new mathematics." (Confrey & Krupa, 2010, p. 10).

This statement in particular supports that of Hull, Miles, and Balka (2012) in their discussion of the purpose of the practices and the transformation from the NCTM Principles and Standards (2000) and the NRC Proficiencies (2001). The intention of all three sets of standards/proficiencies/practices is to increase students' conceptual understanding of mathematics to be able to apply it in differing mathematical contexts.

Two suggestions for acting on the recommendation to lead with the practices were made that particularly apply to the context of this research. First, one suggestion encourages curriculum and instructional resources to “make connections to the practices more explicit” and a second suggestion implores educators and assessment designers to “ensure that all new assessments evaluate students’ proficiency in the practices with connections among the content standards” (Confrey & Krupa, 2010, p. 10).

Krupa (2011) also prepared a report on the “Moving Forward Together: Curriculum & Assessment and the CCSSM” conference organized by the Consortium for Mathematics and Its Application. Multiple panels were formed to discuss the implementation of the CCSSM in instruction and assessment. One key recommendation resulting from the conference implored assessment designers to include the mathematical practices in assessments (Krupa, 2011). Krupa (2011) reiterated that research shows teachers choose instruction content and strategies based upon what they know will be tested and that schools often design intervention services for children in mathematics based upon preparation for tests. Thus, to be sure teachers will design instruction and curricular materials for students to be proficient in mathematics beyond procedural skills, stakeholders and the assessment consortia must design assessments with the practices in mind. An additional recommendation resulting from the conference indicated “there must be clear models of what it means for students to engage in the mathematical practices at different grade levels or for them to show proficiency of a mathematical topic using one or more of the mathematical practices” (Krupa, 2011, p. 10).

Multiple authors support the reasoning behind assessing the practices and the importance of doing so (Kepner & Huinker, 2012; McCallum, 2012); however, there is

little literature available regarding the development of the assessments that will test them. Websites such as the Illustrative Mathematics website or the Partnership for Assessment of Readiness for College and Careers (PARCC), provide examples of items that utilize the practices. The Illustrative Mathematics site provides the definitions of each of the practices with an example video of students working a mathematics problem while utilizing a practice. The PARCC site, as one of the consortia that received Race to the Top funding for development of one of the national assessments for the CCSSM, provides examples of possible constructed response test items. However, in regards to the design of an assessment item or the analysis of the alignment of an assessment item to the CCSSM, there is no literature or guide to define the specifics.

Importance of Research Exploring the Alignment of the Assessments to CCSSM

If alignment among the three assessments explored in this proposed study can be made, then states wishing to prepare students for the Common Core State Standards in Mathematics examinations may see the value of using interim assessments to check progress on the current state standards. Aligning NAEP and EXPLORE will allow for strengths to be determined in each assessment and, in particular, provide valuable information on specific accountability measures' validity to the curriculum which reflects the CCSSM.

ACT and NAEP, along with NCTM, have addressed some of the standards put forth by the CCSS. This point is imperative in the literature review for this study since it confirms the choice of utilizing the EXPLORE and NAEP assessments in alignment to the CCSSM and interim assessment utilized by JCPS.

NAEP to CCSSM link. A brief released by Achieve (2010) provided information regarding the comparison of frameworks between NAEP and CCSS. The intention of CCSS is primarily for directing instruction and describing what students need to know to become college and career ready. However, the intention for NAEP is to provide a research-based assessment for what students should know at a specified age level using all state standards as a guideline. The link between the two is likely the strongest for administrators and policy makers to consider when preparing for CCSSM assessments. Achieve (2010) analyzed the “rigor, coherence, and focus” of the comparison of CCSSM and NAEP. The findings confirmed that the CCSSM is similarly focused on skill, problem solving, modeling, and understanding levels in mathematics (Achieve, 2010, p. 2). The report also stated that “both the CCSSM and the NAEP Frameworks have similar expectations for students by the end of 8th grade” with a greater focus in algebra and equations (Achieve, 2010, p. 3). Achieve concluded, through their comparison, that CCSSM and NAEP have similar levels of rigor and expectations at each grade level; thus, students mastering the CCSSM should perform well on NAEP (2010).

ACT to CCSSM link. The CCSS initiative used longitudinal studies completed by ACT to identify the “knowledge and skills essential for success” in life after high school as well as using ACT’s College Readiness standards as a resource for creating the CCSSM (ACT, 2010a). In their 2010 study, ACT explored questions related to determining estimates of current performance using the CCSSM and ACT data even though states had not implemented the CCSSM yet (p. 1). Like this study, the ACT study was also used in determining the current state of students’ strengths and weaknesses on the CCSSM (ACT, 2010a, p. 1). After having clustered the CCSSM content into testable

and comparable sections (i.e. determining which CCSSM content standards compare to the ACT content strands), ACT was able to identify the percentage of 11th graders who met the proficiency standards set forth by the ACT assessment frameworks (ACT, 2010a, p. 2). Similarly, this study will also cluster assessment items based upon strand areas identified by CCSSM for eighth grade.

ACT identified only one-third to one-half of 11th graders performed at acceptable skill ranges on the mathematics and language arts CCSS domains, strands, and the clusters determined by ACT (2010a, p. 3). The ACT data was separated into the following strand areas that were similar to but not exactly the same as the current CCSSM frameworks noted above. The strand areas were number and quantity, algebra, geometry, functions, and statistics and probability. Concerns were noted in the level of understanding in foundations of mathematics. Specifically, concerns were mainly focused on number and operations, since it is evident upon review of the CCSSM that this standard is woven throughout all of the strand areas in all grade levels (ACT, 2010b, p. 7).

Although this study will also analyze alignment between the interim assessment used by JCPS and the ACT EXPLORE assessment, ACT has recently conducted its own alignment between the CCSSM and the ACT EXPLORE. ACT (2010b) states,

“For states initiating the process of adopting the Common Core State Standards, the report can help education and policy leaders: (1) Demonstrate the alignment between the new Common Core State Standards and the more commonly understood ACT College Readiness Standards and the ACT Course Standards, and, (2) provide evidence of current student performance relative to the Common Core State Standards by using scores from ACT assessments” (p. 1).

The report did not include a methodology for conducting this evaluation with only one reference to utilizing their experts in standards to develop this report. ACT (2010b), finding a thorough agreement of alignment between the two assessments, reported that there was 100% alignment between the EXPLORE standards and the eighth-grade CCSSM content standards (p. 4 and 5). The study also found an 88% alignment between EXPLORE and the eighth-grade Mathematical Practices for CCSSM (ACT, 2010b).

Study state to CCSSM link. Carmichael, Martino, Porter-Magee, and Wilson (2010) completed an assessment of the state standards compared to the Common Core in order to obtain an overall view of state standards content when held to the CCSSM standard of rigor and covered content. Carmichael et al. (2010) indicated Kentucky earned a “D” for overall mathematics organization, clarity and specificity, content and rigor (pp. 141-143). For this comparison, a group of content experts were utilized to grade standards based on a specified rubric to determine the state of the standards for each state. Carmichael et al. (2010) determined Kentucky, with a grade of “D”, was among the worst in the country indicating that the CCSSM were far superior to that of Kentucky’s Program of Study (p. 143).

More recently, the KDE conducted its own comparison of their standards to the CCSSM. A “Crosswalk” (KDE, 2012a) table showing the relationship between their own state standards and the CCSSM revealed grading for each of the main mathematics standards in the CCSSM. The KDE assessed themselves by first indicating the CCSSM standard and matching their own standards according to which standards applied to that specific CCSSM standard. There was no indication by the KDE with regard to the

method of determining this comparison, nor was there any indication of who completed the comparison or their credentials.

The state graded the state standards alignment to CCSSM using a zero, one, two or three. The numbers were represented with the following definitions:

- 0 = No match
- 1 = Weak match, major aspects of the CCSS not addressed
- 2 = Good match, with minor aspects of the CCSS not addressed
- 3 = Excellent match between the two document

According to the most current website information for the KDE and their system of analysis, when the eighth-grade standards were compared to the CCSSM, the state gave themselves 16 excellent matches, 12 good matches, and 5 weak matches. There were overall 33 CCSSM broad standards, and each Kentucky state standard was paired to one CCSSM standard. Some state standards fell into more than one CCSSM standard.

The Kentucky state standards' match to the Common Core State Standards was not completely clear. Although the study by Carmichael et al. (2010) is older than the crosswalk released by the KDE, it may be more accurate. This may be due to the Carmichael et al. (2010) not having had a relationship, which would form a conflict of interest during the study. Thus, a validity threat limitation should not interfere with the results of the Carmichael et al. (2010) study.

Content Validity

Messick (1989) identified a definition of validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other

modes of assessment” (p. 13). In addition to Messick’s (1989) initially published definition, there are multiple other definitions supported by other authors (For example: Messick, 1993; Nunnally & Bernstein, 1994; Suen, 1990; and, Walsh, 1995). Of all of these definitions, an operational definition can be formed for use in this study. Thus, content validity can be defined as the degree that assessment items reflect the intended purpose of the assessment. In this study, content validity of the three assessments, NAEP, ACT EXPLORE, and JCPS interim assessments, is established by showing the degree to which the assessments represent the CCSSM.

The American Educational Research Association (AERA), The American Psychological Association (APA), and The National Council on Measurement in Education (NCME) (1999), in their *Standards for Educational and Psychological Testing*, have placed significant emphasis on the importance of content validity in measurement processes. In fact, AERA, APA, and NCME (1999) have emphasized validity as the most essential quality in measurement. Thus, in this study the focus on the content validity of the three assessments is indeed an important characteristic.

Content Validity and Standards Alignment

By studying the alignment of these measures of achievement, the role of content validity in the assessments can be addressed. Moreover, this study uses NAEP mathematics as a data source for alignment. Although the study district may not use NAEP as an accountability measure, the NAEP is still viewed as a national comparison tool by many and projects growth in achievement. However, the issue of alignment with the standards compared to what is taught remains. In this case, these standards are the CCSSM. Additionally, if there is not an acceptable amount of content validity evidence,

the intended use of the exams is possibly jeopardized even despite the high level of assessment reliability established. A position statement released by AERA (2000) corroborates this argument.

Both the content of the test and the cognitive processes engaged in taking the test should adequately represent the curriculum. High-stakes tests should not be limited to that portion of the relevant curriculum that is easiest to measure. When testing is for school accountability or to influence the curriculum, the test should be aligned with the curriculum as set forth in standards documents representing intended goals of instruction.

Although the assessments evaluated in this study are not high-stakes for students, they may be considered high-stakes for the those held accountable.

Evidence of Content Validity in an Assessment

Content coverage. Particular to this study, content validity is established by comparing test items to the CCSS mathematics content standards. McMillan (2008) cites an example of establishing content validity of one high-stakes test. In order to establish strong evidence of the content validity of the test: “experts in the subject areas reviewed the tests and made systematic judgments about whether the items represented the content. These experts also made judgments about whether the percentage of items in different areas was appropriate and whether some areas that would be important were not on the test” (McMillan, 2008, p. 25).

In a U.S. Office of Special Education Programs (OSEP) (2005) document on validity evidence, Webb’s (1999) alignment tool is cited as one tool which adequately assesses alignment of items as well as producing “statistical indicators about the degree to which the test as a whole matches the standards” (p. 4). Additionally, content coverage is the first step involved in establishing content validity in the OSEP (2005) document. While the document does imply there are other tools for alignment OSEP (2005),

examples of content validity are established through the use of Webb's alignment tool. The document describes criteria of Webb's tool for alignment: depth-of-knowledge, range of knowledge and balance of knowledge. Proof of each criterion is given through the use of content experts by coding each assessment.

In this study, in order to establish content validity of the assessments, a tool for alignment must be used. By establishing a level of alignment of the three assessments to the CCSSM, a qualitative and quantitative evaluation of the degree to which the assessments represent the CCSSM will be presented.

Standards Alignment

In some cases, alignment of standards has been conducted to assist states in comparing their students' achievement to that of another state. In others, alignment studies have been completed to direct attention to areas in assessments that do not cover certain standards. In the case of this study, the alignment study will be conducted in order to determine the extent to which the CCSSM aligns with the NAEP, EXPLORE, and in the JCPS benchmark assessment. The alignment will focus on strands that are not covered on the assessments but are in the CCSSM. This strategy will inform administrators about the strengths or problem areas of included, excluded, or under-represented strand items on their own interim assessment as compared to the CCSSM.

Several models for studying degree of alignment of standards to assessments have been proposed. The following models have been approved by the Council of Chief State School Officers (CCSSO) (2002) and continue to be used in studies: (1) the Webb model; (2) the Surveys of Enacted Curriculum (SEC) model; (3) Achieve model; and (4) The Council for Basic Education (CBE) model. The CCSSO (2012b) has been particularly

involved in studying two models, the SEC and the Webb alignment tool. The CCSSO approved models are the focus of the current literature review since the CCSSO oversees the Common Core State Standards initiative, which is the focus of this study.

Webb's Alignment Tool (WAT)

In the Webb alignment model, reviewers are trained through a half-day session to identify standards and match them to assessment items (CCSSO, 2002, p. 2). Flowers, Browder and Ahlgrim-Delzell (2006) describe the Webb model as a procedure that “combines qualitative expert judgments and quantified coding for evaluating the alignment of standards and assessments (p. 202). After training, four to six reviewers first determine the depth of knowledge level of each of the standard objectives. Then reviewers code the items represented on the assessment, match them to the standard objective; and finally rate the item on the assessment for the depth of knowledge required from the student to complete the task (CCSSO, 2002). In order to develop the reliability and validity of the test, information from each reviewer is entered into a spreadsheet to be “statistically analyzed across the reviewers to produce statistics and tabular reports on four criteria of alignment for each standard: 1) Categorical Concurrence, 2) Depth Of Knowledge Consistency, 3) Range of Knowledge Correspondence, and 4) Balance of Representation” (CCSSO, 2002, p. 3). The CCSSO (2002) reported that reliability ratings across studies utilizing this method were consistently high at 0.60 to 0.90 (p. 3).

Utilizing his alignment tool, Webb (1999) studied the alignment of content standards to four states. The main goal of the study was in “developing a reliable and valid process for analyzing alignment among standards to assessments” (Webb, 1999, abstract). Webb (1999) found varying alignments among the states’ number of items

representing the number of standards identified in each strand area. In addition, Webb (1999) found low levels of cognitive complexity across all four state assessments. Webb (1999) determined that the tool for alignment of standards and assessments was indeed “viable for detecting the degree of alignment between assessments and standards and how alignment can be improved” (p. 28). This finding supports the strength of this study’s intention to provide feedback to the JCPS administrators and district developers with regard to improvement to gain data on student strengths and weaknesses in preparing for the CCSSM assessment.

Flowers, Browder, and Ahlgrim-Delzell (2006) successfully used Webb’s model of alignment to evaluate several states’ alternative assessments, meaning those designed for children with exceptionalities, to their standards. Their study found acceptable percentages of content alignment, although not strong. They also found a lack of alignment with regard to required cognitive complexity of assessment items to the cognitive complexity level indicated by the standards (Flowers, Browder, & Ahlgrim-Delzell, 2006). After training 13 content experts to be reviewers on the alignment process, the researcher had them independently code five items. After coding, the reviewers convened to discuss their results and to reach consensus on the alignment of the five items. Reviewers then coded the remaining items independently. The study revealed 80% to 94.1% intra-class rater reliability on content standards alignment as well as 69% to 97% intra-class correlation of rater reliability on DOK (Flowers, Browder, & Ahlgrim-Delzell, 2006).

Survey of Enacted Curriculum (SEC)

The CCSSO (2002) reported that the SEC model of alignment had been tested across 11 states and several large urban districts. The SEC model was utilized “to produce an alignment analysis of standards, assessments, instruction, or curriculum materials that were systematically categorized according to a common framework of content topics by cognitive demand” (CCSSO, 2002, p. 3). Differing from Webb’s model, which included four levels of depth of knowledge, the SEC model had five levels: 1) Memorize; 2) Perform procedures; 3) Communicate understanding; 4) Generalize/prove; and 5) Solve non-routine problems (CCSSO, 2002, p. 3). Similar to Webb’s model of alignment, SEC utilizes four content experts as reviewers who code each assessment item onto the two-dimensional matrix (CCSSO, 2002, p. 3). Currently available is an online process for teachers to enter data and obtain information regarding the alignment of curriculum and classroom assessments.

This study was not completely supported with the SEC model since it is more aptly used for analyzing classroom instructional practices (Blank, Porter, & Smithson, 2001). Although several components of alignment in the SEC model relate to this study, it is not generally used for standards and achievement test alignment. Multiple studies utilize the SEC model and report it as a strong alignment model for classroom instruction to standards, classroom instruction to classroom assessment, and support for professional development (e.g. Smithson & Blank, 2006).

Although the SEC model was used for content alignment in mathematics for the 2007 NAEP, inter-rater reliability was 0.58 across the documents. However, Blank

reported at the time that typical rater reliability was 0.50 (2007, p. 2). Blank (2007) also noted that SEC does not typically lend itself to inter-rater reliability analyses. The study indicated the alignment of the NAEP assessment to the content standards in the frameworks as being a minimum of one standard deviation higher for NAEP than the typical state standards to its own state assessment (Blank, 2007, p.2).

Achieve

As with the other two models, a set of content expert reviewers were used to determine the level of alignment between standards and assessment items (CCSSO, 2002, p. 4). “There are five criteria: 1) Content centrality; 2) Performance centrality; 3) Challenge; 4) Balance; and 5) Range.” (CCSSO, 2002, p. 4) The process involves content experts performing the initial qualitative analyses of the five criteria. The Achieve staff then complete a statistical analysis of the data on each of the criteria and then on the entire assessment (CCSSO, 2002).

In a 2001 study of Massachusetts’s benchmark assessments, Achieve (2001) indicated no inter-rater reliability statistics. The alignment strategy however was able to identify strengths and weaknesses of the assessments. This study, intending to indicate strengths and weaknesses of NAEP, EXPLORE, and the JCPS interim assessments, is relatable to the Achieve (2001) study. The study on Massachusetts’s benchmark assessments gives precise indications of areas for improvement on the assessments. Moreover, by making improvements on the assessments administrators can better prepare students for their high-stakes assessments (Achieve, 2001). Similar research conducted by Achieve (2003) was conducted with regard to one district’s assessment and standards alignment. Achieve followed the same format as their 2001 study with no indication of

inter-rater reliability among its experts but a clear picture of strengths and weaknesses of the alignment.

The Council for Basic Education (CBE)

The CBE alignment model is different from others in that the process will “identify test items or framework specifications that match benchmarks and record the degree of match in content and performance level” (CCSSO, 2002, p. 5). In this model, content pairs of experts review assessment items to “determine degrees of match” on a rubric (CCSSO, 2002, p. 5). In comparison with the other three alignment models, the CBE model utilizes a rubric and exemplars to assist with reliability of collected information. CBE also utilizes other assessments, such as NAEP, to give feedback to administrators regarding progress on standards in application to the assessment.

Alignment Research Applicable to this Study

Among those studies that used alignment research focused on the CCSSM are two recent studies related to this study. Porter, McMaken, Hwang, and Yang (2011) aligned the CCSS with states’ standards across all grades. The findings noted “considerable differences” among the state content standards for both language arts and mathematics CCSS (Porter et al., 2011). Although Kentucky was not included in their study, the NCTM standards were a part of the study. The study indicated the average level of alignment for states was similar level to that of the alignment between the NCTM standards and the CCSSM. This study exhibits an example of incorporating an alignment method using the CCSSM as a basis for comparison. However, the study by Porter et al. (2011) did not compare assessments, nor did it compare national assessments, as is the case in this study.

In a preview of determining assessment alignment to the CCSS, Carmichael, Martino, Porter-Magee, and Wilson (2010) conducted a study focused on multiple assessments. The NAEP was one of few assessments that included both national and international student samples. The NAEP earned an overall “C” in mathematics (Carmichael, Martino, Porter-Magee, & Wilson, 2010, p. 25). The method for determining grades involved a group of coding experts in each content area and a “common grading metric upon which all standards could be measured” (p. 10). Thus, the method for determining a grade for each of the assessments reviewed was unexplained. However, using content expert coders led readers to believe a qualitative study took place.

A recent study completed by Schmidt and Houang (2012) using data from all state standards and focusing on grades 1 through 8, responded to alignment studies completed by Porter, McMaken, Hwang, and Yang (2011) and comments referring to the Porter et al. (2011) study by Cobb and Yang (2011). The intent of the study was to answer the lingering and widely expressed question: Will the CCSSM make a difference in student performance? (Schmidt & Houang, 2012) Within the publication, Schmidt and Houang (2012) recognize that the answer to this central question cannot be fully answered until the CCSSM has been fully implemented and the Common Core State Assessments have been released. In the Schmidt and Houang (2012) study, standards from all 50 states were compared to the CCSSM focusing on coherence and focus of the standards. Schmidt, Wang, and McKnight (2005) define coherence as standards “articulated over time as a sequence of topics and performances that are logical and reflect, where appropriate, the sequential and hierarchical nature of the disciplinary content from which the subject

matter derives” (p. 528). Additionally, Schmidt and Houang (2012) later refer to the TIMSS to define focus as the “number of topics covered at each grade that was also aggregated over the first eight grades, by counting the total number of topic-by-grade combinations” (p. 295).

Similar to this study, the research involved the coding of standards and cognitive demand. The results of the study revealed Kentucky as one of the lowest states covering the CCSSM. Schmidt and Houang (2012) cite the reasoning for the low ranking, identified as “Least like CCSSM” (p. 301), as the lack of coverage of standards topics at grade levels not matching that of the CCSSM. Specifically, it was explained that Kentucky, in 2008 when the data were gathered, was one of eight states that included multiple standards in a certain grade level that the CCSSM did not have identified. Thus, Kentucky had a low level of congruence with the CCSSM (Schmidt & Houang, 2012, p. 302).

Additionally, similar to this study, cognitive demand of the CCSSM were analyzed. Schmidt and Houang (2012) determined that “61% of the intended level of cognitive demand associated with each topic was at the most basic knowledge-memorizing definitions and performing routine procedures levels. Only 3% reached the highest level” (p. 303). It should also be mentioned that the eighth grade, as should be expected, had the highest level of high cognitive demand topics, but was still only around 10% (Schmidt & Houang, 2012).

In an effort to determine whether the CCSSM may have an effect on student achievement, Schmidt and Houang (2012) utilized data from the 2009 NAEP. The examination of the NAEP data was an effort to attempt to identify any relationship

between achievement on the NAEP and levels of state standards congruency. The study, when using a linear regression model, did reveal statistically significant results of a positive association of congruence to the CCSSM and NAEP scores (Schmidt & Houang, 2012). However, it should be mentioned that the Schmidt and Houang (2012) study did not analyze congruence between the 2009 NAEP and the CCSSM. Although the Schmidt and Houang (2012) study referred to the analysis as determining congruence of the standards to the CCSSM, determining congruence, in effect, is the same as determining alignment such as that done in this study.

Cognitive Complexity

History and Relevant Research of Cognitive Complexity Models

A review of literature on cognitive complexity models revealed references to multiple taxonomies that have been developed over the years since the publication of Bloom's Taxonomy of Educational Objectives (1956). The following taxonomies, or hierarchies, were referenced multiple times in comparison to others, which were not included here, which were referenced two or fewer times.

With a committee of 33 members, Bloom created and tested a method for determining educational objectives. Those applicable to cognitive domain are described in his Taxonomy of Educational Objectives (Bloom, 1956). Bloom (1956) identified the level of cognitive demand required within classroom objectives in a hierarchical, ascending manner. These cognitive domains "provide the clearest definition of educational goals expressed in terms of descriptions of student behavior" (Reeves, 1990). Bloom developed his hierarchy of cognitive domains to be used during the planning of instruction to draw attention to the behaviors students would express in relation to the

curriculum (Reeves, 1990). By using the taxonomy, teachers could build curriculum that required critical thinking from students and build comprehension of the content. Within the taxonomy, Bloom developed a set of categories, beginning with the most simple of elements and key words, indicating what level of learning/understanding would be needed for a student to respond. However, Hess, Jones, Carlock, and Walkup (2009) identify the use of verbs as a limitation of the taxonomy developed by Bloom. The authors indicate there are often verbs repeated throughout questions that indeed do not match the level of complexity by identifying the verb alone (Hess, Jones, Carlock, Walkup, 2009). Thus, within assessments, identification of cognitive domain could be made more difficult.

Enright, Allen, and Kim (1993) expanded Bloom's (1956) taxonomy for use in analyzing test items in conjunction with item difficulty. They narrowed the hierarchy created by Bloom (1956) to three skills measuring cognitive demand to indicate whether or not the student knows, uses, or integrates their content knowledge (Enright, Allen, & Kim, 1993). The complexity analysis study of NAEP science questions included both judgments of difficulty and cognitive demand needed for answering the questions (Enright, Allen & Kim, 1993). Several frameworks that are discussed within this literature review were used to determine question difficulty and cognitive complexity (Bloom, 1956; Glaser, Lesgold, & Lajoie, 1987; Messick, 1984). Enright et al. (1993) posited that item complexity and the cognitive level needed to answer questions should be somehow correlated. However, they found that "there was no evidence that the items' cognitive demand was related to item difficulty" (Enright et al., 1993, pp. 26-27). The

authors suggest that the idea of cognitive demand, at that time, should be more specifically defined and explored in relation to item complexity.

The study by Enright et al. (1993) is important to this study because the complexity level framework was driven initially by Bloom's taxonomy (1956). The framework for assessing cognitive complexity in this study is also based upon Bloom's taxonomy (1956) since the taxonomy and educational objectives identified by Bloom are also embedded in the Webb depth-of-knowledge framework.

Benton, Tremaine, and Scher (2004) studied item difficulty in computer-generated questions embedded in a distance computer-learning course. They cited a hierarchy from Enright et al. (1993) along with their own semantics for a system for analyzing semantic and cognitive demand of computer-generated questions. The portion of the Benton et al. (2004) study that relates to that done by Enright et al. (1993) references item difficulty due to the wording utilized in that item, as well as concepts tested (Benton et al., 2004). Additionally, the study by Benton et al. (2004) identified a lack in relationship between the items and the concepts being tested; thus, the standards were not strongly represented by the test items. Moreover, the study concluded that cognitive complexity levels of standards and the test item difficulty did not match (Benton et al., 2004).

Marzano (2001) also developed taxonomy for identifying cognitive demand characteristics from assessment questions used by several other researchers (Webb, 2002a; Lombardi, Seburn, Conley, and Snow, 2010). Marzano's (2001) taxonomy included six levels: (1) retrieval, (2) comprehension, (3) analysis, (4) knowledge utilization, (5) metacognition, and (6) self-system thinking. Similar to Bloom's

taxonomy, Marzano's levels apply to educational discourse rather than identification of cognitive level of assessment questions. However, Lombardi et al. (2010) narrowed Marzano's taxonomy by utilizing only the first four levels in the hierarchy (p. 6). The research by Lombardi et al. (2010) revealed that the indicators were reliable and "elicited more immediate responses with ease and objectivity." (p. 15) Thus, the Marzano taxonomy allowed researchers to analyze cognitive demand of items quickly and, according to Lombardi et al. (2010), allowed those assessing cognitive demand a concise method for choosing levels of demand.

Marzano and Kendall (2007) released the most recent version of Marzano's taxonomy. The new taxonomy is also represented in a two-dimensional format. The three domains of knowledge included: psychomotor procedures, mental procedures, and information (Marzano & Kendall, 2007, p. 13). The second dimension included six levels of cognitive demand referred to by Marzano and Kendall (2007, p.13) as mental processing. The cognitive demand dimension included the following domains: retrieval, comprehension, analysis, knowledge utilization, metacognitive system, and self-system. This new two-tiered system was a culmination of their previous work.

Webb (1997, p. 15) defined depth of knowledge as "the number of concept connections and ideas a student needs to make in order to produce a response, the level of reasoning, and the use of other self-monitoring processes." However, his 1997 research monograph did not identify the specific elements of the depth-of-knowledge (DOK) levels, as we know them today. Webb thoroughly examined the expectations of what is needed to consider full alignment of curriculum to assessments. He indicated that verbs should not be the only identifier in determining item complexity but, instead, "each level

should be dependent upon how deeply students understand the content in order to respond.” (Hess, 2006, p. 4)

Porter (2006) developed a two-tiered system for aligning standards, curricula, and assessments. One part of the tier is the level of cognitive demands based upon the topics presented. Porter’s Survey for Enacted Curriculum was designed to assist teachers in obtaining data about their teaching with the initial tier of the system identifying links of classroom lesson plans and classroom assessments to the strands of the standards. The second tier reported the overall level of cognitive demand represented by the lesson plans and assessments. Jones and Tarr (2007) summarize cognitive demand concepts in the following way: “Cognitive demands distinguish memorizing; performing procedures; communicating understanding of concepts; solving non-routine problems; and conjecturing, generalizing, and proving.” (p. 7) Available on the internet are surveys for teachers to enter information regarding specific standards and the cognitive demand level for which they teach that standard area (Blank, 2005; WCER, 2011). The result of utilizing the online tool allows teachers to “address educators’ needs for comparable, reliable data and analyses of math and science instruction.” (Blank, 2005, p. 5)

Utilizing Bloom’s taxonomy, Hancock (1994) studied and analyzed item construction as it was linked to cognitive complexity. The study results revealed items were constructed with high levels of cognitive demand (Hancock, 1994). The level of cognitive complexity for multiple-choice questions and constructed-response questions were both high. It is important to note that, in the Hancock study, the researcher was the only rater of test items. Herman, Webb, and Zuniga (2005) and Webb (1997) noted that

having a single rater does not support reliability in a study; thus, Hancock's (1994) work is considered questionable.

During alignment evaluations, cognitive complexity of a test item should be equal or greater than the cognitive complexity needed to learn the standard for which the item was designed to assess. In previous research on assessment of cognitive complexity and alignment of standards, the cognitive complexity of the test questions on standardized assessments were lower than the indicated standards (Resnick, Rothman, Slattery, and Vranek, 2003; Webb (1999, 2002a). Webb (1999, 2002a) conducted studies comparing state assessments by using his alignment hierarchy and found lower cognitive demand on the state assessments. The conclusion supported in several studies regarding cognitive demand, was that cognitive demand on assessments was lower than cognitive demand implied in state standards (Polikoff, Porter, and Smithson, 2011; Resnick, Rothman, Slattery, & Vranek, 2003; Webb, 1999, 2002a).

Anderson and Krathwohl (2001) refined a model that furthered the knowledge dimensions presented in a previous taxonomy. The added knowledge dimensions allow teachers or administrators to track the type of knowledge required from students: factual, conceptual, procedural, or meta-cognitive. Paterson (2002) used the Anderson and Krathwohl (2001) two-dimensional system of cognitive demand to analyze an online assessment system. Paterson (2002) intended to determine whether the verbs used during the learning of mathematics could identify levels of cognitive skill required to learn the mathematics. Paterson's (2002) conclusions indicate that the assessment items focused mainly on the two lowest cognitive levels and that there is doubt that online assessments can assess skills at higher levels of cognitive demand. Additionally, Paterson (2002)

recommends that, to use online assessment systems, rewording of items would need to take place in addition to reconfiguring the presentation of the items in order to assess at the top four levels of Bloom's Taxonomy.

Theoretical Constructs of Cognitive Complexity

Cognitive complexity has been defined using different terms. For the purpose of this study an operational definition is described as the determination of whether assessment items are appropriately challenging for the standard or standards the test is intended to assess. Thus, there should be a match of cognitive demand required to answer test items as is required to learn the standards. Terms used to reference cognitive complexity include: performance centrality (Resnick, Rothman, Slattery, & Vranek, 2003, p. 6); depth of knowledge (Webb, 1997a); level of demand (Achieve, Inc., 2006) and cognitive demand (AERA, 2006).

Webb's alignment of depth of knowledge includes four levels. Level 1 (Recall) is the lowest level of knowledge that requires the use of a memorized fact or understanding of a simple procedure. Level 2 (Skill/Concept) requires a student to use a simple procedure but also requires the student to determine where to start or which process to use. Level 3 (Strategic Thinking) requires students to plan a process give reasoning or evidence for their answer, or analyze a graph. The final level, Level 4, the highest level of thinking represents a response to an extended-response question requiring students to plan, produce, reason, or analyze.

Smith and Stein (1998), Webb (1999), Porter (2002), Anderson et al. (2000), and Anderson and Krathwohl (2001) developed frameworks for aligning standards with instruction, learning materials, and/or assessment items. A commonality of each

alignment framework was to determine the level of cognitive complexity of assessment items or standards. No matter the alignment process followed, a team of reviewers must reach some level of consensus regarding each item or standard.

The American Federation of Teachers (1998) compared five mathematics assessments to the NAEP content frameworks in mathematics to evaluate the tests with regard to classification of content and difficulty level of each item. The difficulty levels of the items were determined based upon three levels: Easy, Middle-Level, and Hard. Each level was defined as follows:

- Easy items require students to recognize and substitute numbers into a given formula. The student can complete the item without having to know relationships or synthesize information.
- Middle-level items require students to formulate a solution plan. They require thinking and the coalescing of knowledge. They usually require students to produce additional information before the final solution, and many require some generalizing.
- Hard problems require creation of an abstract model, understanding the problem, and what the problem requires. The problem demands effort since it is not well defined and the student cannot look at it and immediately know what to do. The context is meaningful and necessary to solving the problem. The student may have to establish a procedure and draw upon logic, theory, and proven principles. A hard problem requires students to know which theorems are relevant and how to apply them. (American Federation of Teachers, 1998, pp. 11 and 12)

Webb was one of the expert panel reviewers assisting with the study (AFT, 1998). As the knowledge levels increase in the AFT study, it becomes clearer which levels of Webb's DOK are linked in the cognitive process and level of thinking required. By utilizing the definitions, rater reliability was very high. There was very "little disagreement about the levels assigned to each item, and consensus was reached on 99 percent of all items on all tests." (American Federation of Teachers, 1998, p. 11)

In the AFT (1998) study, out of the five assessments that were rated with regard to item difficulty, none of the assessments included questions that were considered hard. Even a test, with all constructed-response items, contained no questions with an item difficulty level defined as hard (American Federation of Teachers, 1998, p. 15). The lowest percentage of easy questions was on a test where 79% of items were defined as easy (American Federation of Teachers, 1998).

The NAEP assessment utilizes a different construct for determining cognitive complexity. In NAEP assessment design, cognitive complexity is measured on three levels: low complexity, moderate complexity, and high complexity. The NAEP 2011 Framework (NAGB, 2010) indicated that the following percentage of testing time at each level is: moderate complexity 50%, low complexity 25%, and, high complexity 25% (p. 38). These percentages are identified to create a balance among the levels of complexity (NAGB, 2010, p. 38). The NAGB (2010) defines the levels of complexity which are summarized below:

- Low-complexity items expect students to recall or recognize concepts or procedures specified in the framework. Items typically specify what the student is to do, which is often to carry out some procedure that can be performed mechanically. The

student is not left to come up with an original method or to demonstrate a line of reasoning. (NAGB, 2010, p. 38)

- Items in the moderate complexity category involve more flexibility of thinking and choice among alternatives than do those in the low-complexity category. The student is expected to decide what to do and how to do it, bringing together concepts and processes from various domains. Students might be asked to show or explain their work but would not be expected to justify it mathematically. (NAGB, 2010, p. 42)
- High-complexity items make heavy demands on students because they are expected to use reasoning, planning, analysis, judgment, and creative thought. Students may be expected to justify mathematical statements or construct a mathematical argument. Items might require students to generalize from specific examples. Items at this level take more time than those at other levels due to the demands of the task, not due to the number of parts or steps. (NAGB, 2010, pp. 46 and 47)

The theoretical construct for the ACT EXPLORE mathematics assessment items is to “emphasize quantitative reasoning rather than memorization of formulas or computation skills.” The technical manual indicates that the items “cover four cognitive levels: knowledge and skills, direct application, understanding concepts, and integrating conceptual understanding.” (ACT, 2007, p. 5) The frameworks do not explain or define the levels. This study will determine the extent to which the released items match the cognitive complexity levels identified by Webb (2002c). Those levels are identified as Level 1 through Level 4 or, respectively, Recall, Skill/Concept, Strategic Thinking, and Extended Thinking.

The department of education in the state in which the study is located cited Hess's (2006) exploration of cognitive complexities as a source they used as a tool in designing their achievement tests. Hess (2006) commented, "Webb's Depth Of Knowledge (DOK) Levels are also being used more and more by local schools and districts to develop curriculum materials and performance assessments to demonstrate learning." (p. 3) Thus, the Hess (2006) statement lends support for utilizing Webb's DOK levels in this study.

Theoretical Framework of this Study

Selected Frameworks for Alignment

Many researchers agree that alignment must consist of both subject matter alignment with state standards, also referred to as content match, as well as depth of cognitive demand to answer the items (LaMarca et al., 2000). With high-stakes assessment serving as a key criterion in a state's status as a pass or fail for Adequate Yearly Progress requirements (NCLB, 2002), aligning assessments and standards has become an increasingly important in determining the validity and reliability of the scores representing the achievement of the students in public schools (LaMarca, 2001, p. 4). Thus, student assessments developed by state agencies or districts must clearly represent the adopted standards. Kentucky was one of the first to adopt the Common Core State Standards, and as such, in preparation for the upcoming CCSS assessment, their current accountability and diagnostic measures must align to the CCSS as well.

Standards Alignment. In this study, it is the goal to align test items to the CCSSM standards across NAEP, EXPLORE, and JCPS's interim assessments. If the items on each assessment in this study are weakly aligned to the CCSSM, then their use as tools for accountability measures will be problematic. Strong alignment is vital to the

integrity of this study because, without doing so, no viable correlation would exist among the standards studied on the assessments and the items on the eighth-grade CCSSM. In short, the utility of the information gathered from the comparison of the CCSSM to the assessments is based upon content validity.

Cognitive Complexity. Many studies use Webb’s depth of knowledge as a method for identifying cognitive complexity (Brown & Conley, 2007; Flowers, Browder, & Ahlgrim-Delzell, 2006; Porter, 2002; Roach, Elliot & Webb, 2005; Webb, 1999; Webb, 2002b). By incorporating depth of knowledge to identify cognitive complexity, this study “indicates a minimum level of consistency between the cognitive demands of state standards and the cognitive demands of the assessment items.” (Webb, Alt, Ely, Cormier, & Vesperman, 2005, p. 24)

Webb’s (2002c) DOK design was chosen because the study state has used the design previously (KDE, 2007). An additional reason for the choice is that the CCSSO has shown support for Webb’s (2002) version of the Depth Of Knowledge levels by sharing multiple studies utilizing the WAT (La Marca, Redfield, & Winter, 2000; Rabinowitz, Roeber; Schroeder, & Sheinker, 2006; Vockley & Vockley, 2009; Webb, 1997b). With the CCSSO as the main contributor and publisher of the Common Core State Standards, DOK is a logical choice as the basis for alignment and review. The following table presents application of Webb’s Depth of Knowledge to assessment items:

Table 1

Webb's (2002c) Depth Of Knowledge for Mathematics

Recall DOK 1	Skill/Concept DOK 2	Strategic Thinking DOK 3	Extended Thinking DOK 4
<p>Recall of information like a fact, definition, term or simple procedure</p> <p>Performing a simple procedure from a given formula.</p> <p>A one-step problem.</p> <p>Possible verbs used in the task: “identify,” “recall,” “recognize,” “use,” and “measure.”</p>	<p>Engagement of some mental processing beyond a habitual response.</p> <p>Requires some decision making by the student.</p> <p>Requires students to show work or describe the process.</p> <p>Requires more than one step.</p> <p>Possible verbs used in the task: “classify,” “organize,” “estimate,” “make observations,” “collect and display data,” and “compare data.”</p>	<p>Requires reasoning, planning, using evidence, and a higher level of thinking than in DOK 1 and DOK 2.</p> <p>Usually requires explanation of thinking.</p> <p>The task generally has more than one possible answer and requires the student to justify their answer.</p> <p>The task may require: citing evidence and developing an argument for the concepts.</p>	<p>Requires complex reasoning, planning, developing, and thinking over a more extended period of time.</p> <p>Students are required to relate ideas within the content area or across content areas and select one approach amongst many for completing the task.</p> <p>The task may require: designing and conducting an experiment, critiquing, or finding relationship among related concepts and phenomena.</p>

(Webb, 2002c, pp. 4-5)

Conceptual Framework

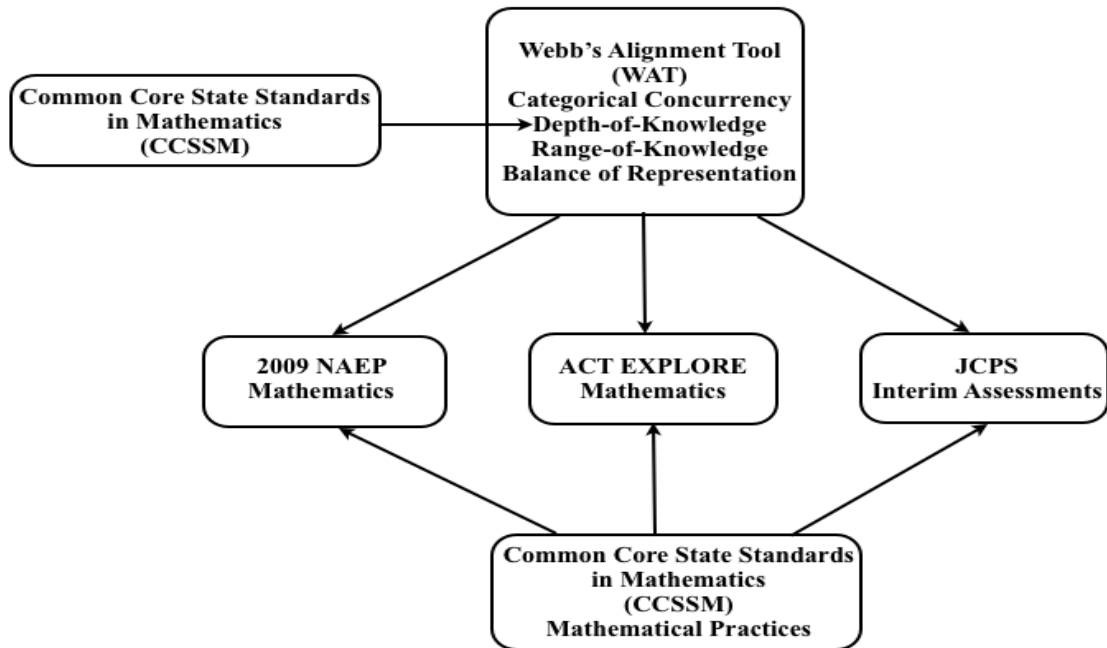
The level of representation of the CCSSM content for eighth-grade on the NAEP, ACT EXPLORE, and interim assessments, must be determined using an alignment method with previous research support. This study utilizes the WAT, which has four alignment criteria. In Figure 1, the conceptual framework of this study is depicted. Initially the three exams will undergo the process of the WAT: Categorical Concurrency and Depth-of-Knowledge. Both of the first two criteria are completed using content experts. The remaining two criteria of the WAT, Range of Knowledge and Balance of

Representation, are calculated based upon the information gained from the Categorical and DOK criteria. However, to be able to complete the DOK match, the CCSSM content standards must also be examined to determine DOK levels for each standard listed. Since the CCSSM also includes a list of mathematical practices, this study will also include a comparison of these CCSSM Mathematical Practices with the assessment items of the three exams. The completion of the WAT and the CCSSM Mathematical Practices comparison will identify the exam that shows the highest content alignment and the largest representation of the practices.

Figure 1

Theoretical Framework Model for Alignment of Assessment Items to the CCSSM

Alignment for Assessment Content Validity



Conclusions

Chapter II Summary

To summarize, this literature review discussed the three types of assessments explored here and the history of each: ACT EXPLORE, NAEP, and the JCPS interim assessment data collection system. The review also included information about how the reliability of each of these assessments was established and to what level of reliability each has achieved. For ACT EXPLORE and NAEP, any comparisons of tests to state assessments revealed lower proficiency scores as acceptable in many states than those of the NAEP. Additionally, it became apparent that mathematics content strands covered by each assessment and the study state presented differing levels of importance given to each strand.

Comparison of each of the assessments requires a study of their alignment to the CCSSM frameworks, the balance of representation for each of the strands, and item DOK. As was indicated, NAEP and ACT EXPLORE were not designed to reflect a specific curriculum taught by teachers; thus, they are more likely to reflect the future assessments based on the CCSSM. Producers of the CCSSM assessment will not be able to design the test based on a ACT EXPLORE have established reliabilities and been tested repeatedly and used in research studies, it makes sense to use those assessments as logical tools in comparing the interim assessments in JCPS.

Research Questions

- 1) To what extent do the (1) eighth-grade mathematics released items in the 2009 NAEP assessment, (2) ACT EXPLORE released items, and (3) items from the

Jefferson County Public School District interim assessments align with the eighth-grade CCSSM?

- 2) To what extent are the CCSS Mathematical Practices represented in each strand of the eighth-grade 2009 NAEP mathematics assessment released items, ACT EXPLORE, and the Jefferson County Public School District interim assessments?
- 3) Which of the tests shows the most content validity and which shows the most mathematical practices represented compared to CCSSM?

CHAPTER III

METHODOLOGY

The initial section of this chapter describes the design of the study and includes a restatement of the purpose and research questions. The next section describes the population of the study, the methods for obtaining the sample, and the sample itself. The third section describes the instrument used for alignment of NAEP, ACT EXPLORE and the interim assessments with the Common Core State Standards in Mathematics. Included in the description of the instrument is the method for establishing its validity and reliability for use in this study. Following the description of instruments is a section that describes the procedures for conducting the alignment and for collecting data. This chapter also includes a section describing the method for analyzing the data collected. Finally, the chapter will describe limitations of the study as they stand prior to conducting research.

Design of the Study

Restatement of Purpose and Research Questions

The purpose of this study was to determine the alignment of the 2009 NAEP, ACT EXPLORE, and JCPS interim assessments with the 8th-grade CCSSM. The Webb Alignment Tool (WAT) has been used frequently for studying test alignment among standards and assessment items. Webb's method is divided into four parts: (1) categorical concurrence; (2) depth of knowledge consistency; (3) range of knowledge correspondence; and (4) balance of representation; thus, the WAT will supply

information in regards to the amount of content validity each assessment shows when compared to the CCSSM. Along with the content standards and assessment of the items to identify the CCSS Mathematical Practices will also take place. Additionally, this study intends to present current information regarding each of the assessments by using a principal components analysis.

This study will answer the following research questions:

- 1) To what extent do the (1) eighth-grade mathematics released items in the 2009 NAEP assessment, (2) ACT EXPLORE released items, and (3) items from the Jefferson County Public School District interim assessments align with the eighth-grade CCSSM?
- 2) To what extent are the CCSS Mathematical Practices represented in each strand of the eighth-grade 2009 NAEP mathematics assessment released items, ACT EXPLORE, and the Jefferson County Public School District interim assessments?
- 3) Which of the tests shows the most content validity and which shows the most mathematical practices represented compared to CCSSM?

Study Design

This study was a mixed methods research design. In this study, the quantitative principal components analysis (PCA) uses a sample of student test answers. Although the PCA is one quantitative measure used in this study, a qualitative component exists. A qualitative coding of the content of tests items in three assessments will take place. By using several content experts, qualitative analyses will be utilized during the use of the Webb Alignment Tool. Additionally, the end result of the qualitative analysis of written documents during the WAT coding process is an empirical analysis representing the

amount of content validity in each assessment. Supported by Teddlie and Tashakkori (2009), the use of both qualitative and quantitative components in a study represents a mixed methods research design.

Teddlie and Tashakkori (2009) describe five types of mixed methods research designs. This study represents two of these designs. First, a parallel mixed methods design is described as the qualitative and quantitative methods occurring simultaneously and the “planned and implemented qualitative and quantitative phases answer related aspects of the same questions.” (Teddlie & Tashakkori, 2009, p. 151) The PCA and WAT can be used to answer portions of the first research question at the same time; thus, the specific type of design represented is a Parallel Mixed Methods design. However, Teddlie and Tashakkori (2009) identify another mixed methods design that includes the parallel design, the Conversion Mixed Method design. Initially a parallel design occurs but “mixing occurs when one type of data is transformed and analyzed both qualitatively and quantitatively; this design answers related aspects of the same questions.” (Teddlie & Tashakkori, 2009, p. 151) This study utilized both quantitative and qualitative data. Quantitative data were obtained from both the factor analysis and by converting the qualitative data gained from coding standards and cognitive complexity during use of the WAT into quantitative results. Additionally, qualitative analyses took place through meetings with the content experts after the coding of the CCSSM practices.

Patton (2002) states that “qualitative findings grow out of three kinds of data collection: (1) in-depth, open-ended interviews; (2) direct observation; and (3) written documents” (p. 4). The definition of a qualitative study is represented in this study by using Webb’s Alignment Tool to analyze items in assessment documents; thus, meeting

Patton's (2002) third data collection method, written documents. This study represents a qualitative component and meets Patton's (2002) definition of a mixed methods design.

Application to this study. In the past, studies about state standards have focused on: comparisons of state standards (Carmichael, Martino, Porter-Magee & Wilson, 2010); alignment of state standards to assessments (AERA, 2003; Flowers, Browder & Ahlgrim-Dezell, 2006; Polikoff, Porter & Smithson, 2011; Polikoff, Porter, Zneidner & Smithson, 2008; Resnick, Rothman, Slattery & Vranek, 2003-2004); analysis of state standards with regard to their cognitive complexity (Herman, Webb & Zuniga, 2005; Webb, 1997a, 1997b, 2002a); and comparisons of different assessments (Timms, Schneider, & Rolfus, 2007; Webb, 2002b). LaMarca, Redfield, and Winter (2000) made a strong case for alignment among state standards, "valid and meaningful data-based decision making depends on the degree of alignment between standards and assessments" (p. iii). This idea supports the primary goal of this study in determining the level to which the NAEP, ACT EXPLORE, and the JCPS interim assessments are aligned with the Common Core State Standards in Mathematics. Several studies have verified the importance of having quality assessments that are fair, valid, and reliable; however, none of these assessments were focused on an overall view of multiple assessments at one time while determining the relationship among each of them to the CCSSM.

Using the WAT alignment process as an analysis of content validity allows this study to present findings in empirical formats. With the addition of using a principal components analysis to determine the strands represented by the items in the ACT EXPLORE and the JCPS assessments prior to alignment to the CCSSM, this study was identified as a mixed methods research design.

Population and Sample

Population

This study will be most generalizable to those using the NAEP and ACT EXPLORE as accountability measures and to those participating in assessments designed to monitor achievement of their students.

Sample and Sampling Procedures

The sample for data collection will be taken from three assessments employed in local school district, JCPS. The 2009 8th-grade NAEP released mathematics items are accessible via the NCES website; an ACT EXPLORE assessment was previously obtained from a member of school JCPS; and the JCPS interim assessments will be requested from the district after the study meets the University of Louisville Institutional Review Board requirements for completing research. After completing a request for data through the district's proposal for research protocol and the district's research permissions documents, the researcher will collect all data for research.

Table 2 shows the number of items from each assessment and the distribution of the item type. All items were from eighth-grade mathematics assessments.

Table 2

Number of Items and Item Type

Distribution of Sample

Item Type	2009 NAEP	ACT EXPLORE	2011-2012 JCPS Interim Assessments
Multiple Choice	22	30	81
Constructed Response	12		8
Total Items	34	30	89

Instrumentation

Description of Instrument

Webb’s (1999) tool, the WAT, will be employed in this study. The tool includes four criteria for alignment: Categorical Concurrence, DOK Consistency, Range of Knowledge, and Balance of Representation. The following definitions describe each of criteria:

1. Categorical Concurrence provides a general indication of alignment if both documents (standards and assessment) appear to incorporate the same content. It is judged by determining whether the assessment included items measuring content from each standard.
2. Depth Of Knowledge Consistency describes the expected cognitive requirement of each standard and assessment item. In this case, the standard DOK of the assessment item is determined to be below level, at the same level, or above the level required when the item is compared to the required

complexity of the standard. For each level of DOK, the following level definitions are given below:

- Level 1 (recall) - Recalls information such as a fact, definition, term, or a simple procedure, as well as performs a simple algorithm or applies a formula.
 - Level 2 (skill/concept) - Engages in some mental processing beyond habitual response. Requires students to make some decisions as to how to approach the problem or activity.
 - Level 3 (strategic thinking) - Requires more reasoning, planning, using evidence, and a higher level of thinking than in the previous two levels. This level of complexity does not come from a question that has multiple answers but because it requires more demanding reasoning.
 - Level 4 (extended thinking) - Requires complex reasoning, planning, developing, and thinking most likely over an extended period of time. The extended period of time is a distinguishing factor only if it requires higher order thinking, not because multiple repeated steps are used.
3. Range of Knowledge Correspondence - This criterion is used to judge whether a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need to correctly answer the assessment items/activities.

4. Balance of Representation - This criterion is used to determine whether certain standards have been given more emphasis on an assessment than others. (Webb, 2007, pp. 10-14)

In order to determine which 8th-grade CCSSM strands are represented most in each assessment, the four criteria of the Webb model will be utilized. Table 3 describes the acceptable, weak and unacceptable levels for each of the criteria. Following this guideline, the study will identify strands and standards most represented on each assessment.

Table 3

Alignment Levels Using the Four Criteria

Alignment Level	Categorical Concurrency	Depth of Knowledge	Range of Knowledge	Balance of Representation
Acceptable	6 items per standard	50%	50%	70%
Weak	4 to 5	40% - 49%	40% - 49%	60% - 69%
Unacceptable	Less than 4 items per standard	Less than 40%	Less than 40%	Less than 60%

(Webb, Alt, Ely & Vesperman, 2005)

The following definitions are presented to further explain the concepts in Table 3. The definitions are quoted from the training manual for Webb’s Alignment Tool with the following caveats for each describing the use of the criteria in this study.

- “Categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents.” (Webb et al., 2005,

p. 121) Webb determined that six items of each standard is appropriate for a fair and reliable test.

- Depth of knowledge consistency is met between standards and assessment “when at least 50% of targeted objectives are hit by items of the appropriate complexity.” (Webb et al., 2005, p. 122)
- Range of knowledge is determined by the number of items that represent each standard; thus, it is “used to judge whether a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities.” (Webb et al., 2005, p.123)
- Balance of representation “indicates the degree to which one objective is given more emphasis on the assessment than another.” (Webb et al., 2005, p. 124) The intention of the balance of representation measure is to describe whether an assessment has an equal number of items for each standard. Additionally, in this study, balance of representation will describe the level to which each strand is represented on the assessments.

Webb, Alt, Ely and Vesperman (2005) presented the rubric for acceptable limits for each of the criterion used in the WAT to show alignment. They used several methods for determining the required passing rates. The Categorical Concurrence criterion was determined based upon the methodology of Subkoviak (1988) using means and standard deviations of cutoff scores. Any fewer items than six would require students to get all but one answer correct in that particular section of the assessment to qualify as proficient for that standard; thus, the requirements of that assessment would be very stringent.

To determine the cutoff score of the depth-of-knowledge criteria, the manual describes the 50% cutoff as conservative. They say it “is based on the assumption that a minimal passing score for any one standard of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding objectives” (Webb et al., 2005, p. 122). They also explain the range-of-knowledge criteria acceptable level “is based on the assumption that students’ knowledge should be tested on content from over half of the domain of knowledge for a standard” (Webb et al., 2005, p. 123). Finally, the balance of representation criteria is determined based upon an index value determining the level of distribution of items among standards.

Instrument Reliability

This study uses mathematics experts to provide alignment information using the Webb model of alignment. Therefore the reliability of the instrument relies heavily upon the degrees of agreement of the raters. Porter, Polikoff, Zeidner, and Smithson (2008) performed a study on rater reliability in alignment studies. The study used the SEC as the alignment tool; however, the focus of the study was to determine the number of content experts to use for coding versus the reliability of the instrument itself. The authors discovered differences in reliability of content agreement due to the raters’ backgrounds (Porter, Polikoff, Zeidner & Smithson, 2000, p. 3). Textbook publishers (Buckendahl, Plake, Impara, & Irwin, 2000), college faculty and high school teachers (Herman, Webb, & Zuniga, 2005) each rated the alignment of standards and assessments differently. Porter et al. (2008) indicated that training and raters’ backgrounds had a profound effect on rater reliability. In addition, the number of raters should be considered for inter-rater

reliability. No definitive number of individuals for coding has been identified as the most appropriate for the highest level of reliability; however, five or six coders would be sufficient for using the model for alignment (LaMarca, Redfield, Winter & Bailey, 2000; Porter et al., 2008; Webb 1997a & 1999).

Porter et al. (2008) indicated a lack of generalizability of the SEC model of alignment results to other models because the procedures differed. The SEC was directed toward educators in reflecting upon what they teach and how they teach it. This study was not intended to describe what teachers do or how they teach. Further, the SEC model requires educators to enter data into a computerized system so that experts employed by the SEC can rate them.

Given the frequent use of Webb's procedures in alignment studies and the thorough capacity of the model's application to contexts other than standards to tests, test publishers as well as the Council of Chief State School Officers (CCSSO) have employed the Webb model as an alignment method (Case, Jorgensen, & Zucker, 2004, p.7).

Data Collection Procedures

Institutional Review Board (IRB) approval will be obtained from both the university and the JCPS. This study will be submitted as exempt status since the research is using existing data that is already de-identified.

Alignment Procedures

Following IRB approval, tests and/or released test items will be obtained from NAEP, ACT EXPLORE, and JCPS. After obtaining the tests or items, the study will use five expert coders to use Webb's method for assessment-to-standards alignment. Each expert will be a mathematician, mathematics educator, or classroom teacher having

taught more than five years, with at least two connected to the middle school grade level. The content experts will be familiar with the Common Core State Standards in Mathematics prior to this study. Five expert reviewers serve as an acceptable number of reviewers according to Webb (1997a, 1999) and LaMarca et al. (2000).

Prior to coding, the researcher will explain the method for coding to all five raters. A spreadsheet developed specifically for the assessments to be studied, was developed and based upon Webb's design located in the WAT training manual (Webb, 2005). Prior to the meeting, coders will learn how to record their findings of CCSSM alignment to items and DOK level for each item. Additionally, a discussion of identifying CCSSM mathematical practices of each item will take place prior to coding.

After training, coding will take place for three items on each assessment on both standards alignment and depth-of-knowledge alignment in order to ensure agreement and understanding among all coders. After the training session, all coders will code the remaining items independently. Any items having less than 80% agreement will be discussed regarding inconsistencies. If consensus cannot be reached, the raters will elect either to not use the specific item or to use the alignment established by the largest number of coders.

Data Collection

Data will be collected using the alignment tool with regard to standards alignment as a comparison to the eighth-grade mathematics content portion of the CCSSM. The alignment data will be gathered from the NAEP released-items database on the internet, an ACT EXPLORE assessment employed in the 2011 academic year, and the assessments used by JCPS in 2011-2012 and released only for the purpose of this study.

In addition, a principle components analysis will be performed on the assessments included in the alignment study. Thus, a sample of data from students' answers on items used from one form of the ACT EXPLORE will be used. Additionally, to keep the sample and effect size large, all student cases that have all answers for all five of the JCPS interim assessments will be used. Those cases that have blank answers or do not have data for all five JCPS assessments will be removed from the sample to avoid errors from missing data. The same process will be used to select student cases from the ACT EXPLORE data. Any students with blank items will be removed from the sample to address errors that would skew the statistical analysis.

Data Analysis

Data analysis will be completed in accordance with the recommendations made by the U.S. Office of Special Education Programs (OSEP) (2005) document on content validity evidence. Since the purpose of this study was to determine the feasibility of using the NAEP, ACT EXPLORE, and the JCPS interim assessments as progress monitoring or accountability measures, the analyses of data takes the form of supporting content validity of each assessment compared to the CCSSM. Therefore, I will describe the methods in accordance with the recommendations of the OSEP (2005) regarding evidence of content validity. The areas described are as follows: content coverage; response processes; and test internal structure. The research questions are stated below followed by data analysis methods.

- 1) To what extent do the (1) eighth-grade mathematics released items in the 2009 NAEP assessment, (2) ACT EXPLORE, and (3) items from the Jefferson

County Public School District interim assessments align with the eighth-grade CCSSM?

- 2) To what extent are the CCSS Mathematical Practices represented in each strand of the eighth-grade 2009 NAEP mathematics assessment released items, ACT EXPLORE, and the Jefferson County Public School District interim assessments?
- 3) Which of the tests shows the most content validity and which shows the most mathematical practices represented compared to CCSSM?

Methods Used to Analyze Data

Research Question One

In order to answer research question number one, a determination of extent of content match between standards and assessment items, an analysis of the content coverage of the items on each exam must be performed (The U.S. Office of Special Education Programs, 2005). The following analysis method describes the process of providing evidence that substantiated the level of content validity in each of the assessments when compared to the CCSSM.

Content Coverage

This initial indicator of content validity, Content Coverage, can be described as the percentage of test content that covers the standards set forth by the state standards document, the CCSSM.

WAT analysis. The WAT allows expert coders to determine the standard for which each test item applies, in addition to each item's cognitive complexity level required for an answer by the student. Alignment data will be compiled using coder

information and the alignment levels displayed in Table 4. The results of the data analysis from coding will be displayed according to the expectations noted in Table 4. Each of the categories represented by the WAT will be reported for each assessment aligned with the CCSSM.

Principal component analysis. In addition to using the WAT to align the NAEP, ACT EXPLORE, and the JCPS interim assessments, the researcher completed a principal component analysis on the ACT EXPLORE and the JCPS interim assessments to determine the extent to which the mathematic strands of each assessment matched the strands of the CCSSM. Principal component analyses for each of the two assessments could reveal the strands represented in each exam separately by identifying the factor structures for the items.

Research Question Two

To provide answers to research question number two, each assessment item must be analyzed to determine which, if any, mathematical practice is identified in the item. Content experts using a method similar to that used for identifying DOK did this. Inter-rater reliability was again established for this identification process.

Research Question Three

After the alignment process and the identification of mathematical practices in each of the assessment items for each exam have been completed, a review of the data was completed. By doing this review of data, identification of the assessments that have the highest level of content alignment and the most mathematical practices represented provided information regarding the assessments that were indicated as the strongest in each area.

Reliability

By utilizing content experts to code the standards and test items, the WAT can provide inter-rater reliability. Thus, by completing statistics to find inter-rater reliability coefficients, reliability will be established. Reliability was projected to be sufficient due to the use of the WAT. In addition, by completing a PCA of the ACT EXPLORE and interim exam test items, internal consistency reliability using Krippendorff's alpha was determined.

Both NAEP and ACT EXPLORE have pre-established reliability procedures. NAEP developers field tests items with small groups of students before using them in an assessment. These items are then checked and a reliability coefficient is established. For establishing reliability content raters are used and intra-class correlations are determined in content validity. ACT EXPLORE uses a similar system for establishing reliability of their mathematics assessment. They use content specialists to write items and then assess content specifications and statistical specifications. ACT EXPLORE assesses level of difficulty and the minimum acceptable level of discrimination (ACT, Inc., 2011, p. 6).

Limitations

One limitation of this study was the lack of availability of multiple forms of the ACT EXPLORE as well as the limited availability of NAEP items. Additionally, the NAEP data from 2011 were not available this study began. Therefore, the alignment data could only be obtained from the 2009 released items available on the internet.

An additional limitation lies in the limited number of students who have taken the specific form of the ACT EXPLORE. This limitation reduces the number of students in the sample for drawing data. Related to this limitation is the fact that not all teachers give

the JCPS interim assessments in the same way. Since the interim assessments are not standardized, assessments are given in various ways with different emphases. In addition, not all teachers are consistent in inputting data which is a limiting factor in the sample.

The WAT Depth-of-Knowledge criteria requires a comparison of levels of cognitive complexity between assessment items and standards items. Upon an initial glance, the CCSSM content standards are not of high level DOK. This would skew the information gained from the comparison. To alleviate this issue, research question number two was developed to review the number of mathematical practices which are addressed in each of the assessments.

Also, the JCPS interim assessments are not tested for reliability. Thus, the information provided by the PCA may be unreliable. However, with the alignment data established by this study, the validity of the test will be established.

An additional limitation of this study applies to the ACT EXPLORE sample. The copy of the assessment obtained for the purpose of this study is dated 2008, but was utilized during the fall of 2010. A more recent version reflecting the CCSSM is available, but it was not obtainable at the time of the study.

Finally, given that assessments are in the process of changing to reflect the CCSSM, this study represents only a snapshot of information. Changes will be made to the CCSSM as more information is gained regarding student achievement. Thus, this study may need to be repeated within a short timespan.

CHAPTER IV

RESULTS

Chapter IV has the results of the quantitative and qualitative analyses of the JCPS Interim Assessments, ACT EXPLORE, and 2009 NAEP mathematics items. This chapter is organized into six sections: (a) introduction, (b) principal components analysis of the JCPS interim assessments and ACT EXPLORE data, (c) secondary analysis of each assessment for content validity using the WAT, (d) tertiary analysis of CCSS Mathematical Practices, (e) qualitative data from the final meeting with content experts on the process for identifying the mathematical practices, and (f) conclusions.

Introduction

The purpose of this study was to determine the level of content alignment represented in each of three tests (2009 NAEP mathematics items, ACT EXPLORE, and JCPS interim assessments) when compared to the CCSS in Mathematics. In this study the alignment was performed using Webb's Alignment Tool (WAT) and Webb's four criteria for alignment: Categorical Concurrence, Depth-of-Knowledge, Range of Knowledge, and Balance of Representation. The three research questions for this study are:

1. To what extent do the (1) eighth-grade mathematics released items in the 2009 NAEP assessment, (2) ACT EXPLORE, and (3) items from the Jefferson County Public School District interim assessments best align with the eighth-grade CCSSM?

2. To what extent are the CCSS Mathematical Practices represented in each strand of the eighth-grade 2009 NAEP mathematics assessment released items, ACT EXPLORE, and the Jefferson County Public School District interim assessments?

3. Which of the tests shows the most content validity and which shows the most mathematical practices represented compared to CCSSM?

To answer these questions, the researcher used a mixed-methods design performing: (a) qualitative document analysis completed by content experts, (b) quantified analysis of the coded data, and, (c) a quantitative statistical analysis of the factor structures of the JCPS interim assessments. The sample for this study consisted of items from three mathematics assessments: 34 released mathematics items from the 2009 NAEP, 30 items from an ACT EXPLORE examination, and items from five of the 2011-2012 JCPS interim assessments.

For this study, the researcher used a principal components analysis through the factor analysis function in SPSS to determine the representation of strands revealed by factor structures on the JCPS interim assessments. The researcher then trained content experts to use the WAT. Experts identified the standards represented for each assessment item and the depth-of-knowledge (DOK) level that was represented in each assessment item. After the content experts coded each assessment item the data were used to determine the extent that each of the three mathematics tests represented the CCSS in Mathematics (CCSSM). The four criteria of the WAT were used to determine the extent of alignment between tests and standards. For example, for an assessment to meet the Categorical Concurrence criteria, each standard must be represented by no less than six assessment items. Since the WAT does not include a method for determining the extent

assessment items represent the CCSS Mathematics Practices, content experts attempted to identify the practices in each assessment item using the same approach as identifying the standards.

Primary Analysis

Principal Component Analysis

A factor analysis identified as a Principal Components Analysis was utilized in SPSS to identify the primary factor structures in the five JCPS interim assessments. In effect, the analysis was conducted to identify the mathematics strands represented by the examination items; these strands are referred to as extracted components. Although this process was not performed to answer a research question directly, the researcher intended to strengthen the purpose for conducting the study. By completing the analysis, results not showing extracted components representative of the CCSSM strands would identify one reason for the need to gather data on content validity in another format. Additionally, if the strands were represented, the study would allow the researcher to compare the results of the factor analysis to that of the results from the WAT data.

This analysis initially intended to include data from the ACT EXPLORE. An analysis of the student answers for each of the 30 items on the ACT EXPLORE assessment were to be analyzed through a principal component analysis. However, the researcher was unable to obtain the data. Attempts were made over a three-month time frame to obtain the data. Initially, the researcher was given access to the data; however, when the researcher began the process of adapting the data to be in a format that would be usable in the statistical software, the researcher discovered that student answers were not given for 14 of the items. The 14 items were missing for all students. The researcher

proceeded to contact the assigned JCPS representative for assistance in gaining access to the data. The JCPS representative attempted to find the data requested but was unable to do so. The researcher was then referred to a Kentucky Department of Education (KDE) representative. Communication through email correspondence occurred with the KDE representative indicating that the district receives individual student data for that district whereas the state does not receive student data. The researcher returned to the JCPS representative for further assistance. The researcher was given an email address for a different individual to assist in finding an ACT representative for the state or district to contact. After multiple email communication among JCPS and KDE staff, an ACT representative was contacted. After this, the researcher communicated through email with the ACT staff person regarding what was needed to complete the data file. The ACT contact person responded that he would discuss the issue with other staff members and would communicate with the JCPS staff person with whom the researcher was continuing email correspondence. Email correspondence between the researcher and the JCPS staff person occurred at minimum once per week regarding the timeline of data access. Despite consistent communication over approximately an eight-week period, the data was not obtained from ACT. Additionally, there was no communication regarding a timeline for data to be sent to the researcher. Thus, at the recommendation of the researcher's dissertation committee chair, the researcher was allowed to continue the study without the ACT EXPLORE data.

The process for completing this analysis included the receipt of the sample data from the Jefferson County Public School District, the deletion of cases that could not be used in the analysis, and the re-coding of data from text to numerical data for use in

SPSS. The sample consisted of more than 7,000 cases; in this context a case consists of assessment answers for all test items for each student for the 2011-2012 school year. The sample was then simplified to include only those cases that had all test item answers. Therefore, any student's data, which did not include information for all five tests was deleted from the data to be used in SPSS. The deletion of students with missing data resulted in 825 cases that were usable for the factor analysis of the JCPS interim assessments.

There are several assumptions that must be met in order to determine whether a factor analysis will reveal results (Pallant, 2011). These assumptions are: 1) sample size, (2) level of correlation of items, (3) linearity, and, (4) outliers. Sample size is above the acceptable limits. Since the general acceptable sample size is a minimum of 200, although a factor analysis can be done on lower samples, this sample of 825 cases is sufficient. Table 3 describes the sample used in the analysis.

Table 3

Description of Sample Demographics

	N	%
Race		
White	434	52
Black	281	34
Hispanic	76	9
Other	34	4
Gender		
Male	439	53
Female	386	47
Lunch		
Paid	256	31
Reduced	72	9
Free	497	60

In determining the level of correlation of items, the analysis initially reveals acceptable results indicating the data are suitable for completing a factor analysis. This can be seen through Kaiser-Meyer-Olkin (KMO) statistic. The closer the KMO statistic is to 1.00, the better, and values above 0.60 are considered appropriate (Stevens, 2009). In the case of this data the KMO statistic was 0.764, which meets the general statistic for suitability. Additionally, Bartlett's Test of Sphericity indicates a significant value of 0.000 ($p < .05$); again, this indicates that this data would be appropriate for a factor analysis. However, upon examination of the correlation table, none of the correlations were in the acceptable range of 0.30 or higher. Thus, the un-correlation of test items makes a factor analysis unusable. Since the KMO and Bartlett's statistics were within acceptable limits, the researcher checked the remaining factor analysis output results to be certain that no usable data was obtained.

Generally, a Total Variance Explained table shows extracted components with Eigenvalues above 1.00 and indicating what components account for the most variance. Upon inspection of this table, the interim assessment data indicated a total of 32 components accounting for approximately 30% of the variance. An additional reference method for determining the number of components is the use of a scree plot, a graph that shows reference points identifying components to be extracted. The graph should show a visually identifiable slope from one component to the next; the slopes should not be close to equal or equal to one another. The research data revealed one steep slope and the remaining components had slopes that were fairly equal, shown by the appearance of a horizontal line with little to no steepness of the line, to one another. Although the initial component appeared steep, by returning to the total variance explained table the

researcher was able to determine that the extracted component only accounted for 7% of the variance. The low level of variance, in this data, determined the component to be of little value. The extracted component that accounted for 7% of the data referred to test items that did not appear related. The researcher was unable to identify a pattern or theme among the test items to identify a specific strand that the items applied to. This is likely due to the low correlation between each of the items.

Finally, a review of the initial component matrix revealed no usable data and a rotated component matrix was not completed by the SPSS output because of the lack of correlation among the data. The analysis was conducted to identify mathematics strands supported by the assessment items. This process revealed no output usable in identifying extracted components. Thus, the use of the student answers for the assessment items in the JCPS interim assessments lacked statistical representation to confirm specific mathematics strands.

Secondary Analysis

Webb Alignment Tool (WAT)

Completing the process with the WAT allowed the researcher to identify data for answering research question 1 regarding the extent that each assessment represented the content on the CCSSM. Additionally, research question 3 was represented through the process by identifying the assessment that most fully represented the content of the CCSSM. Each assessment was first analyzed by content experts and coded for item match to standards and cognitive complexity. Following the coding process, identification of 8th-grade content standards and cognitive complexity of each item, the researcher compiled data for analysis.

After data from the three assessments were gathered from the content experts' completion of the WAT, the researcher analyzed the standards and DOK identified for each assessment item, 89 items on the JCPS interim assessments, 30 items on the ACT EXPLORE, assessment, and 34 mathematics items from the 2009 NAEP. During the analysis of data supplied by the content experts, final selection of standards and DOK for each assessment item was determined by the most popular response among all five content experts. For example, an assessment item was assigned a Level 2 DOK level, if three of five coders assigned the item a Level 2. However, all matches identified by coders were submitted when inter-coder reliability (ICR) was determined. Thus, final tallies of the results were based upon the assignments made by the majority of the coders.

A Krippendorff's Alpha was utilized in identifying the extent to which this study obtained ICR for the WAT. After data for each expert were collected and entered into a spreadsheet, the data were then converted to a CSV file in order to submit the data for analysis through the ReCal website. One of the content expert's data was identified as an outlier when the results were analyzed for coding of standards. For example, pairwise agreement between three coders and coder number five was noted at 10% and average pairwise agreement for all five coders was 57%. After consultation with the researcher's committee, it was determined that, for the WAT process, the researcher should exclude the fifth coder.

All initial data submitted by the coders were analyzed to determine the level of inter-coder reliability (ICR). For the standards identification process a Krippendorff's Alpha revealed an ICR of 0.82 when data from four coders were used. Additionally, an average pairwise percentage agreement revealed 86% agreement among all four coders.

For DOK coding matches a Krippendorff's Alpha revealed an ICR of 0.80 for four coders and an average pairwise percentage agreement of 89%. The results of the Krippendorff's Alpha were within the parameters considered to be good for inter-coder reliability.

Categorical Concurrence. Categorical concurrence shows the extent to which the standards are represented on the assessment. In the Webb Alignment Tool Training Manual (Webb, Alt, Ely, & Vesperman, 2005), categorical concurrence is considered acceptable if a minimum of six assessment items corresponds to one standard. If four to five items on the assessment represent a standard it is considered weak and when three or less items are represented it is considered unacceptable. For the purposes of this study, if multiple CCSSM standards were identified for an assessment item, the assessment item was considered a part of the tally for the number of assessment items representing a standard. For example, for one assessment item representing three standards, each of the three standards identified would receive credit for the one assessment item. Only 8% (12 items of 136 items) of the assessment items represented multiple standards for the standard and none of the 12 items affected the meeting of categorical concurrence on any of the assessments. Therefore, the researcher included all data despite the 12 crossover instances. The results presented in Table 4 show the extent that each of the three tests studied met the categorical concurrence requirements for alignment to the CCSS in Mathematics.

Table 4

8th-grade CCSSM Categorical Concurrence

Categorical Concurrence		# of Objectives	# of Hits	Categorical Concurrence Level	% Acceptable
JCPS Interim Assessments	The Number System	1	7	Acceptable	100
	Expressions and Equations	2	46	Acceptable	
	Functions	2	21	Acceptable	
	Geometry	2	22	Acceptable	
	Statistics	1	10	Acceptable	
ACT EXPLORE	The Number System	1	0	Unacceptable	0
	Expressions and Equations	2	5	Weak	
	Functions	2	1	Unacceptable	
	Geometry	2	1	Unacceptable	
	Statistics	1	0	Unacceptable	
2009 NAEP	The Number System	1	0	Unacceptable	0
	Expressions and Equations	2	3	Unacceptable	
	Functions	2	0	Unacceptable	
	Geometry	2	5	Weak	
	Statistics	1	0	Unacceptable	

To summarize, the JCPS interim assessments are the only tests that meet the requirements for categorical concurrence according to the specifications noted in the WAT Training Manual. The ACT EXPLORE and 2009 NAEP tests do not meet the categorical concurrence specifications; however, each assessment has one category considered a weak alignment of categorical concurrence. In addition, after results were analyzed, it was discovered that 25 of the 30 items on the ACT EXPLORE did not match any of the 8th-grade CCSSM standards. Similarly, 28 of the 34 2009 NAEP released

items did not match any of the 8th-grade CCSSM standards. With the lack of representation of 8th-grade CCSSM clusters, the researcher also analyzed categorical concurrence of the ACT EXPLORE and 2009 NAEP to the 7th-grade CCSSM standards. The JCPS interim assessments were not included in the further examination of alignment with the 7th-grade CCSSM content because the JCPS interim assessments had already been determined to meet acceptable levels of Categorical Concurrence for 8th-grade. The results of the categorical concurrence analysis for the 7th-grade CCSSM content alignment are represented in Table 5.

Table 5

7th-grade CCSSM Categorical Concurrence of ACT EXPLORE and 2009 NAEP

Categorical Concurrence		# of Objectives	# of Hits	Categorical Concurrence Level	% Acceptable
ACT EXPLORE	Ratios and Proportional Relationships	1	4	Weak	40
	The Number System	1	13	Acceptable	
	Expressions and Equations	2	7	Acceptable	
	Geometry	2	2	Unacceptable	
	Statistics and Probability	3	1	Unacceptable	
2009 NAEP	Ratios and Proportional Relationships	1	2	Unacceptable	0
	The Number System	1	1	Unacceptable	
	Expressions and Equations	2	4	Unacceptable	
	Geometry	2	4	Weak	
	Statistics and Probability	3	2	Unacceptable	

The data on Table 5 shows an increased acceptable level of categorical concurrence for the ACT EXPLORE of 40% overall and an additional CCSSM 7th-grade strand rated at a weak level. After the alignment of 7th-grade standards, data obtained from the ACT EXPLORE resulted in a total of 15 items that did not represent any standards for 7th-grade; thus, there were fewer items that had no standards represented compared to the results of the 8th-grade CCSSM comparison. The data on Table 3 shows that the NAEP results do not show an increased categorical concurrence level overall for the assessment; however, analysis of the 7th-grade standards did result in the Geometry strand obtaining a weak representation of standards. However, despite the increased categorical concurrence for the Geometry strand, this did not increase the acceptability of the NAEP assessment for the number of assessment items representing the CCSSM. In the analysis of categorical concurrence 20 of the 34 items from the sample of the 2009 NAEP did not represent any of the 7th -grade CCSSM. Additionally, both the ACT EXPLORE and the NAEP had items that measured neither the 8th-grade nor the 7th-grade CCSSM standards. A total of 10 items on the ACT EXPLORE and a total of 16 items on the NAEP did not represent either set of standards.

DOK Consistency. DOK consistency, the second requirement of the WAT, determines the extent to which an item matches the cognitive complexity level expected of an assessment item to assess a standard. Each assessment item and each CCSSM standard was coded for DOK level. This resulted in a determination of percent match or percent above DOK represented by the assessment items. DOK consistency is met if 50% or more of the items' levels match the DOK levels of the standards they represent. According to the requirements set forth in the WAT Training Manual, a strand is

considered weakly met if the comparison is 40-49% and if the strand shows 39% or less consistency it does not meet DOK level consistency and is noted as unacceptable. Table 6 presents the results of the matching of DOK of assessment items compared to the DOK of the CCSSM standards.

Table 6

8th-grade CCSSM DOK Consistency

DOK Consistency		% at DOK	% above DOK	% at or above	DOK Level
JCPS Interim Assessments	The Number System	60	0	60	Acceptable
	Expressions and Equations	45	37	82	Acceptable
	Functions	90	0	90	Acceptable
	Geometry	57	5	62	Acceptable
	Statistics and Probability	45	0	45	Weak
ACT EXPLORE	The Number System	0	0	0	Unacceptable
	Expressions and Equations	60	20	80	Acceptable
	Functions	100		100	Acceptable
	Geometry	0	0	0	Unacceptable
	Statistics and Probability	0	0	0	Unacceptable
2009 NAEP	The Number System	0	0	0	Unacceptable
	Expressions and Equations	50	25	75	Acceptable
	Functions	0	0	0	Unacceptable
	Geometry	0	0	0	Unacceptable
	Statistics and Probability	0	0	0	Unacceptable

The data on Table 6 shows an overall acceptable DOK consistency level for the JCPS interim assessments. The Statistics and Probability strand shows a weak consistency but only by 5%. The ACT EXPLORE and 2009 NAEP data represents a weaker DOK consistency compared to the 8th-grade CCSSM. ACT EXPLORE data demonstrates consistency in both the Expressions and Equations strand and the Functions

the strand, while 2009 NAEP data demonstrates consistency in only one strand. Table 7 represents the results of the DOK consistency for the additional 7th-grade CCSSM analysis.

Table 7

7th-grade DOK Consistency for ACT EXPLORE and 2009 NAEP

DOK Consistency		% at DOK	% above DOK	% at or above	DOK Level
ACT EXPLORE	Ratios and Proportional Relationships	100	0	100	Acceptable
	The Number System	53	30	83	Acceptable
	Expressions and Equations	67	33	100	Acceptable
	Geometry	50*	0	50	Acceptable
	Statistics and Probability	0	100**	100	Acceptable
2009 NAEP	Ratios and Proportional Relationships	100	0	100	Acceptable
	The Number System	0	0	0	Unacceptable
	Expressions and Equations	50*	25	75	Acceptable
	Geometry	25	25	50	Acceptable
	Statistics and Probability	100	0	100	Acceptable

*50% of 2 assessment item hits for the strand.

**100% for only 1 assessment item hit for the strand.

Table 7 illustrates that an increase in the DOK consistency for the ACT EXPLORE and the 2009 NAEP when the alignment with the 7th-grade CCSSM. As seen in the table ACT EXPLORE met acceptable levels in all strand areas and the NAEP met all except in the Number System strand.

Range of Knowledge Consistency. The range of knowledge (ROK) criterion for content alignment is determined by the percentage of objectives for each standard that are hit by an item. The goal is to obtain an equal distribution of items hitting the clusters across the standards. This criterion is considered acceptable if at least 50% of the

objectives under each standard have at least one corresponding assessment item. The data on Table 8 shows the distribution percentage of items to objectives and identifies the strands that have been noted as acceptable for each assessment studied.

Table 8

8th-grade CCSSM Range of Knowledge Consistency

Range of Knowledge Consistency		% of Objectives Hit	ROK Level	% Acceptable
JCPS Interim Assessments	The Number System	100	Acceptable	100
	Expressions and Equations	88	Acceptable	
	Functions	80	Acceptable	
	Geometry	100	Acceptable	
	Statistics	75	Acceptable	
ACT EXPLORE	The Number System	0	Unacceptable	0
	Expressions and Equations	37	Unacceptable	
	Functions	20	Unacceptable	
	Geometry	11	Unacceptable	
	Statistics	0	Unacceptable	
2009 NAEP	The Number System	0	Unacceptable	0
	Expressions and Equations	37	Unacceptable	
	Functions	20	Unacceptable	
	Geometry	44	Unacceptable	
	Statistics	0	Unacceptable	

The range of knowledge table shows that the JCPS interim assessments meet the criteria in full while the ACT EXPLORE and 2009 NAEP meet none of the criteria showing 0% acceptable levels. Table 9 presents the ROK levels for the alignment of the ACT EXPLORE and the NAEP to the 7th-grade CCSSM.

Table 9

7th-grade CCSSM Range of Knowledge Consistency

Range of Knowledge Consistency		% of Objectives	ROK Level	% Acceptable
ACT EXPLORE	Ratios and Proportional Relationships	33	Unacceptable	60
	The Number System	100	Acceptable	
	Expressions and Equations	75	Acceptable	
	Geometry	50	Acceptable	
	Statistics and Probability	12	Unacceptable	
2009 NAEP	Ratios and Proportional Relationships	33	Unacceptable	20
	The Number System	33	Unacceptable	
	Expressions and Equations	75	Acceptable	
	Geometry	16	Unacceptable	
	Statistics and Probability	25	Unacceptable	

Although the ACT EXPLORE and 2009 NAEP assessment items fare better for ROK when compared to the 7th-grade CCSSM, they still are not measuring acceptable levels for each strand. ACT EXPLORE shows a change from 0% on the 8th-grade CCSSM to 60% overall on the 7th-grade CCSSM. However, ACT EXPLORE still remains far below an acceptable level in two strands.

Balance of Representation. The WAT uses the balance of representation criteria to determine whether the hits on the objectives from the assessment items are balanced evenly under the standards. Lombardi (2006) states, “a high number of duplicate hits will overwhelm the number of test items, creating a spurious proportional index that is unusable” (p. 9) which supports the exclusion of the Balance of Representation criterion. The researcher elected not to include this criterion when completing the data analyses because of the high number of hits coded for multiple objectives from one assessment item. Although the WAT formula for determining balance of representation does allow

for secondary hits, it does not consider the number of items that do not identify an objective nor does it support items covering a multitude of objectives.

Summary of WAT Findings

Table 10 provides a summary of the results of the use of the four content experts’ to alignment of each of the assessments to the 8th-grade CCSSM.

Table 10

Summary of Alignment for each Assessment

		Alignment Criteria		
		Categorical Concurrence	DOK Consistency	ROK Consistency
JCPS Interim Assessments	The Number System	Acceptable	Acceptable	Acceptable
	Expressions and Equations	Acceptable	Acceptable	Acceptable
	Functions	Acceptable	Acceptable	Acceptable
	Geometry	Acceptable	Acceptable	Acceptable
	Statistics	Acceptable	WEAK	Acceptable
ACT EXPLORE	The Number System	Unacceptable	Unacceptable	Unacceptable
	Expressions and Equations	WEAK	Acceptable	Unacceptable
	Functions	Unacceptable	Acceptable	Unacceptable
	Geometry	Unacceptable	Unacceptable	Unacceptable
	Statistics	Unacceptable	Unacceptable	Unacceptable
2009 NAEP	The Number System	Unacceptable	Unacceptable	Unacceptable
	Expressions and Equations	Unacceptable	Acceptable	Unacceptable
	Functions	Unacceptable	Unacceptable	Unacceptable
	Geometry	WEAK	Unacceptable	Unacceptable
	Statistics	Unacceptable	Unacceptable	Unacceptable

The data in Table 10 shows the JCPS interim assessments strongly measure the content of the 8th-grade CCSSM. Both the ACT EXPLORE and the 2009 NAEP clearly do not measure the CCSSM content standards. The data in Table 11 represents a summary of the results of the 7th-grade alignment for the ACT EXPLORE and the 2009 NAEP. The data

represents a stronger alignment between the two assessments and the 7th-grade content standards versus the 8th-grade content standards; however, the data indicate that both assessments still lack in alignment in Categorical Concurrence and ROK Consistency.

Table 11

Summary of 7th-grade Alignment for ACT EXPLORE and 2009 NAEP

		Alignment Criteria		
		Categorical Concurrence	DOK Consistency	ROK Consistency
ACT EXPLORE	Ratios and Proportional Relationships	WEAK	Acceptable	Unacceptable
	The Number System	Acceptable	Acceptable	Acceptable
	Expressions and Equations	Acceptable	Acceptable	Acceptable
	Geometry	Unacceptable	Acceptable	Acceptable
	Statistics and Probability	Unacceptable	Acceptable	Unacceptable
2009 NAEP	Ratios and Proportional Relationships	Unacceptable	Acceptable	Unacceptable
	The Number System	Unacceptable	Unacceptable	Unacceptable
	Expressions and Equations	WEAK	Acceptable	Acceptable
	Geometry	WEAK	Acceptable	Unacceptable
	Statistics and Probability	Unacceptable	Acceptable	Unacceptable

Tertiary Analysis

CCSS Mathematical Practices Alignment

As recommended by the researcher’s committee, this study included an analysis of the representation of the CCSS Mathematical Practices by the items on the assessments. The purpose for conducting an analysis of mathematical practices in assessment items was to answer both research question 2 and research question 3. Thus, the representation of the practices would provide answers for the extent of representation

of the practices in application to each of the mathematics strands. Moreover, the analysis of the practices would identify the assessment with the greatest representation of items utilizing components of the mathematical practices in assessing student content knowledge.

A search of the literature revealed no current tool for identifying the practices within assessment items. Under the guidance of the researcher's committee, the researcher was instructed to complete the analysis in a similar manner to the content analysis of assessment items completed using the WAT. As a result, content experts received instructions to assign each assessment item at least one mathematical practice in the same manner the CCSSM content standards were assigned. With no supporting tool, the researcher with committee guidance, requested the content experts to choose practices in a similar manner to the method for assigning a DOK level. Thus, since a DOK level is assigned based upon what amount of cognitive demand to solve a problem, the content experts were asked to choose CCSS Mathematical Practices based upon what skills the student would need to use in order to answer the question. These assignments were made in a similar manner to that of assigning the content standards when using the WAT.

Three content experts coded independently while two content experts coded collaboratively. The two content experts coding collaboratively were allowed to view the mathematical practices assigned to each item completed by the other content experts in order to attempt to identify a consistent definition of each of the practices. Despite the ability to reference the data provided by the other three experts, the two experts that coded collaboratively still had differing opinions than the other experts. These differences were obvious when inter-coder reliability statistics were obtained.

After coding of mathematical practices by content experts, an inter-coder reliability analysis was completed. A Krippendorff's Alpha was obtained at $\alpha=0.454$. An acceptable alpha level is 0.80; thus, the mathematical practices alignment was not reliable. This is further supported by the low average pairwise percentage of agreement 55.7%. Although a 55.7% may seem reasonable, this is a slightly skewed statistic due to two of the coders working collaboratively on their assignment of practices to items. Thus, the two coders had 100% agreement in data supplied that also had several instances of match with other coders. The 100% agreement between two coders due to working collaboratively affected the overall percentage. However, despite an overall 55.7% agreement, pairwise agreement was low among all coders. For example, in one case of a pairwise agreement, the percentage obtained was 10% and in another the agreement was 11%.

Due to the low level of reliability found amongst the coders, all five content experts met to discuss definitions of the practices in application to assessment items. At the recommendation of the researcher's committee, a final meeting of content experts took place; the meeting resulted in notes to use as reference points during data analysis. The five content experts discussed each of the mathematical practices to attempt to define how an assessment item could represent a practice or more than one. While the discussion regarding the operational definitions took place, the comment, "but that is my opinion based upon teaching X grade and X content". Thus, each content expert, having had different instructional experiences, had differing opinions about the practices and what an example of an item testing the practice would show. Specific items were discussed to attempt to identify the procedure for assigning the practices to the

assessment items. Table 12 shows the practices that were identified by each content expert for Part B of the test item in Figure 2. Each of the numbers in Table 10 represents the CCSS Mathematical Practice (CCSSMP) each coder assigned. For example, a “2” represents CCSSMP: Reason Abstractly and Quantitatively.

Figure 2

JCPS Interim Assessment Question

Open Response: Saving Pennies

Justin’s parents are trying to encourage him to start saving for college. To show how quickly pennies can add up, they offer to start saving pennies in a jar. They will give him 1 penny the first day, 2 pennies the second day, 4 pennies the third day, and continue to double the previous day allowance for thirty days.

A. Complete the table to show the number of pennies placed in the jar for the first ten days.

Day	1	2	3	4	5	6	7	8	9	10
# of pennies	1	2	4							

B. Graph the data for days 1 through 10. Label and title your graph appropriately.

C. Is the relationship between the day and the number of pennies linear? Explain why or why not.

Table 12

Coding Results for the Assessment Item in Figure 1 for Part B Only

Coder 1	Coder 2	Coder 3	Coder 4	Coder 5
2	2	2	2	4
4	1	1	0	0
7	6	6	0	6

Note: 0's represent no additional practice assigned by that expert.

Even though there are several agreements, there are no instances where all 5 content experts agreed on one specific practice. Additionally, while four of the five experts assigned the assessment item the CCSS Mathematical Practice Number 2, Reason Abstractly and Quantitatively, the rationale behind the assignment was different. One content expert assigned Mathematical Practice 2 because the student would have to create an equation from contextual information; the argument was that this represents the practice definition portion, “quantitative reasoning entails habits of creating a coherent representation of the problem at hand” (Content Expert 1, personal communication, December 5, 2012). However, another content expert identified this as a plug and chug type of problem requiring the student to plug a number in for a variable and use basic mathematical computation to find an answer; thus, the expert determined that item did not represent reasoning and representation. Another content expert who assigned the same practice said that the student would need to choose a correct procedure and be flexible in solving the problem in multiple ways. The practice definition states that students should be able to attend “attend to the meaning of quantities, not just how to compute them and knowing and flexibly using different properties of operations and

objects” (CCSSO, 2010, 7). Throughout the conversation with the content experts, the content experts could not come to consensus or identify a reason for assigning a particular practice for an assessment item. Although experts would assign an assessment item the same mathematical practice, they cited differing opinions for their reason for identifying that practice. Thus, the data collected from experts assigning the practices to test items was inconsistent. The causality and relationship between definitions of each practice was inconsistent. The content experts continued to work to determine operational definitions for using the practices in assessment items; however, upon completing definitions for each practice and reviewing several additional problems, it was determined that the definitions continued to apply only to instruction and not to assessments items.

Summary of Mathematical Practices Findings

After the results were analyzed and findings showed a lack of reliability among the content experts, a discussion was warranted to determine areas of disagreement in the coding of the items in regards to the Mathematical Practices. When multiple assessment items were discussed individually, each content expert agreed to having assigned each item a CCSS Mathematical Practice due to the initial requirements of the study. Each content expert expressed concern regarding the data that were submitted to the researcher. Based upon evidence gathered from the final meeting, each expert commented that the CCSS Mathematical Practice data provided was not a true reflection of his or her opinion, and, given the chance, would assign most assessment items no mathematical practice.

Due to these results, research question 2, identifying the assessment with the highest representation of the CCSS Mathematical Practices, was not a possibility. Moreover, since research question 3 is in regards to finding the assessment that meets acceptable levels of alignment of content validity and represents the mathematical practices, it is also unanswerable. Thus, in order to discern findings that explicitly answer research questions 2 and 3, the framework of the study must be readdressed. The theoretical framework would need to include a method such as the SPUR method developed by Usiskin (2007) to address a different framework for assessing mathematical practices. The SPUR method assesses students' mathematical understanding by identifying each assessment items as a Skills, Properties, Uses, or Representations item. Bleiler and Thompson (2013) studied the SPUR method during their assessment of results from the International Project on Mathematical Attainment (IPMA). Researchers across multiple countries worked together to assign each assessment item a SPUR identifier. Through analyzing student achievement dependent upon the type of task identified by the SPUR method teachers were provided with thorough information regarding what their students understood about mathematics and how they could reason through a mathematics task. Bleiler and Thompson (2013) claim that students who can move fluently through each of the SPUR dimensions show that they "are developing a robust and connected understanding of mathematics" (p. 300). Thus, if the CCSS Mathematical Practices are designed to ensure students show conceptual understanding and contextual reasoning skills, then the use of the SPUR dimensions could be an acceptable framework for identifying representation of mathematical practices in assessments.

Conclusion

The findings provided in this chapter support the need to further study mathematics assessment items on tests to determine the level that they are correlated to one another. By doing this it can be determined whether the assessment items are representing specific themes that are represented as mathematics strands. The PCA results of the JCPS interim assessments showed little correlation of items to one another; therefore, the results showed no strong representation of items for the strands recognized in the mathematics standards. Additionally, although it was the intention of the researcher to include a PCA analysis of the ACT EXPLORE assessment items, the data was unobtainable despite correspondence over several weeks. Therefore, the PCA analysis of the ACT EXPLORE data could not be included within the scope of this study.

With no correlation immediately visible through the PCA, an alignment through the use of content experts was supported. Each of the three assessments in this study was analyzed using the Webb Alignment Tool and, for 8th-grade, only the JCPS interim assessments met acceptable levels in each of the criteria to be considered aligned with the 8th-grade CCSSM content standards. The data made it evident that the ACT EXPLORE and the 2009 NAEP were not meeting alignment requirements; therefore, the study extended into examining the 7th-grade CCSSM for the ACT EXPLORE and the 2009 NAEP. After the 7th-grade alignment was completed, it was evident that both assessments represented more of the 7th-grade standards than they did for the 8th-grade CCSSM content standards. Thus, the JCPS interim assessments show the greatest representation of the 8th-grade CCSSM content standards.

Finally, the content experts worked to determine whether the items on the three assessments represented the CCSS Mathematical Practices. Findings of this research portray little to no representation of the practices by the items on any of the three assessments. However, the research reflects the opinions of the five content experts utilized for this study. The opinions of those experts are what determined the data and results presented here. Due to the qualitative nature of this data, if alternate experts to be utilized, the results of the mathematical practices data could vary significantly. Thus, sharing the initial results of the coding of practices to assessment items would be misleading to those attempting to find possible items to assess student skill in application to the mathematical practices. Therefore, the findings of the analysis of mathematical practices represented in assessment items revealed evidence that none of the three assessments is a good choice for use in representing student content knowledge while utilizing the mathematical practices.

The findings and conclusions support teachers and administrators who use these three assessments to determine student knowledge in application to the CCSSM content standards and CCSS Mathematical Practices. Chapter V provides information regarding the implications, limitations, and further research suggestions for study in application to the results found here.

CHAPTER V

DISCUSSION

Introduction

This chapter provides a summary of the study and results, presents conclusions based upon the findings presented in chapter four and a discussion of the implications from the data analyses. The chapter concludes with the limitations of the study and suggestions for future research.

Summary of the Study

Restatement of the Problem Statement

A requirement of the No Child Left Behind legislation (2002) to reach 100% proficiency in mathematics and language arts by 2014 has created a national focus on state standardized testing outcomes. Studies on the state of eighth-grade mathematics proficiency and eighth-grade content standards in the United States have revealed concerning results. For one, national and international exams have repeatedly indicated the lack of student mathematics proficiency in the United States when compared to other countries (ACT, Inc., 2005; American Federation of Teachers, 1998). Additionally, ACT (2007, a, b, c) has indicated a lack of high school readiness by eighth-graders in addition to the lack of college and career readiness of high school students.

Reactions to the continued lack of proficiency in mathematics by US students, has led to multiple studies attempting to discover reasons for this deficiency. The American

Federation of Teachers (AFT) (1998) discovered that US mathematics exams lacked rigor and assessed broad amounts of content when compared to assessments used in three of the highest performing countries on the TIMSS exams. Moreover, an additional study of all 50 states' content standards indicated a lack of cohesiveness in content coverage indicating only 20% of mathematics standards as similar (Porter, McMaken, Hwang, & Yang, 2011). With studies supporting lack of rigor and the broad level of content in mathematics in the US, national standards were developed, drafted, and released in 2010. The Common Core State Standards in Mathematics (CCSSM), adopted by 45 states (Common Core State Standards Initiative, 2012) are now being implemented across the nation with national assessments created by PARCC and Smarter Balanced due to be released by 2014.

While the CCSSM national assessments have not yet been released, under NCLB legislation, states are still expected to maintain accountability expectations. The school district studied in this research, Jefferson County Public Schools (JCPS), simultaneously began to transition their interim assessments in 2010 to reflect the state changes. The district assessments, used for tracking of student performance, school accountability, and teachers' assessment of students, have been designed based upon test banks and input from teachers and district support staff. JCPS also participates in the National Assessment of Educational Progress (NAEP) in mathematics every two years and receives district level score reports since it is a participant in the Trial Urban District Assessment (TUDA). Additionally, the state utilizes the ACT EXPLORE assessment of eighth grade students for accountability purposes and tracking of student college and career readiness progress.

Restatement of the Purpose and Research Question

The purpose of this study is to determine the level of alignment of 2009 NAEP released mathematics items, the ACT EXPLORE used in 2010, and the 2011-1012 JCPS interim assessments to the CCSSM. The research questions addressed in this study were:

- 1) To what extent do the (1) eighth-grade mathematics released items in the 2009 NAEP assessment, (2) ACT EXPLORE, and (3) items from the Jefferson County Public School District interim assessments align with the eighth-grade CCSSM?
- 2) To what extent are the CCSS Mathematical Practices represented in each strand of the eighth-grade 2009 NAEP mathematics assessment released items, ACT EXPLORE, and the Jefferson County Public School District interim assessments?
- 3) Which of the tests shows the most content validity and which shows the most mathematical practices represented compared to CCSSM?

Findings and Relation to the Literature

Research Question 1

The purpose of question 1 was to determine the extent of alignment with the 8th-grade CCSSM content standards for the 2009 NAEP mathematics released items, the ACT EXPLORE, and the JCPS interim assessments. Therefore, the resulting findings address the extent of content validity in application to the 8th-grade CCSSM content standards are represented by each assessment. Prior to the use of the WAT for alignment, the researcher conducted a principal components analysis (PCA) to determine whether evidence of assessment items represented the strands of the CCSSM. The results of the PCA for the JCPS interim assessments revealed a lack of correlation among items on the assessments.

The statistical extraction of components to identify CCSSM strands was not possible within the scope of this study. However, this statistical method may not be the only manner possible for determining whether strands were discernable. The researcher combined all items into one file whereas the interim assessments are actually separated into five assessments. Each assessment is specific to the content covered during the time frame preceding the students' taking the assessment. Thus, by combining all items from the assessments into one file, the correlations may have been weakened. So for example if the analysis of each of the 5 assessments had been conducted separately it may have contributed to stronger correlations among the items.

The researcher made the decision to analyze all five assessments as one document to be consistent with the research that was conducted throughout the study. In order to use the WAT in a format that would result in a thorough representation of the content assessed within the interim assessments, the assessments had to be examined as one document. If the research had not been conducted using this method, none of the assessments examined independently would have had enough items to represent the CCSSM content standards as a whole. Therefore, to remain consistent with the methodology used in the WAT, the researcher examined data from all five assessments in one PCA.

Results from the WAT, the method used for analyzing content validity of the assessments applicable to the 8th-grade CCSSM, were presented in Chapter IV within multiple tables representing each of the alignment criteria. The alignment data from this study with the inclusion of the links to the literature are summarized in the following paragraphs.

The JCPS interim assessments met each of the requirements on the WAT to show acceptable alignment to the 8th-grade CCSSM content standards. Thus, categorical concurrency, depth-of-knowledge, and range-of-knowledge requirements were met. Davis, Caros, Grossen, and Carnine (2002) found that interim assessments that are determined to be reliable and have internal consistency, which can also be referred to as content validity (The U.S. Office of Special Education Programs, 2005), can be used to predict performance on large-scale assessments. Therefore, because the findings of this research determined that the JCPS interim assessments do indeed represent the content of the 8th-grade CCSSM, JCPS could utilize the data gathered to predict performance on other assessments that are also considered to have 8th-grade CCSSM content validity.

The 2010 ACT EXPLORE mathematics and the 2009 NAEP mathematics released items did not meet all requirements of the WAT to be considered to have acceptable alignment for the 8th-grade CCSSM. ACT (2010b) indicates that administrators can use ACT EXPLORE scores as evidence of students' success in application to the CCSSM. The findings of this study indicate that administrators should use caution when attempting to apply the ACT EXPLORE scores to specific grade levels. In this case, strong alignment was not indicated for use of ACT EXPLORE scores applicable to the 8th-grade CCSSM. Additionally, ACT (2010b) reports findings from ACT experts that examined alignment of the EXPLORE items to the 8th-grade CCSSM. Those findings indicated 100% content alignment; however, the findings from this research contradict the ACT (2010b) report indicating a 0% alignment with 8th-grade CCSSM content standards.

The Carmichael, Martino, Porter-Magee, and Wilson (2010) study that included the examination of CCSSM content standards included on NAEP assessments, determined that the NAEP assessed the CCSSM content at a grade of C due to lack of rigor and unclear language in the assessment items. It is difficult to determine if the grade of C aligns with the findings of this study due to unclear methodology as to what the grade indicated in the Carmichael et al. (2010) study. However, the results of this research do not corroborate an average level of alignment, the indications of the research provide evidence that the 2009 NAEP is lacking in any content validity with the 8th-grade CCSSM content strands.

Due to the results of limited alignment with the 8th-grade content, the researcher and content experts also conducted analyses of the ACT EXPLORE and the 2009 NAEP released items with the 7th-grade CCSSM content standards. When the 7th-grade CCSSM content standards were aligned to the fall 2010 ACT EXPLORE mathematics assessment, WAT data demonstrated acceptable alignment levels for 67% of the requirements, 7% were considered weak alignment, and 26% were considered unacceptable when aligned to the 7th-grade CCSSM content standards. The ACT EXPLORE assessment reviewed in this study was given in the fall of 2010. After alignment processes were completed for the ACT EXPLORE using both 7th- and 8th-grade CCSSM content standards, 15 of the 30 items on the assessment did not represent any standards for either of the two grade levels analyzed. Recalling that the intended use of the ACT EXPLORE is to represent student potential to succeed in high school, the researcher believes that the ACT EXPLORE does not assess unimportant skills. However, if the ACT EXPLORE is to be used as a predictor for CCSSM success,

particular attention needs to be paid to the areas and grade levels for which the ACT EXPLORE scores will be applied.

When the 7th-grade CCSSM content standards were utilized to align the 2009 NAEP mathematics released items, WAT data revealed acceptable levels for 33% of the requirements, 13% were considered to be at a weak alignment level, and 54% were considered to be at an unacceptable level of alignment. After alignment processes were completed for the 2009 NAEP mathematics released items using both the 7th- and 8th-grade CCSSM content standards, 20 of the 34 items on the assessment did not represent any standards for either of the two grade levels analyzed. The NAEP as it currently is used, does not assess any one state's or any one association's (e.g. NCTM) principles and standards for mathematics content. Instead, the NAEP is designed to assess what students should know at a given grade, in this case 8th-grade. A study conducted by Daro, Stancavage, Ortega, DeStefano, and Linn (2007) found that the NAEP assessments for 4th- and 8th-grade consisted of items that broadly covered the content frameworks set forth in the design of the NAEP. Moreover, the Daro et al. (2007) study cited the main issue with the content validity was the lack of items representing certain strands.

The research findings in this study corroborate the findings of Daro et al. (2007). The data collected from the WAT show that in both 7th- and 8th-grade content, the NAEP items examined more thoroughly both the Expressions and Equations and the Geometry strands than any others. The 7th-grade CCSSM content had at least one NAEP item in each strand area; however, the 8th-grade CCSSM content strands had items only in the two strands, the Expressions and Equations and the Geometry strands. However, because

the NAEP does not utilize a set of standards that matches neither a state nor national standards, aligning the NAEP content to the CCSSM is not intended to identify validity issues with the NAEP mathematics assessment.

Since these assessments were created prior to the release of the CCSSM neither the ACT EXPLORE nor the NAEP could have been specifically aligned to the 7th- or 8th-grade CCSSM content standards. The previously discussed ACT EXPLORE (2010b) report was published prior to the release of the CCSSM with an analysis of their ACT EXPLORE assessments showing 100% 8th-grade alignment. However, ACT (2010b) does not state from which years the assessments they reviewed were released. The researcher speculates that the assessments reviewed would be those published prior to the release of the CCSSM since the report was published in June of 2010. Assuming this is the case, the results of this research are valid. However, the researcher was interested in the alignment of the ACT EXPLORE with the NCTM Principles and Standards for School Mathematics (NCTM PSSM) since NCTM was also a part of the CCSSM design team. Upon perusal of the NCTM PSSM grades 6 – 8 expectations, a better alignment of the ACT EXPLORE was obvious to the researcher.

For example, a certain ACT EXPLORE item requested students to determine the perimeter of a rectangle (no formula was given). The content experts all agreed that there were no 7th- or 8th-grade CCSSM content standards matching this assessment item. However, the NCTM PSSM Measurement standard identifies that students in grades 6-8 should “select and apply techniques and tools to accurately find length, area, volume, and angle measures to appropriate levels of precision” (NCTM, 2013). Although this specific item did not meet the CCSSM content standards for either of the grade levels

examined, they do meet the 6-8 grade level expectations of the NCTM content standards. Due to the NCTM standards inclusion of grade 6, the researcher did look at the grade 6 CCSSM content standards to determine whether the item met requirements to be considered a grade 6 item. Indeed, there were no standards specifying the student learn or understand perimeter. Therefore, the ACT EXPLORE problem is identified at a lower grade level than 6-8 according to the CCSSM content standards.

Due to the consideration of both the ACT College and Career Readiness Standards and the use of the NCTM PSSM in the design of the CCSSM the researcher would anticipate that the ACT EXPLORE item be within the 6-8 grade band standards on the CCSSM. So, the researcher believes that assessments created prior to the release of the CCSSM would likely include standards in lower grade levels than were previously recognized. Prior to the study being conducted, the researcher believed that the designers' use of the ACT EXPLORE, the NCTM PSSM, and the NAEP frameworks as guides for the creation of the CCSSM would have meant each of the national assessments would have more closely aligned with the CCSSM. Therefore, the researcher advises administrators to use continued caution in utilizing national assessments as accountability measures for the CCSSM, progress monitoring tools in application to the CCSSM, or predictors of student achievement on the CCSSM.

In the context of this study, information was provided for administrators who are interested in using the NAEP as a progress-monitoring tool. The JCPS district, as a TUDA district, can continue to use the NAEP data in longitudinal manner monitoring student proficiency as it applies to the NAEP framework from one year NAEP assessment to another. However, the findings reported in this study do not support the

use of NAEP as a progress-monitoring tool or predictor of success on the 8th-grade CCSSM content. There is currently no literature specifically linking NAEP assessment items to the content standards for any grade level of the CCSSM. These findings indicate that the NAEP items are not robust enough to represent middle school CCSSM content; however, it may be the case that, if other grade levels were examined, the NAEP covers broad grade levels of CCSSM content. Thus, the grade of C given by Daro et al. (2007) would be further corroborated due to the broad range of items covering the CCSSM. Using the Survey of Enacted Curriculum (SEC) alignment tool, Polikoff, Porter, and Smithson (2011) determined that only 21% of the NAEP items were aligned to the CCSSM content. Although the findings of this study indicate a 0% alignment, this study was also limited to only 7th- and 8th-grade CCSSM content.

Although the overall analysis of the ACT EXPLORE and the NAEP items were not considered to be at a 100% acceptable alignment with the 8th- or 7th- grade CCSSM content standards, each of the three assessments met acceptable cognitive complexity requirements. Specifically, Table 13 shows DOK level representations for each assessment. The ACT EXPLORE and NAEP assessments were not found to meet acceptable levels of content validity to be utilized as predictors of student achievement on the CCSSM; however, it is possible that each exam may be usable as a preview of student reasoning and sense making skills. Therefore, the ACT EXPLORE and NAEP assessments could provide administrators and educators with information regarding the level to which their students are progressing on applying conceptual understanding skills. Thus, were each assessment aligned first to content followed by DOK representation of the assessment items, the researcher posits that student scores on the

items in each grade level would indicate their understanding. For example, consider the following hypothetical situation:

- 1) 10 assessment items determined to be 7th-grade content
- 2) Five of those items identified at a level 3 DOK (Strategic Thinking), 4 items at DOK 2 (Skill/Concept), and 1 item at DOK 1 (Recall)
- 3) A student gets all of the DOK 3 items correct and 2 of the DOK 2 items correct

This situation would indicate to an educator that the student does well applying the concept and has the skill to solve the problem. However, this student may have missed an item that required a formula or memorization of the difference between rational and irrational numbers. Therefore, this information would be valuable for educators and administrators to pin point specific areas of skill that students need to further develop as well as what the district's or school's teachers are doing well.

Table 13

DOK Representations

Assessment	% of Items in Level 1 (Recall)	% of Items in Level 2 (Skill/Concept)	% of Items in Level 3 (Strategic Thinking)	% of Items in Level 4 (Extended Thinking)
JCPS Interim Assessments (89 items)	37%	60%	3%	0%
ACT EXPLORE (30 items)	40%	60%	0%	0%
2009 NAEP (34 items)	59%	29%	9%	3%

With the cognitive complexity of each of these assessments represented, one can see that all three assessments' items are focused mainly in the first two levels of DOK. This indicates a low level of cognitive demand required of students to answer questions that apply to the content. This does not mean the content is easy to learn, however. Instead, it indicates that the wording of the CCSSM content standards only indicates a low level of DOK. Wording of both standards and items can reduce the level assigned to assessment items and content standards; thus, it is important for assessment designers to consider the wording utilized when writing assessment items dependent upon the level to which they want students to contemplate mathematics questions. The creation of the CCSS Mathematical Practices, being more descriptive regarding the level to which each grade must be able to apply the mathematics, may allow assessment and curriculum designers to develop more cognitively demanding assessment items. Thus, furthering this research to include links of cognitive complexity to that of the CCSSM Practices may reveal the NAEP assessment provides items that can indicate to administrators and educators that their students are successful in applying mathematics in more intellectually demanding situations.

Moreover, both the JCPS interim assessments and NAEP mathematics items included constructed-response type items while the ACT EXPLORE consisted only of multiple-choice items. When question format is taken into consideration, the researcher believes that higher cognitive complexity items are typically constructed-response items. If a goal of the CCSS is to become more specific in content and focus on student use within life contexts, the cognitive complexity levels may be more applicable to the validity of the assessments. With the addition of the CCSS Mathematical Practices

within the national standards, a focus has become reasoning, sense making, and application in mathematics; thus, the DOK percentages in each assessment show that the NAEP is in fact the strongest of the three assessments. This finding indicates an additional argument for furthering research in linking item format, cognitive complexity, and, now, the mathematical practices.

Research Question 2

The purpose of research question 2 was to analyze the extent to which each of the assessments items demonstrated the use of the mathematical practices. Analysis of the data collected from all five content experts revealed a lack of inter-coder reliability. Thus, a final meeting among the experts was conducted to determine the reason for the inconsistencies among coders. The following paragraphs identify pertinent results from the discussion and data analysis.

Content experts were unable to unanimously agree upon a single definition for each of the eight CCSS Mathematical Practices. Thus, no overarching or operational definition could be created in order to re-analyze the items on each of the assessments. Moreover, similar to opinions expressed by others (Confrey & Krupa, 2010; Hull, Miles, & Balka, 2012) experts questioned the application of the practices within assessment items. The experts talked extensively about application of the practices to instruction for students to develop conceptual understanding to apply content in real-world, meaningful contexts. However, the review of the literature identified multiple authors citing the need for implementing the practices within instruction and assessments (Confrey & Krupa, 2010; Kepner & Huinker, 2012; Krupa, 2011; McCallum, 2012), but the information gained in this study identified difficulty in consistent definitions of the practices for

instruction purposes and further difficulty in identifying consistent representations in assessment items.

Based upon the Porter et al. (2008) literature regarding content expert experience and the discussion between the content experts, the researcher believed that, when reviewing assessment items, each grade level should be considered separately when identifying the mathematical practices. That is to say that the mathematical practices cannot be applied at each grade level in the exact same manner, but must be applied in a method that is appropriate for the grade. Moreover, this ideal is loosely supported within the descriptions of each of the practices within the CCSSM. The practice definitions state what types of tasks elementary children versus high school children may execute to represent understanding using each practice. However, although the definition of the practices does give an example, the examples are very broad and are not identified for each grade level.

In order to more accurately and consistently determine mathematical practice representation in assessment items, a definition and identifiers must be declared prior to the coding of the items. In the case of this study, the researcher gave little instruction on the method for determining the practices in order to allow the researcher to gather data on the experts' individual definitions and practice identifiers. Content experts were asked to review each assessment item and assign a practice based upon what practices they thought the student would have to utilize to answer the question. However, each expert felt differently about what the practices meant in instruction and in assessment items. Thus, a more consistent method for determining the practice would need to be identified prior to coding.

A meeting was held to discuss the results of the coding methods of the mathematical practices. The discussion began with dialogue regarding assessment items that experts had assigned the same mathematical practices. The researcher intended to use those items to discern a consistent operational definition for the practices. When those discussions ensued, none of the five coders shared similar opinions for why the practice was assigned. Despite a common practice choice being assigned to an item, there was no apparent common definition. A clear example of this was the JCPS interim assessment item requiring a student to use a given table to create a graph of the data with axes labeled correctly. The majority of the experts had identified the same two mathematical practices for the item; however, each of the experts cited differing definitions and identifiers for those practices.

During the meeting, the researcher referred to descriptions of mathematical proficiency standards explained in the NRC (2001) book *Adding it Up* and the NCTM (2000) standards and principles for mathematics proficiency. A linking chart represented by Hull, Miles and Balka (2012) relating the NRC (2001) proficiency standards and NCTM (2000) process standards the CCSS Mathematical Practices, was utilized as a way to reference the items in different manners (p. 50). For example, MP 2: Reason Abstractly and Quantitatively in the CCSSM, is referenced as Reasoning and Proof by NCTM, and, as Procedural Fluency by the NRC. The use of these references did not assist the experts in linking their mathematical practice choices. Instead, the references brought about more questions regarding the difficulty of interpreting the mathematical practices. Experts argued their choices for choosing the mathematical practice for different reasons citing the differing definitions from both NCTM and NRC references.

As an example, Figure 3 represents item 1 from the JCPS interim assessment first test. On part A, four experts identified the item as MP 7: Look For and Make Use of Structure, and one expert identified the item as MP 6: Attend to Precision. In this example, experts did not agree on the practices, but cited similar definitions. The Hull, Miles and Balka (2012) chart supports this phenomenon by identifying the NCTM principle Connections as covering both of the CCSSM Practices.

Figure 3

JCPS Interim Assessment Test 1 Item

The table below shows the relationship between the numbers of people standing in line at Hot Tickets to the total number of minutes waiting in line.

Number of People in Line (<i>p</i>)	Number of Minutes Waiting (<i>m</i>)
2	3
3	6
4	9
5	12

A. Based on the table, how many minutes will you wait if you are the 10th person in line? Show work or explain how you arrived at your answer.

There are additional reasons presented in the literature for the identification of one practice for an assessment item while the experts referenced dissimilar reasons for assigning the same practice. Previous research indicates that experience and career positions can account for some fluctuation in coding choices (Buckendahl, Plake, Impara, & Irwin, 2000; Herman, Webb, & Zuniga, 2005; Porter, Polikoff, Zeidner & Smithson, 2008). A Porter, Polikoff, Zeidner, and Smithson (2008) study indicated that the background and level of experience of each content expert affects the results of the

coding data. Specific to this study, four of the five content experts had experience with 8th-grade in different settings. One content expert's experience was more focused on children with exceptionalities such as those with specialized learning accommodations. Another expert had eight years of experience and taught 8th-grade honors mathematics and two other experts had differing teaching philosophies. The fifth expert had more years of experience in high school mathematics than in 8th-grade mathematics. The experts referenced that the mathematical practice being discussed would have been represented in their classroom in differing ways.

Discussion among the content experts revealed discrepancies regarding what qualities each assessment item should incorporate to show representation of particular mathematical practices. For one example, in the discussion regarding MP 4: Model with Mathematics, content expert 1 believed that a student identifying an equation to represent a problem was representative of a student modeling with mathematics. Expert 1 believed that students that are able to move between different representations of a problem (e.g. a graph to an equation) would be showing proficiency in understanding how to model with mathematics. The same expert felt that the task did need to involve the student creating his or her own equation and/or graph and that the task needed to involve the student in explaining the representation. An example shared by the expert was that MP 4 would likely be used in conjunction with MP 3: Construct Viable Arguments and Critique the Reasoning of Others, on a frequent basis for students to show proficiency with using MP 4.

Content expert 1 also commented that an assessment item that allowed students to represent their knowledge using MP 4 would be most likely be represented as a

constructed-response item; however, the content expert pointed out several possible examples of multiple-choice tasks on the JCPS interim assessments that, where the tasks connected, could represent MP 4. The multiple-choice items on the JCPS interim assessments were stand-alone items so that students did not need to refer to previous items or responses to answer subsequent tasks. The content expert explained that each item asked students to show mathematical representations in different ways. Each task gave multiple-choice answers, one task asked for a graph to represent an equation; one task asked for an equation representing a graph; and one task asked students to choose the least likely explanation of slope.

Content expert 1 argued that, had those items been linked together with one constructed-response task the fact that students were given multiple-choice options may lower the cognitive complexity but would still be a representation of MP 4. However, two of the content experts believed that a task allowing students to show proficiency in the use of MP 4 would involve an extensive task that would be linked to a real-life situation that students would encounter on their own. The second content expert's suggestion was that the task would result from something found in the newspaper. Thus, the problem would begin from a task that allowed students to connect on a more personal level and that students would understand how they might go about representing their understanding and model their understanding with the mathematics that would be needed to understand the problem. Moreover, content expert 2 argued that the problem did not need to be high in cognitive complexity on an 8th-grade level, but that the students would need to use reasoning to determine the correct mathematics to represent their answers.

A third content expert believed that, depending upon the grade level that the problem was presented to, experts 1 and 2 were correct. This third content expert held that, in 8th-grade, students should need to show their understanding in multiple ways for the same problem; thus agreeing with content expert 1, but that the problem would also need to be an extended, performance type assessment that would likely take a longer time frame to complete. Therefore, content expert 3 agreed with content experts 1 and 2 descriptions of portions of MP 4.

Discussions such as the one regarding MP 4 continued for each of the CCSS Mathematical Practices, and continued for two hours. The discussion amongst the experts connected to the literature in that the practices are difficult to interpret (Rasmussen et al., 2011) and are difficult to determine the methods with which they should be identified (Hull, Miles, & Balka, 2012). The content experts that participated in this study voiced concern regarding the instructions for which the practices were to be identified or applied as well as how to interpret the practices in application to assessments. Consistent with McCallum's (2012) description of the intent of the practices in the literature, the content experts agreed that the practices would not be utilized in the same methods for all problems and in all grades.

The experts shared a similar concern regarding the actual process of coding the data. Each of the experts expressed that they did not feel the data they supplied were true representations of their opinions. The experts stated they had assigned practices only because they were requested to do so. The NRC (2011) describes the mathematical proficiencies, a previous identification of mathematical practices, as ways to ensure students become proficient in mathematics, "enable them to cope with the mathematical

challenges of daily life and continue their study of mathematics in high school and beyond” (p. 118).

The team of content experts also discussed similar thoughts regarding what the practices should enable children to do, but expressed concern as to how assessment items could identify student’s ability to cope with mathematical challenges. The final meeting conducted with the content experts provided invaluable perspective in regards to the purpose of the mathematical practices within assessment items. Interestingly, all of the experts, despite each having taught middle and high school mathematics, had different interpretations of the practices. However, this finding is consistent with the literature in that the coding of the data could likely have been effected by the experiences of each content expert. The researcher did not anticipate this particular finding since all experts had middle school teaching experience. Thus, in future research consideration should be made when selecting content experts and analyzing data. If expert experience is shown to affect coding of assessment items for mathematical practices, the researcher believes expert experience would likely affect the design of assessment items as well. Therefore, when assessment design committees are created, a broad array of experience among the team members should be included in order to ensure perspective from multiple teaching experiences are taken into consideration.

Additionally, the researcher believes that prior to coding of mathematical practices, a meeting should have taken place to discuss operational definitions of the practices in assessment items. A training and collaboration meeting prior to coding of the mathematical practices should have included support resources for experts to refer to during the coding process. Had the meeting taken place prior to coding, the content

experts may have been able to come to consensus on possible mathematical practice identifiers in assessments. Therefore, the coders could have utilized the initial operational definitions or practice identifiers to code the assessment items and would have been able to discuss in more depth the changes that would need to be made in order to continue the research. The data from the initial coding would likely have been more representative of the assessment items that would have allowed the researcher to share initial mathematical practice findings. In order to avoid publishing inaccurate, inconsistent, and unusable data that may effect administrator decisions regarding the assessment data, the researcher determined that the specific mathematical practice quantitative data should not be presented in this context.

Research Question 3

The purpose of research question 3 was to determine which of the three assessments presented the highest level of content validity when aligned with the 8th-grade CCSSM content standards. The third question was also designed to determine which of the three assessments had the highest representation of the CCSS Mathematical Practices in the assessment items. In essence, the intention of this question was to supply administrators with the information they would need to determine which assessment could provide them with the most information in application to the CCSSM.

After each of the analyses were completed from the WAT data and the CCSS Mathematical Practices data, the evidence showed that the JCPS interim assessments had the highest 8th-grade CCSSM content validity. By representing a high level of content validity, the interim assessments could be used to predict future results on large-scale assessment. In this case, the JCPS interim assessment data may be used to predict student

proficiency on the national assessments. However, to ensure this is indeed possible, the state assessment would also need to undergo a similar alignment process to determine the level of content validity represented. Both assessments would need to be similar for the interim assessments to be reliable predictors of achievement.

The final determination of the assessment that showed the highest representation of the CCSSM practices could not be made. The findings from the quantitative and qualitative data reveal that the identification of mathematical practices is a difficult process that warrants further study to achieve consistency in identification methods. The researcher believes that the processes are indeed represented within the assessment items. However, to assign the practices, a consistent method needs to be created in order to obtain reliable data in application to the practices. The mathematical practices are intended to support students in creating intellectual connections between the mathematics procedures to real-world, contextual mathematical situations. Thus, tasks utilizing the mathematical practices would be of higher cognitive demand than those tasks that are based solely on assessing students' procedural knowledge. If one were to consider cognitive complexity levels in lieu of the mathematical practices, both the JCPS interim assessments and the 2009 NAEP released items appear to assess students at a higher level of cognitive demand. Therefore, if cognitive complexity were a consideration, the overall best assessment for determining student progress toward meeting proficiency on the CCSSM would be the JCPS interim assessments.

Depth-of-knowledge levels 3 and 4 are represented within NAEP items despite the evidence showing that content validity of the NAEP assessment items did not align to the 8th-grade CCSSM content standards. Depth-of-knowledge levels 3 and 4 represent the

use of conceptual mathematics skills and real-world applications requiring a higher demand of student thinking to answer the questions. Thus, administrators and educators could consider the use of assessments including items with high DOK levels to indicate whether their students are performing well on conceptual use of mathematics skills. In this case, the NAEP may be usable to determine whether students are making progress in application of mathematics, not just procedural mathematics.

Conclusions and Implications

Although this research focused on overall alignment of the assessments to the 8th-grade CCSSM, several additional important findings emerged from this research. First, to answer research question 1 and 3 a determination of the overall best test must be made. As an overall assessment, the JCPS interim assessments are the most thoroughly aligned to the 8th-grade CCSSM content standards; thus, showing the highest level of content validity. The findings of the study verify the use of the interim assessments for accountability, progress monitoring, and as future predictors of 8th-grade CCSSM achievement.

The researcher intended to strengthen support for conducting this study by utilizing a principal component analysis to determine the statistical representation of CCSSM domains within the JCPS interim assessments and the ACT EXPLORE. Although complete ACT EXPLORE data was not obtained from the district, the researcher was able to perform the analysis on the sample of JCPS interim assessment student item answers. Despite the fact that the interim assessments are given as five separate assessments in five separate instances, the approximately 3000 students' data were input as if the students took one assessment. Thus, student data for 89 items were

entered into SPSS and analyzed as a whole one time in lieu of five separate analyses for each assessment. The researcher believes the method of data input may have affected the results of the PCA. Therefore, researchers should consider methods for using the data in accordance with when the assessments were given. For example, future research reanalyzing the PCA in this study would include the analysis of each of the five assessments separately.

The researcher recommends future use of PCA to determine whether assessment items are correlated to each other. A correlation of items would show that the test does indeed assess student knowledge of similar standards. In essence, the correlation of items would indicate the assessment is reliable for determining student knowledge of standards by showing the relationship among the items. Moreover, a strong correlation of items would allow a researcher to analyze PCA results to determine if the items show variation, or extracted components. The indication of extracted components would allow the researcher to determine the items that represent specific domains or standards. Thus, by doing a PCA, an alignment study is strengthened by allowing the researcher to compare alignment tool results, in this case the WAT data, to that of the PCA results. The comparison of the results from each type of analysis would present evidence of one of two possible scenarios: (1) The comparison would show either similar results identifying items with aligned standards and items aligned within similar domains, or, (2) The comparison would show dissimilar alignment. In the second scenario, the researcher would need to assess the reason for the difference in analysis results. Therefore, the researcher concludes that any researcher utilizing PCA in application to alignment of test items should consider both, input methods of data within a statistical analysis package,

when the PCA should be conducted within the study (before use of the alignment tool with content experts or after), the number of items on the examined assessment, and the method in which the assessments are given to students (at separate intervals or during one session).

The inclusion of lower grade levels did seem to verify a higher level of content validity of the national assessments. The ACT EXPLORE and NAEP assessments are not aligned with a specific set of national standards or state standards; each of the assessments utilizes their own set of standards or frameworks. Therefore, the likelihood of the two assessments meeting all acceptable levels of content alignment with the CCSSM was not high. Since the CCSSM is based upon several frameworks and sets of standards, it is interesting to consider the differences in the CCSSM compared to those standards and frameworks. Although the researcher anticipated lower alignment of the national assessments with the CCSSM due to the inclusion of multiple frameworks, the final results were still surprising.

The researcher believes the findings of cognitive complexity percentages of each assessment were interesting. Each of the assessments met overall DOK requirements for the WAT despite the lack of content alignment for the ACT EXPLORE and NAEP. However, several important conclusions from the cognitive complexity findings were discovered. The NAEP released items and the JCPS interim assessments each included constructed-response items and also showed items representing higher cognitive complexity levels. However, the ACT EXPLORE included only multiple-choice items that were determined to assess cognitive demand at DOK levels 1 or 2. Thus, future research may also include the linking of item formats with alignment processes.

The CCSS Mathematical Practices were determined to be extremely difficult to identify in the method that was used in the research study. The researcher discovered that experience of content experts and limited instructions for identifying the mathematical practices affected the results of the CCSSM Practices alignment. The researcher discovered discrepancies among the data that were explained during the final content expert discussion. First, the same practice – different definition discrepancy occurred in several assessment items where content experts identified the same practice; however, the explanation and definition of the reasons for assigning the practice were different. Conversely, the different practice – same definition discrepancy occurred in several items where content experts identified differing mathematical practices, but cited the same or similar definition. This phenomenon can be linked to the relationship among the CCSSM Practices, NCTM PSSM, and the NRC mathematical proficiencies. Thus, when future studies take place analyzing the mathematical practices within assessment items, the establishment of identifying properties and definitions should occur prior to coding.

The study also supplied evidence to support the Porter et al. finding in regards to the experience level and experience type of content experts. It is important to consider the previous experiences held by content experts. The researcher took into consideration the recommendations from the review of the literature in regards to the number of content experts needed for optimal reliability in the coding process. Moreover, the researcher considered the findings of the Porter et al. study and determined that utilizing educators with experience teaching 8th-grade mathematics would allow for stronger reliability. However, the researcher discovered and would like to draw attention to additional considerations in selecting content experts. The content experts in this study were able to

perform the tasks requested during the content alignment process with very few issues or questions; however, when discussing the mathematical practices, the experiences of the content experts in differing settings did play an important role in decisions made. Therefore, to determine a more consistent and reliable method for identifying the practices in assessment items, experience and teaching philosophies of discussants should also be taken into consideration.

Limitations of this Study

In the design and implementation of this research, there were several issues that occurred that had an impact on the results of the study. First, the PCA conducted in this study was chosen to provide support about how the CCSSM strands were evident in the JCPS interim assessments and the ACT EXPLORE. The process was not possible for two reasons. First, the ACT EXPLORE student data provided to the researcher was not complete. In addition, only half of the student item responses were given for that specific assessment form. The researcher attempted to alleviate this issue through contacting representatives in the district, at the state department of education, and the ACT regional offices. The correspondence occurred for approximately three months with no final conclusion for obtaining the necessary data. Thus, the PCA could not be completed on the data because missing information would provide incomplete and inaccurate results.

Second, the JCPS interim assessments were initially given as five separate exams with 12 to 19 items on each assessment. This study analyzed the interim assessments as one document with 89 items. The lack of correlation of the assessment items, which caused the PCA analysis to be unusable, may have been due to the time lapse between the assessments being given, or due to the low number of items on each separate assessment.

For example, extracted components from the PCA should have identified items within different domains; however, due to the lack of correlation, 32 components were identified. With that many components extracted and the low level of correlation, only two or three assessment items fell within the components, which allows for no identifiable domains.

Results of the research also indicate that the number of items on an assessment may affect the reliability of the Webb Alignment Tool. That is to say, the alignment tool could be considered a limitation for assessments having few items. With a requirement of a standard to have six assessment items testing that standard, it could be, in some circumstances, difficult to meet the six-item requirement on a test with a lower number of items. Moreover, the CCSSM content standards are written in such a way that multiple standards could be represented in one assessment item. In addition, similar to the findings by Lombardi (2006), the number of assessment items testing multiple standards affects the outcome of the quantitative data resulting from the coding. The WAT also includes the use of Webb's DOK definitions and identifiers. Those identifiers did not match the method in which each of the assessment designs identified level of cognitive complexity. Therefore, the four levels of DOK present a limitation in the lack of cohesiveness with the design of the assessments. The DOK levels assigned to each of the CCSSM content standards were also difficult to determine. This is because the wording of the standards on the CCSSM is very broad and limited in specific examples. For instance, the 7th-grade Geometry Domain includes one standard chunking area, volume, and surface area of two- and three-dimensional figures within the same standard. When the WAT model was designed, standards identified student content knowledge requirements much more

specifically. That is to say, there were more state standards for each grade level in the years prior to the CCSSM and the standards were more specific. For instance, area would be one standard, surface area would be a second standard, and volume would be a third standard which are now all included within the one 7th-grade CCSSM standard. Therefore, identifying DOK and standard match of assessment items is not consistent with the current set of standards compared to previous state standards. Thus, the researcher believes the application of the WAT in relation to DOK of the standards may now show lower DOK levels on the CCSSM than the previous, more specific, state standards.

The assessments studied within this research were utilized in previous years. The 2009 NAEP assessment items that were analyzed may be considered a limiting factor to this study. However, because the NAEP is a standardized assessment, in order for the assessment to remain consistent in longitudinal reliability of score comparisons, the assessment cannot change significantly from year to year. The technical manuals for the NAEP report that assessment items must be piloted, field tested, and tested for fairness, reliability, and validity before the items can be added to the assessment. Information regarding the NAEP assessments indicates that only one block of items is changed per assessment year (National Center for Education Statistics, 2010). The new assessment items go through an extensive process well in advance of the year they are used for student proficiency assessment. The assessment items for the 2011 NAEP assessment were field tested in 2009 and no indication of items to be changed for the 2013 NAEP has been noted on the National Center for Education Statistics website (IES, 2011). There is additional support for consideration of the continued usefulness of the findings presented

within this study. The NAEP Mathematics Frameworks for both the 2009 and 2011 assessments report identical information for the distribution of items on the assessments across the strands (NCES, 2009, 2011).

Similar to the NAEP, the ACT EXPLORE assessment analyzed was used to test students in the fall of 2010. It is reasonable to assume the assessment has changed over time; however, since the ACT EXPLORE is a nationally utilized assessment and student progress is monitored from year to year, the assessment cannot change drastically from one year to the next. Otherwise, the statistical reliability of the assessment would be jeopardized. Also, the ACT EXPLORE pilots, field tests, and conducts item analyses prior to use on subsequent assessments. ACT reported in June of 2010 that the ACT EXPLORE aligned 100% with the 8th-grade CCSSM content standards and 88% with the mathematical practices. Since the report was released prior to the release of the final CCSSM, the ACT assessments analyzed for content alignment compared to the CCSSM would have had to be created prior to 2010. ACT also cited that ACT played a leadership role in the creation of the CCSSM and that the Standards for College and Career Readiness were taken into consideration during the design process. Thus, although this research in this study analyzed an ACT EXPLORE assessment from the fall of 2010, the researcher believes the date limitation may not be important. If ACT already reported 100% alignment, even if items were changed, the test item distributions across the CCSSM content standards are not likely to have been affected greatly.

The JCPS interim assessments analyzed were used in the academic year prior to the beginning of this study. Therefore, the assessments were designed in application to the CCSSM but were designed for the second year of state use of the CCSSM. Thus, the

assessment for the current academic year may be more highly aligned with both the content standards and the mathematical practices.

The findings from the discussion of content experts further reinforce the need for more research on methods for and strategies for recognizing the evidence of mathematical practices in assessment items. The review of the literature cited several instances of authors referencing the need for inclusion of mathematical practices in assessment items if teachers are to take them seriously. However, there is not yet a consistent method for doing so. Moreover, although an attempt was made to identify practices in the assessments analyzed in this study, no further progress was made in determining a high-quality method for recognizing the practices. If assessment of student proficiency in use of the mathematical practices is to take place, districts should use caution in utilizing the three exams analyzed in this study until further research on the practices takes place.

An additional limitation to this study is in regards to the fast pace of changes being made to assessments examining student knowledge in application to the CCSSM. This study was limited to the assessments and standards available at the time. Since the CCSSM and respective assessments are being transformed quickly, this research added to the field of alignment and methods for identifying the mathematical practices. As assessment designers continue to transform the ACT EXPLORE and NAEP assessments to reflect changes in the CCSSM, the levels of alignment in application to those assessments will also change.

Recommendations for Future Research

The researcher suggests further content analysis of NAEP and ACT EXPLORE in application to CCSSM content standards for grade levels below the 7th-grade. By doing this, the research would represent a deeper understanding of the focus of each assessment. For example, if assessment items on the NAEP represent a higher number of 5th-grade CCSSM content standards, administrators may be able to utilize information to link to 5th-grade proficiency.

Further examination of the two national assessments' alignment using a different alignment tool could help assure the reliability of the WAT use in this study. For example, the use of the Achieve or Survey of Enacted Curriculum (SEC) tools, although the literature indicates they are more thoroughly aligned to instructional practices and classroom assessments, may reveal different content validity results than those found in this study. The researcher does not have access to the Achieve or SEC tools to determine specific items that may be aligned differently with the CCSSM based upon the method of alignment using the different tools. However, utilizing a different tool would either corroborate the findings of this study or would identify areas that were not supported by the use of the WAT.

The researcher recognizes the need for further study in the method for identifying the mathematical practices within assessments. The researcher believes it is necessary to create assessments of students' conceptual knowledge of mathematics; therefore, inclusion of the practices would be necessary to assess conceptual knowledge. In order to assess the practices, student conceptual knowledge, we need to be certain there is a viable

method for consistently including and identifying the practices in assessment items. Currently, there are support systems being developed for educators to find and use tools in their classrooms to assist in developing tasks for both instructional and assessment purposes that require students to utilize mathematical practices to complete the task. Websites such as the Illustrative Mathematics site provide teachers of different grade levels with example tasks as well as videos showing students working through the task. Additionally, for those teachers in states that have committed to using the national assessments designed by the Partnership for Assessment of College and Career Readiness (PARCC), it is important for those educators to preview the sample items that are being created for the assessments. The Illustrative Mathematics site represents practice tasks that are extended-response, performance-based items for teachers to begin using in classrooms to familiarize students to using their knowledge in in-depth and contextual settings. Whereas the PARCC site supplies constructed-response assessment items in constructed-response, fill-in-the-blank, and multiple-choice formats that use the practices within the tasks. The PARCC items require students to apply initially given information to multiple questions in multiple formats. Both support sites do indicate which mathematical practices are utilized in the tasks; thus, teachers are able to connect the types of classroom tasks and assessment items to the representative practices.

This study did not include student assessment data or the entire pool of NAEP mathematics assessment items. However, it was noted that the released items represent the assessment items fairly equally. Never the less, the researcher feels examining the released items from the 2011 NAEP assessment would increase quality of an alignment study of NAEP to the CCSSM. By examining the newest set of released items, the study

would corroborate or disprove the statement that the released items represent the assessment. The researcher posits that the alignment of additional released items should be similar to the extent that the released items of each year represent the CCSSM content standards. Thus, to support the use of released items for alignment purposes, future research should include NAEP released items of the most current assessment year.

A final suggestion is to complete additional alignment studies of future assessments. The researcher feels that a similar study utilizing the newest versions of the three assessments is necessary due to the rapid transformations of assessments and possible future transformations of the CCSSM content. Despite the fact that both national assessments cannot change immensely from year to year, continued analyses of both exams should be completed. Further studies in contexts similar to this study could lend validity to contribute to the research in this area and provide information to administrators, educators, assessment designers, test data analysts, and curriculum specialists.

Summary

This study informed administrators for JCPS and test developers of the content validity of the JCPS interim assessments, the ACT EXPLORE, and the 2009 NAEP mathematics released items regarding the extent of content validity of each assessment. The researcher felt that the lack of correlation among test items on the JCPS interim assessment was puzzling. The researcher had hoped to be able to compare the strands extracted from the factor analysis to those represented by the findings from the WAT. Since this was not a possibility, the researcher was only able to report what was found through qualitative data that were analyzed using quantitative methods. Thus, the

findings focused on the results from the use of the WAT. Although there were no direct links to the literature specific to the JCPS interim assessments, the researcher felt the results of the data gathered from the WAT were accurate due to the extent of statistical reliability found. Thus, the study was able to inform designers of the JCPS interim assessments regarding areas for improvement and strengths of the interim assessments. In addition, the researcher was surprised at the lack of alignment of the items from the ACT EXPLORE and the NAEP released items to the 8th-grade CCSSM. Adding the analysis of the 7th-grade CCSSM content standards was unanticipated. Moreover, despite the addition of the extra analysis, the findings continued to be disappointing regarding the level of content covering either of the grade levels. Thus, the researcher hopes administrators will use caution when utilizing the ACT EXPLORE and NAEP data as suppliers of student proficiency information on 8th-grade CCSSM content.

Although the researcher agrees that mathematical practices should be considered when designing student assessment items, the researcher found the process of identification of the practices challenging and unhelpful within this study. However, the research does raise awareness of the need to find methods for being consistent in recognizing the practices in assessment items. By completing this task, both assessment and instructional designers can create more appropriate items to test student mathematics proficiency in applying mathematical practices. Thus, a final observation by the researcher, aligned with the literature on teacher instruction, is that if assessment creators pay particular attention to content validity and the mathematical practices of the CCSSM, educators will create better instruction so students can reach proficiency on the tests. Moreover, since the CCSSM was created to help students be prepared for college and

career, colleges, businesses, and communities will be more thoroughly populated with mathematically literate adults.

References

- Achieve, Inc. (2006). *An alignment analysis of washington state's college readiness mathematics standards with various local placement tests*. Big Bend Community College: Washington Public Universities: Whatcom Community College.
- Achieve, Inc. (2010). Comparing the common core state standards in mathematics and the NAEP framework. *Achieving the common core*. Iowa City: IA.
- Achieve, Inc. (2001). *Measuring up: A report on education standards and assessments for Massachusetts*. Achieve, Inc. Retrieved from <http://www.achieve.org/files/MassachusettsBenchmarking10-2001.pdf>
- Achieve, Inc. (2003). *Measuring up: A report on education standards and assessments for Montgomery County*. Achieve, Inc. Retrieved from <http://www.achieve.org/measuring-reort-education-standards-and-assessments-montgomery-county>
- ACT, Inc. (2011). *Affirming the goal: Is college and career readiness an internationally competitive standard?* Retrieved from: <http://www.act.org/research/policymakers/pdf/AffirmingtheGoal.pdf>
- ACT, Inc. (2010a). *A first look at the common core and college career readiness*. Retrieved from <http://www.act.org/research/policymakers/reports/firstlook.html>
- ACT, Inc. (2006a). *Assessing the college readiness in reading of eighth- and ninth-grade students using act's explore*. Iowa City, IA.

- ACT, Inc. (2005). *Do current state standards and assessments reflect college readiness?: A case study*. Iowa City: IA.
- ACT, Inc. (1995). *EXPLORE: Helping students discover a world of possibilities. A content description of ACT's EXPLORE program*. Iowa City: IA.
- ACT, Inc. (2007a). *EXPLORE: Technical manual*. Iowa City: IA.
- ACT, Inc. (2011). *EXPLORE: Technical manual*. Iowa City: IA.
- ACT, Inc. (2009). *Issues in college readiness: Readiness and success: Statewide implementation of EXPLORE and PLAN*.
- ACT, Inc. (2006c). *Reading between the lines: What the act reveals about college readiness in reading*. Iowa City, IA.
- ACT, Inc. (2007b). *High school course rigor study*. Iowa City, IA.
- ACT, Inc. (2006b). *Ready for college and ready for work: Same or different?* Iowa City, IA.
- ACT, Inc. (2007c). *Rigor at risk: Reaffirming quality in the high school core curriculum*. Iowa City: IA.
- ACT, Inc. (2010c). *The alignment of common core and act's college and career readiness system*. Iowa City: IA.
- ACT, Inc. (2008). *The forgotten middle: ensuring that all students are on target for college and career readiness before high school*. Iowa City: IA.
- American Educational Research Association (AERA). (2006). Do the math: Cognitive demand makes a difference. *Research Points*. Author.

- American Educational Research Association (AERA). (2003). Standards and tests: Keeping them aligned. *Research Points: Essential Information for Education Policy, 1(1)*.
- American Federation of Teachers. (1998). *Setting higher sights: A need for more demanding assessments for U.S. eighth graders*. Washington, DC: US Department of Education.
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. Boston: Allyn & Bacon.
- Anderson, L.W. and Krathwohl, D.R. (Eds). (2001). *A taxonomy for learning, teaching and assessing*. New York: Longman
- Arenson, E. (2000). Estimating error in state-to-NAEP linkages, part I: mean plausible value distributions from a simple model. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Arenson, E.A. (1999). Statistical linkages between state education assessments and the national assessment of educational progress. Paper presented at the Annual Meeting of the Sacramento Statistical Association, Sacramento, CA.
- Bandeira de Mello, B., Blankenship, C., McLaughlin, D., & Rahman, T. (2009). *Mapping state proficiency standards onto NAEP scales: 2005-2007*. Institute of Education Sciences: National Center for Education Statistics (NCES 2010-456).

- Benton, M., Tremaine, M., & Scher, J. (2004). *Computer aids for designing effective multiple choice questions*. Paper presented at the Americas Conference on Information Systems.
- Blank, R.K. (2007). *Alignment content analysis of NAEP 2007 mathematics assessment using the surveys of enacted curriculum methodology*. The Council of Chief State School Officers. Retrieved from <http://programs.ccsso.org/content/pdfs/NAEP%20math%20alignment%20report%2007.pdf>
- Blank, R.K. (2005). *Surveys of enacted curriculum: Tools and services to assist educators*. The Council of Chief State School Officers: Washington, DC.
- Blank, R.K, Porter, A. & Smithson, J. (2001). *New tools for analyzing teaching, curriculum, and standards in mathematics & science*. Results from the Survey of Enacted Curriculum Project, Final Report. Washington, D.C.: CCSSO.
- Bleiler, S.K. & Thompson, D.R. (2013). Multidimensional assessment of CCSSM. *Teaching Children Mathematics, 19*(5), 292-300.
- Bloom, B.S. (Ed.) Englehart, M.D., Furst, E.J., Hill, W.H. & Krathwohl, D.R. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: David McKay.
- Brown, R.S. & Conley, D.T. (2007). Comparing state high school assessments to standards for success in entry-level university courses. *Educational Assessment, 12*(2), 137-160.
- Buckendahl, C.W., Davis, S.L., Plake, B.S., Sireci, S.G., Hambleton, R.K., Zenixky, A.L., & Wells, C.S. (2009). *Evaluation of national assessment of educational*

progress: Final report. Lincoln, Nebraska: Buros Institute for Assessment Consultation and Outreach.

Bulkley, K.E., Christman, J.B., Goertz, M.E., & Lawrence, N.R. (2010). Section 2: Case study: benchmarks in the school district of Philadelphia: Building with benchmarks: The role of the district in Philadelphia's benchmark assessment system. *Peabody journal of education*, 85, 186-204.

Carmichael, S.B., Martino, G., Porter-Magee, K. & Wilson, W.S. (2010). *The state of state standards - and the common core - in 2010*. Retrieved from the Thomas B. Fordham Institute website: <http://www.edexcellence.net/publications/the-state-of-state-of-standards-and-the-common-core-in-2010.html>

Carmichael, S.B., Wilson, W.S., Finn, Jr., C.E., Winkler, A.M. & Palmieri, S. (2009). *Stars by which to navigate: Scanning national and international education standards in 2009: An interim report on common core, NAEP, TIMSS and PISA*. Retrieved from The Thomas B. Fordham Institute website: <http://www.edexcellence.net/publications/stars-by-which-to-navigate.html>

Chen, J.C., Reys, B.J. & Reys, R.E. (2009). Analysis of the learning expectations related to grade 1-8 measurement in some countries. *International Journal of Science and Mathematics Education*, 7(5): 1013-1031.

Christman, J.B., Neild, R.C., Bulkley, K., Blanc, S. Liu, R., Mitchell, C., & Travers, E. (2009, June). *Making the most of interim assessment data: Lessons from Philadelphia*. Research for Action, Philadelphia.

- Chudowsky, N. & Chudowsky, V. (2011). *State test score trends through 2008-09, part 3: student achievement at 8th grade*. Center on Education Policy: Washington, D.C.
- Clune, W.H. & White, P.A. (2009) *Policy effectiveness of interim assessment in providence public schools* (WCER working paper no. 2008-10). (2010) Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research. Retrieved from the Wisconsin Center for Education Research website:
<http://www.wcer.wisc.edu/publications/workingPapers/papers.php>
- Cobb, P. & Jackson, K. (2011). Assessing the quality of the Common Core State Standards for Mathematics. *Educational Researcher*, 40, 183-185.
- Common Core State Standards Initiative (CCSSI). (2010). *Draft K-12 common core state standards available for comment*. Retrieved from the CCSSI website:
http://www.nga.org/cms/home/news-room/news-releases/page_2010/col2-content/main-content-list/title__draft-k-12-common-core-state-standards-available-for-comment.html
- Common Core State Standards Initiative (CCSSI). (2012). *In the states*. Retrieved from:
<http://www.corestandards.org/in-the-states>
- Confrey, J. & Krupa, E. (2010). *Curriculum design, development, and implementation in an era of common core state standards: Summary of a conference*. Retrieved from the Center for the Study of Mathematics Curriculum website:
<http://mathcurriculumcenter.org/conferences/ccss/SummaryReportCCSS>

Council of Chief State School Officers (CCSSO). (2012a). *Alignment: Models*. Retrieved from the Council of Chief State School Officers website:

http://programs.ccsso.org/projects/Alignment_Analysis/Models/

Council of Chief State School Officers (CCSSO). (2012b). *The common core state standards initiative (CCSSI)*. Retrieved from the Council of Chief State School Officers website:

http://www.ccsso.org/Resources/Programs/The_Common_Core_State_Standards_Initiative.html.

Council of Chief State School Officers (CCSSO). (2012c). *Common core state standards for mathematics*. Retrieved from the Council of Chief State School Officers website: http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Council of Chief State School Officers (CCSSO). (2002). *Models for alignment analysis and assistance to states*. Retrieved from the Council of Chief State School Officers website: <http://www.ccsso.org/content/pdfs/AlignmentModels.pdf>

Cronin, J., Dahlin, M. Adkins, D. & Kingsbury, G.G. (2007). *The proficiency illusion*. Retrieved from The Thomas Fordham Institute website:

http://www.edexcellencemedia.net/publications/2007/200710_theproficiencyillusion/Proficiency_Illusion_092707.pdf

Daro, P., Stancavage, F., Ortega, M., DeStefano, L., & Linn, R. (2007). *Validity Study of the NAEP Mathematics Assessment: Grades 4 and 8*. American Institutes for Research and the National Center for Education Statistics: U.S. Department of Education.

- Davis, B., Caros, J., Grossen, B., & Carnine, D. (2002). *Initial stages in the development of benchmark measures of success: Direct implications for accountability* (Research Report #11). Washington, DC: U.S. Department of Education.
- Enright, M.K., Allen, N., & Kim, M.I. (1993). *A complexity analysis of items from a survey of academic achievement in the life sciences*. Princeton, NJ: ETS Office of Corporate Planning and Development.
- Erickson, V., Ho, A., Holtzman, D., Jaciw, A., Lukoff, B., Shen, X., Wei, X., & Haertel, E. (2007). *Closing the gap? A comparison of changes over time in white-black and white-hispanic achievement gaps on state assessments versus state NAEP* (CSE Report 721). National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Los Angeles, CA.
- Ercikan, K. (1997). Linking statewide tests to the national assessment of educational progress: Accuracy of combining test results across states. *Applied Measurement in Education, 10*(2), 145-159.
- Flowers, C., Browder, D., & Ahlgrim-DeLzell, L. (2006). An analysis of three states' alignment between language arts and mathematics standards and alternate assessments. *Exceptional Children, 72*(2), pp. 201-215.
- Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R.R. Ronning, Glover, J.A., & Witt, J.C. (Eds.). *Buros-Nebraska Symposium on Measurement and Testing: Vol. 3. The Influence of Cognitive psychology on testing*. Hillsdale, NJ: Lawrence Erlbaum.

- Goertz, M.E., Olah, L.N., & Riggan, M. (2009, December). *From testing to teaching: The use of interim assessments in classroom instruction* (CPRE Research Report # RR-65). The Consortium for Policy in Research in Education.
- Greene, J.P., Winters, M.A., & Forster, G. (2003). *Testing high stakes tests: can we believe the results of accountability tests?* Manhattan Institute for Policy Research, New York, NY.
- Hancock, G.R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *Journal of Experimental Education*, 62(2), 143-157.
- Herman, J.L., Webb, N.M., & Zuniga, S.A. (2005). *Measurement issues in the alignment of standards and assessments: A case study* (CSE Report 653). National Center for Research on Evaluation, Standards, and Student Testing, CA.
- Hess, K.K. (2006). *Cognitive complexity: Applying Webb DOK levels to Bloom's taxonomy*. National Center for Assessment, NH.
- Hess, K.K., Jones, B.S., Carlock, D., & Walkup, J.R. (2009). *Cognitive rigor: Blending the strengths of bloom's taxonomy and Webb's depth of knowledge to enhance classroom-level process*. National Center for the Improvement of Education Assessment, NH.
- Hull, T.H., Miles, R.H., & Balka, D.S. (2012). *The common core mathematics standards: Transforming practice through team leadership*. Thousand Oaks, CA: Corwin.
- Institute for Education Sciences (IES). (2010). *An introduction to NAEP: National assessment of educational progress*. National Center for Education Statistics, US

Department of Education. Retrieved from nces.ed.gov/nationsreportcard/pdf/parents/2010468.pdf

Institute for Education Sciences (IES). (2009). *National assessment of educational progress: An overview of NAEP*. National Center for Education Statistics, US Department of Education. Retrieved from nces.ed.gov/nationsreportcard/pdf/parents/2009490rev.pdf

Institute for Education Sciences (IES). (2011). *NAEP: Pretests*. National Center for Education Statistics, US Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/tdw/overview/naep_pretests_2009.asp

Institute of Education Sciences (IES). (2012). *The nation's report card: Mathematics 2011: Trial urban district assessment results at grades 4 and 8* (IES Publication No. 2012-452). National Center for Education Statistics, US Department of Education. Retrieved from nces.ed.gov/nationsreportcard/pdf/dst2011/2012452.pdf

Jacobs, J.K., Hiebert, J., Givvin, K.B., Hollingsworth, H., Garnier, H., & Wearne, D. (2006). Does eighth-grade mathematics teaching in the United States align with the NCBM Standards? Results from the TIMSS 1995 and 1999 video studies. *Journal for Research in Mathematics Education*, 37, 5-32.

Jefferson County Public School District (JCPS). (2010). *CASCADE assessment system: Analysis and usage report*. Jefferson County Public School District Accountability, Research and Planning.

- Jones, D.L. Tarr, J.E. (2007). An examination of the levels of cognitive demand required by probability tasks in middle grades mathematics textbooks. *Statistics Education Research Journal*, 6(2), 4-27.
- Johnson, E.G. (1992). The national assessment of educational progress. *Journal of educational measurement*, 29(2), 95-110.
- Kentucky Department of Education (KDE). (2012a). *Mathematics crosswalk*. Retrieved from:
<http://www.education.ky.gov/KDE/Instructional+Resources/Curriculum+Documents+and+Resources/Program+of+Studies/Crosswalk+-+Common+Core+State+Standards+%28CCSS%29+Comparison+to+Kentucky+State+Standards.htm>
- Kentucky Department of Education (KDE). (2012b). *Kentucky core academic standards*. Frankfort: Kentucky Department of Education. Retrieved from:
www.education.ky.gov/users/otl/POS/kentuckycommoncore_mathematics.pdf.
- Kentucky Department of Education (KDE). (2007). *Support materials for core content for assessment version 4.1*. Frankfort: Kentucky. Retrieved from
http://www.education.ky.gov/users/otl/CCA_DOK/CCA_DOK_SUPPORT_808_Mathematics.pdf
- Kepner, H.S. & Huinker, D. (2012). Assessing students' mathematical proficiencies on the common core. *Journal of Mathematics Education at Teachers College*, 3, 26-31.

- Kiplinger, V.L. & Linn, R.L. (1994). *Linking statewide tests to the national assessment of educational progress: Stability of results*. Paper presented at the Annual Meetings of the American Educational Research Association in New Orleans, LA.
- Krupa, E. (2011). *A summary report from the conference "Moving forward together: Curriculum & assessment and the common core state standards for mathematics"*. Retrieved from:
http://www.mathismore.net/resources/MovingForward/MFT_Final_Report.pdf
- LaMarca, P.M. (2001). Alignment of standards and assessments as an accountability criterion. *ERIC Clearinghouse on Assessment and Evaluation College Park MD*. Eric Digest. Retrieved from www.eric.ed.gov.
- LaMarca, P.M., Redfield, D., & Winter, P.C. (2000). *State standards and state assessment systems: A guide to alignment*. Council of Chief State School Officers: Washington, D.C.
- Linn, R. (1998). Validating inferences from national assessment of educational progress achievement-level reporting. *Applied Measurement in Education, 11*(1), 23-47.
- Linn, R.L., Graue, E. & Sanders, N.M. (1990, January). *Comparing state and district test results to national norms: Interpretations of scoring "above the national average"* (CSE Technical Report 308). Center for Research on Evaluation, Standards, and Student Testing.
- Lombardi, A., Seburn, M., Conley, D., & Snow, E. (2010). *A generalizability investigation of cognitive demand and rigor ratings of items and standards in an alignment study*. Presented at the annual conference of the American Educational Research Association in Denver, CO.

- Lombardi, T.J. (2006). An alignment study of the minnesota comprehensive assessment-II with state standards in reading and literature for grades 3-8 and 10. Retrieved from the Minnesota Department of Education website:
<http://www.google.com/search?client=safari&rls=en&q=An+alignment+study+of+the+minnesota+comprehensive+assessment-II+with+state+standards+in+reading+and+literature+for+grades+3-8+and+10.&ie=UTF-8&oe=UTF-8>.
- Lubienski, S.T. (2006). Examining instruction, achievement, and equity with NAEP mathematics data. *Education Policy Analysis Archives*, 14(14). Retrieved from <http://epaa.asu.edu/epaa/v12n12/>
- Marzano, R.J. (2001). *Designing a new taxonomy of educational objectives. Experts in assessment*. Thousand Oaks, CA: Corwin Press.
- Marzano, R.J. and Jesse, D.M. (1987). *A study of general cognitive operations in two achievement test batteries and their relationship to item difficulty*. Aurora, CA: Mid-continent Regional Educational Laboratory.
- Marzano, R.J. & Kendall, J.S. (2007). *The new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.
- McCallum, W. (2012). The U.S. common core state standards in mathematics. A Paper Presented at the 12th International Congress on Mathematics Education in Seoul, Korea on July 12, 2012.
- McMillan, J.H. (2008). *Assessment essentials for standards-based education*. Thousand Oaks, CA: Corwin Press.

- Messick, D. (1984). Educational achievement testing: The assessment of dynamic cognitive structures. In Plake, B.S. (Ed.). *Social and Technical Issues in Testing*. Hillsdale, N.J. Lawrence Erlbaum.
- Messick, D. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement* (3rd Edition). New York: American Council on Education Macmillan Publishing Company.
- Messick, S. (1993). Validity. In R.L. Linn (Ed), *Educational Measurement* (2nd ed. pp. 13-104). Phoenix: American Council on Education and Oryz Press.
- Mullis, A.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLSS International Study Center, Lynch School of Education, Boston College.
- Nabors Olah, L., Lawrence, N.R., Goertz, M.E., Weathers, J., Riggan, M., & Anderson, J. (2011? or 2009). *Testing to the test? Expectations and supports for interim assessment use*. Retrieved from the Consortium for Policy Research in Education website: <http://www.cpre.org/testing-test-expectations-and-supports-interim-assessment-use>
- National Assessment for Educational Progress. (2012). *The nation's report card: Trial urban district assessment*. Retrieved from: <http://nationsreportcard.gov/tuda.asp>
- National Assessment Governing Board (NAGB). (2008). *Mathematics framework for the 2009 national assessment of education progress*. Washington, DC: U.S. Department of Education.

- National Assessment Governing Board (NAGB). (2003). *Mathematics framework for the 2003 national assessment of education progress (College Board Rep. RN91084001)*. Washington, DC: U.S. Government Printing Office.
- National Assessment Governing Board (NAGB). (2010). *Mathematics framework for the 2011 national assessment of educational progress*. Council of Chief State School Officers.
- National Assessment Governing Board (NAGB). (2011). The national assessment of educational progress authorization act. Retrieved: October 15, 2011, from <http://www.nagb.org/who-we-are/naip-law.htm>
- National Center for Education Statistics (NCES). (2010). *Development of the mathematics cognitive items and instruments*. NAEP Technical Documentation. Retrieved from: <http://nces.ed.gov/nationsreportcard/tdw/instruments/cog.asp>
- National Center for Education Statistics (NCES). (2011) *NAEP: Grade 8 mathematics/reading/science: Sample questions: General information about the nation's report card 2011*. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/about/2011SQB_G8_MRS.pdf
- National Center for Education Statistics (NCES). (2009). *NAEP: Sample questions general information about the nation's report card: Grade 8 2009 mathematics, reading, and science*. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/demo_booklet/109SQ-G08-MRS.pdf
- National Council of Teachers of Mathematics (NCTM). (2013). Measurement Standard. Retrieved from: <http://www.nctm.org/standards/content.aspx?id=316>

- National Research Council (NRC). (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- No Child Left Behind Act of 2001, Pub.L. No. 107-110, 115, Stat.1425 (2002).
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA.
- National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. National Council of Teachers of Mathematics, Inc., Reston, VA.
- National Council of Teachers of Mathematics. (2012a). *Curriculum focal points: questions and answers*. Retrieved: January 2, 2012, from <http://www.nctm.org/standards/focalpoints.aspx?id=274>
- National Council of Teachers of Mathematics. (2004). *Rise in NAEP math scores coincides with NCTM standards*. Retrieved from www.nctm.org/news/release.aspx?id=766
- National Council of Teachers of Mathematics. (2012b). *Standards overview*. Retrieved: January 2, 2012, from <http://www.nctm.org/standards/content.aspx?id=26798>
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Partnership for Assessment of Readiness for College and Careers (PARCC). (2012a) 3-8 *assessments*. Retrieved from <http://www.parcconline.org/crd-pld-survey>

- Partnership for Assessment of Readiness for College and Careers (PARCC). (2011). *PARCC model content frameworks for mathematics*. Retrieved from <http://www.parcconline.org/sites/parcc/files/PARCCMCFfor3-8MathematicsFall2011Release.pdf>
- Partnership for Assessment of Readiness for College and Careers (PARCC). (2012b). *PARCC states*. Retrieved from: [http://www.parcconline.org/\(study+state\)](http://www.parcconline.org/(study+state))
- Paterson, J.S. (2002). *Linking on-line assessment in mathematics to cognitive skills*. IN: Proceedings of the 6th CAA Conference, Loughborough: Loughborough University. Retrieved from <http://hdl.handle.net/2134/1887>
- Pedulla, J.J., Abrams, L.M., Madaus, G.F., Russell, M.K., Ramos, M.A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston: Boston College, The National Board on Educational Testing and Public Policy, Lynch School of Education.
- Perie, M., Marion, S, Gong, B., & Wurtzel, J. (2007). The role of interim assessments in a comprehensive assessment system. *Measures that Matter*. Retrieved from the Achieve website website: <http://www.achieve.org/files/TheRoleofInterimAssessments.pdf>
- Polikiff, M.S., Porter, A.C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, 48(4), 965-995.

- Porter, A.C. (2006). Curriculum Assessment. In J. Green, G. Camilli & P. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 141-159). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Porter, A.C. (2002). Measuring the content of instruction: uses in research and practice. *Educational Researcher*, , 31(7), 3-14.
- Porter, A.C., Kirst, M.W., Osthoff, E.J., Smithson, J.L., & Schneider, S.A. (1993). Reform up close: An analysis of high school mathematics and science classrooms. New Brunswick, NJ: Consortium for Policy Research in Education.
- Porter, A.C., McMaken, J., Hwang, J, & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher* 40(3), 103-116.
- Porter, A.C., Polikoff, M.S. & Smithson, J. (2009). Is there a de facto national intended curriculum? Evidence from content standards. *Educational Evaluation and Policy Analysis*, 31, 238-268.
- Porter, A.C. Polikoff, M.S. Zeidner T., & Smithson, J. (2008). The quality of content analyses of state student achievement tests and state content standards. *Educational Measurement: Issues and Practice*, 27(4), 2-14.
- Rabinowitz, S., Roeber, E., Schroeder, C., & Sheinker, J. (2006). *Creating aligned standards and assessment systems* (Issue Paper 3). Council of Chief State School Officers.
- Rasmussen, C., Heck, D.J., Tarr, J.E., Knuth, E., White, D.Y., Lambdin, D.V., Baltzley, P.C., Quander, J.R., Barnes, D. (2011). Trends and issues in high school education: Research insights and trends. *Journal for Research in Mathematics Education*, 42, 3, 204-219.

- Resnick, L.B. and Resnick, D.P. (1991). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford and M.C. O'Connor (Eds.), *Changing assessments: Alternative view of aptitude, achievement, and instruction* (pp. 35-73). Boston: Kluwer Academic Publishers.
- Resnick, L.B., Rothman, R., Slattery, J.B., & Vranek, J.L. (2003). Benchmarking and alignment of standards and testing. *Educational assessment*, 9(1&2), 1-27.
- Reys, R., Reys, B., Lapan, R., Holliday, G., and Wasman, D. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, 34(1), 74-95.
- Roach, A.T., Elliott, S.N., Webb, N.L. (2005). Alignment of an alternate assessment with state academic standards: Evidence for the content validity of the wisconsin alternate assessment. *The Journal of Special Education*, 38(4), 218-231.
- Robinson, G.E. (1997). *What do NAEP state comparisons and rankings really mean?* Paper presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment in Colorado Springs, CO.
- Rochford, J.A., O'Neill, A., Gelb, A. (2010). *Ready to go: Using the EXPLORE test to increase 8th grade readiness for success.*
- Sanford, E.E. & Fabrizio, L.M. (1999). *Results from the north carolina -- NAEP comparison and what they mean to the end-of-grade testing program.* Paper presented at the Annual Meeting of the American Education Research Association, Montreal, Quebec, Canada.

- Schmidt, W. (2002). *Missed opportunities: How mathematics education in the U.S. puts our students at a disadvantage and what can be done about it* (Policy Report number 7). The Education Policy Center at Michigan State University. East Lansing, MI.
- Schmidt, W.H., Houang, R. & Shakrani, S. (2009). *International lessons about national standards*. Retrieved from The Thomas Fordham Institute website:
http://www.edexcellencemedia.net/publications/2009/200908_internationallessons/20090826_International_Lessons_Report.pdf
- Schmidt, W.H., Wang, H.C., & McKnight, C.C. (2005). Curriculum coherence: an examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies*, 37(5), 525-559.
- Shadish, W.R, Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Company.
- Shapley, K.L. & Brite, J. (2008). *Aligning mathematics assessment standards: Oklahoma and the 2009 national assessment of educational progress (NAEP)* (REL Technical Brief, REL 2008 - No. 010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education and Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee.edlabs>
- Sireci, S.G., Robin, F., Meara, K., Rogers, H.J., & Swaminathan. (2000). An external evaluation of the 1996 Grade 8 NAEP science framework. In N. Raju, J.W. Pellegrino, M.W. Bertenthal, K.J. Mitchell & L.R. Jones (Eds.), *Grading the*

nation's report card: Research from the evaluation of NAEP (pp. 74-100).
Washington, DC: National Academy Press.

Smithson, J. & Blank, R. (2006). *Indicators of quality of teacher professional development and instructional change using data from surveys of enacted curriculum: Findings from NSF MSP-RETA project*. The Council of Chief State School Officers. Retrieved from

http://jsmithson.wcerus.org/Reference/Indicators_TchrPD_Smithson-Blank.pdf

Stecher, B.M. and Barron, S. (2001) Unintended consequences of test-based accountability when testing in milepost grades. *Educational Assessment*, 7(4), 259-282.

Stoneburg, B.D. (2009) *A quick review of NCLB (state) and NAEP achievement levels and how they match*. Paper presented at NAEP state Service Center Spring Assessment Literacy Workshop, Bethesda, Maryland.

Subkoviak, M.J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.

Suen, H.K. (1990). *Principles of test theories*. Hillsdale, NJ: Erlbaum.

Tarr, J.E., Reys, R.E., Reys, B.J., Chavez, O., Shih, J., and Osterlind, S.J. (2008). The impact of middle-grades mathematics curricula and classroom learning environment on student achievement. *Journal for Research in Mathematics Education*, 39(3), 247-280.

Tashakkori, A. & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.

- Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.
- Timms, M., Schneider, S., Lee, C., & Rolfus, E. (2007). *Aligning science assessment standards: New Mexico and the national assessment of educational progress (NAEP)* (Issues & Answers Report, REL 2007-No. 021). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment for Educational Progress (NAEP). (2012). *The nation's report card*. Retrieved from: http://nationsreportcard.gov/math_2011/summary.asp?tab_id=tab2&subtab_id=Tab_1#chart
- U.S. Office of Special Education Programs (OSEP). (2005). *Ideas that work: Validity evidence*. Retrieved from the U.S. Department of Education Website: http://www.osepideasthatwork.org/toolkit/tk_validityevidence_a.asp
- Usiskin, Z. (2007). The case of the university of Chicago school mathematics project – secondary component. In C.R. Hirsch (Ed.), *Perspectives on the Design and Development of School Mathematics Curricula* (pp. 173-182), Reston, VA: National Council of Teachers of Mathematics.
- Uzzell, R., Simon, C., Horowitz, A., Hyslop, A., Lewis, S. & Casserly, M. (2011). *Beating the odds. Analysis of student performance on state assessments and*

NAEP: Results from the 2009-10 school year. Council of the Great City Schools, Washington, DC. Retrieved from:

<http://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/4/BTOX.pdf>

Vockley, M. & Vockley, L. (2009). *Alignment and the states: Three approaches to aligning the national assessment of educational progress with state assessments, other assessments, and standards.* Washington, DC: The Council of Chief State School Officers.

Walsh, W.B. (1995). *Tests and assessment.* New York: Prentice-Hall.

Webb, N.L. (1999). *Alignment of science and mathematics standards and assessments in four states* (Research Monograph No. 18). Madison, WI: National Institute for Science Education.

Webb, N.L. (2002a). *Alignment study of language arts, mathematics, science, and social studies of state standards and assessment in four states.* Washington, DC: Council of Chief State School Officers.192

Webb, N.L. (2002b). *An analysis of the alignment between mathematics standards and assessments for three states.* A paper presented at the American Educational Research Association Annual Meeting in New Orleans, Louisiana April 1-5, 2002.

Webb, N.L. (1997a). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Washington DC: Council of Chief State School Officers.

- Webb, N.L. (1997b). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 8). Madison, WI: National Institute for Science Education, University of Wisconsin-Madison.
- Webb, N.L. (2002c). *Depth-of-knowledge levels for four content areas*. Retrieved from <http://facstaff.wcer.wisc.edu>
- Webb, N.L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7-25.
- Webb, N.L., Alt, M., Ely, R., Cormier, M., & Vesperman, B. (2005). The webb alignment tool: Development, refinement, and dissemination. In Council of Chief State School Officers (Ed.), *Aligning assessment to guide the learning of all students*. Retrieved from http://www.pacificmetrics.com/white-papers/Aligning_Assessment_to_Guide_the_Learning_of_All_Students.pdf
- Webb, N.L., Alt, M., Ely, R., & Vesperman, B. (2005). *Web alignment tool (WAT): Training manual*. Retrieved from: www.wat.wcerus.org/Training%20Manual%202.1%20Draft%20091205.doc
- Wei, X., Shen, X., Lukoff, B., Ho, A.D., & Haertel, E.H. (2006). *Using test content to address trend discrepancies between NAEP and california state tests*. Paper presented at the annual meeting of the American Educational Research Association, San Fransisco, CA.
- Wells, C.S., Baldwin, S., Hambleton, R.K., Sireci, S.G., Karatonis, A., & Jirka, S. (2009). Evaluating score equity assessment for state NAEP. *Applied Measurement in Education*, 22: 394-408.

Wenglinsky, H. (2004). The link between instructional practice and the racial gap in middle schools. *Research in Middle Level Education Online*, 28(1). Retrieved May 12, 2012, from

<http://www.nmsa.org/Publications/RMLEOnline/tabid/101/Default.aspx>

Wiley, D.E. and Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 california learning assessment system (CLAS). *Educational Evaluation and Policy Analysis* 17(3): 355-370.

Wisconsin Center for Education Research (WCER). (2012). *Surveys of enacted curriculum: State collaborative on assessment and student standards*. Retrieved from seconline.wceruw.org/secwebhome.htm

Ryan Michelle Higgins

4210 Black Oak Court
New Albany, IN 47150

812-989-5474
rmhigg01@louisville.edu

EDUCATION

University of Louisville, College of Education

Expected May 2013

Ph.D. in Curriculum and Instruction

Dissertation title: Examining Alignment: National and Local Assessments and the Common Core State Standards in Mathematics

Committee Chair: E. Todd Brown, Ph.D.

Committee Members: Dr. William S. Bush, Dr. Samuel Stringfield, Dr. Melissa Evans-Andris, and Dr. Dena Dossett

Xavier University

May 2008

M.Ed. Secondary Education

Thesis: Motivating Teens to Motivate Themselves

Adviser: Robert Townsend, Ph.D.

Indiana University Southeast

December 2004

B.S. Elementary Education

Endorsement in Middle School Mathematics

PROFESSIONAL EXPERIENCE

University of Louisville

Louisville, KY

Graduate Assistant

2012 - present

- Conducted periodic observations of classroom science teachers who had attended science professional development seminars in the previous summer.
- Gathered data from the above classroom observations utilizing the Electronic Quality of Inquiry Protocol (EQUIP) and assisted with qualitative coding for formative assessment strategies utilized in the classrooms.
- Collaborated in the coding and development of a formative assessment rubric for use in university classroom to be utilized by instructors in assessing the pictorial representations of pre-service teachers' perceptions of the use of

formative assessment. In addition, developed preliminary drafts for conference proposals based on gathered data.

- Organized the general day-to-day needs of Dr. Melissa Shirley's BESST Practices summer seminars and academic school year professional development seminars, including, but not limited to: assisting in development of exit slips for teachers to respond to the quality of the seminars, assisting in the development of questionnaires for gathering data on teacher knowledge in a variety of subjects, and the gathering of demographic data and organization via Microsoft Excel.

University of Louisville

Louisville, KY

Pre-service Teacher University Supervisor

2012 - present

- University supervisor for MAT students during their methods and student teaching experiences; worked with teacher mentors and pre-service teachers in middle school mathematics placements in a large, urban district.
- Conducted classroom observations during both methods and student teaching, led conferences between mentors and pre-service teachers, gave feedback on weekly reflections, lesson plans, and assessed the lessons given during observations.
- Attended cognitive coaching seminars to support positive conferences and reflective feedback as well as to give support for any situation the pre-service teacher may have had questions about or difficulty with.

McGraw-Hill

Cleveland, OH

Freelance Performance Task Writer

2013 – present

- Writer of tasks for the Time to Know virtual program that include extended response items over two days – items included mathematical practices from Common Core State Standards Mathematics.
- Collaborate with the program director regarding item construction.
- Attend to time frames for project completion.
- Utilize Microsoft PowerPoint to create storyboards, recommendations for animation, narration suggestions, graphics, and appropriate set up for students to complete the task.
- Assign appropriate standards for teacher.

Community Montessori Charter Public School (Pre-K to 12th)

New Albany, IN

7th - 8th Grade Generalist - Focus in Mathematics

2005 - 2012

9th - 12th Grade - Algebra, Geometry, and Algebra II

2007 - 2012

- Selected along with one other teacher to guide the development of the middle and high school program for the school, I developed a Scope and Sequence addressing the integration and interdisciplinary combination of Middle School Indiana State Standards across the curriculum including the Montessori philosophy-based curriculum for 6th through 9th grade.
- Developed multiple research-based parent and staff workshops on adolescent development including: cognitive, physical, social, moral and emotional progress of teens.
- Due to the structure of the Montessori philosophy-based system of educating students within three-year age groupings, I developed multiple years of curriculum across subject matter as a generalist educator similar to that of an elementary educator.
- Developed engaging and interdisciplinary mathematics curriculum to present in a format similar to college education (i.e., delivered two days a week in one hour and fifteen minute sessions).
- Worked as a committee member on the General Education Intervention team to identify strategies to assist teachers in developing methods to address learning needs of struggling children. In addition, the committee determined when special education testing was necessary. Moreover, I was a “go to” person for the special education staff when resources, strategies and brainstorming were needed.
- Worked with a diverse set of learners; in particular, special needs children on many levels. Developed and adapted curriculum that was both engaging and appropriate for each child’s needs.
- Developed both afterschool and summer enrichment opportunities for middle and high school students in the following topic areas: SAT/ACT math assessment preparation, teen entrepreneurship (including grant writing), mathematics and architecture, and forensic science.
- Worked as a liaison for gaining information regarding the Common Core State Standards and PARCC assessment through webinars, conferences meetings, and seminars at the state level and sharing pertinent information with school faculty at meetings and professional development days.
- Developed and trained staff on using the Live Binders program online to create curriculum materials in each core content area. Further, I led discussions on making each of the “binders” accessible, engaging, and interdisciplinary for all learners.

- Engaged teachers in learning to utilize technology in curriculum development and incorporated technology into personal curriculum, including but not limited to: Voice Thread, Prezi, Edmodo, a variety of applets, Movie Maker, and Wiki Spaces.

TECHNOLOGY EXPERIENCE

Edmodo

Voice Thread

Illuminations

Live Binders

Live Text

Black Board

SmartBoard Technology & Clickers

Various Applets

Face Book

Blogger

Apple Office Applications for MAC

Microsoft Office Suite

EQUIP iPad Applications for Classroom Observation

CERTIFICATIONS

Indiana Department of Education

Elementary Teacher's License and Middle School Math Certification - #922442

American Montessori Society

Secondary I and II (ages 12 to 18)

July 2007

PRESENTATIONS

Higgins, R.M. (2013, April). *Socratic seminar in math: Development of math reasoning collaboratively*. A presentation at the National Council of Teachers of Mathematics, Denver, CO. **(Accepted)**

Shirley, M.L. & **Higgins, R.M.** (2012, January). *Development of a scoring guide checklist to assess pre-service science teachers' concepts of formative assessment.* Poster session presented at the Annual Conference of the Association of Science Teacher Educators, Charleston, SC.

Shirley, M.L. & **Higgins, R.M.** (2012, September). *Assessing pre-service science teachers' concepts of formative assessment.* A poster presentation at the Regional Conference of the Mid-Atlantic Association for Science Teacher Educators, Pembroke, VA.

Higgins, R.M. (2011, October). *Authentic assessment in math.* Session presented at the Fall Conference of the Indiana Council of Teachers of Mathematics, Indianapolis, IN.

Higgins, R.M. (2011, October). *Young entrepreneurs = math learning.* Session presented at the Fall Conference of the Indiana Council of Teachers of Mathematics, Indianapolis, IN.

Higgins, R.M. (2011, October). *Creating entrepreneurs and teaching math.* Session presented at the Annual Conference of the Kentucky Council of Teachers of Mathematics (KCTM), Bowling Green, KY.

Higgins, R.M., Bush, S., & Hunter, A.E. (2011, April). *Math in the real world: Geometry on a budget.* Session presented at the Annual Conference of the National Council of Teachers of Mathematics, Indianapolis, IN.

SERVICE

American Education Research Association

Reviewer, 2013 Conference Paper Submissions. (2012).

Reviewer, Award, Analysis of Methodological Advancements. (2011).

Reviewer, 2012 Conference Paper Submissions. (2011).

Indiana Department of Education

Collaboration for development of the Biology ECA – Questar. (2011).

Collaboration for development of 8th grade Math ISTEP - Questar. (2010).

Curriculum alignment and Depth of Knowledge for Math ISTEP. (2010).

National Council of Teachers of Mathematics

Book Review (2012)

Collins, A. (2012). *NCTM assessment resources for the professional learning community: A practical guide.*

OTHER TRAINING

Electronic Quality of Inquiry Protocol (EQUIP) Training

ASTE Annual International Conference

Charleston, SC 2013

PROFESSIONAL ORGANIZATIONS

American Educational Research Association

Association for Mathematics Teacher Educators

The Association for Science Teacher Educators

Greater Louisville Council of Teachers of Mathematics

Indiana Council of Teachers of Mathematics

Kentucky Council of Teachers of Mathematics

National Council of Teachers of Mathematics

HONORS AND ACTIVITIES

National Center for Education Statistics

NAEP Database Training

Summer 2011

Proposal written and selected from 100 candidates; only 30 were taken.

Kappa Delta Pi Honor Society for Educators

Secretary 2005-2006