

Running head: ASSESSMENT PROJECT CAPE PUBLIC SCHOOLS

A STUDY OF THE RELATIONSHIP OF THE DIAGNOSTIC READING
ASSESSMENT - 2
TO THE
MISSOURI ASSESSMENT PROGRAM IN COMMUNICATION ARTS

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
RUTH ANN ORR
Dr. Ruth Ann Roberts, Dissertation Supervisor
DECEMBER 2013

Assessment Project Cape Public Schools

The undersigned, appointed by the dean of the Graduate
School, have examined the dissertation entitled

A STUDY OF THE RELATIONSHIP OF THE
DIAGNOSTIC READING ASSESSMENT - 2 TO THE
MISSOURI ASSESSMENT PROGRAM IN COMMUNICATION ARTS

presented by Ruth Ann Orr,

a candidate for the degree of doctor of education,

and hereby certify that, in their opinion, it is worthy of
acceptance.

Professor Ruth Ann Roberts

Professor Lisa Bertrand

Professor Margaret Dalton

Professor Jan Ward

Professor Paul Watkins

Assessment Project Cape Public Schools

Special thanks to

my husband, Kevin;

my children, Heath, Meghan, Elizabeth and Cory;

my grandson, Aiden;

my parents, Ruth and Robert George;

my colleagues and friends, Sharon Bennett, Betty Chong,

Stephanie Depro, Susie Reinagel, and Julia Unnerstall;

all my prayer warriors, both family and friends;

my daddy, Bud Brockmeier, who is always so close to my

heart;

and my Heavenly Father who sustained me throughout this

entire process.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ruth Ann Roberts for all her knowledge, help and support throughout my doctoral program. Dr. Jerry Waddle and Dr. Paul Watkins have also been a wealth of information and support while I was completing my coursework. I would also like to thank Dr. Jeremy Heider for all his assistance in helping me with the statistical analysis portion of this paper. Finally, I would like to thank my colleagues in Cohort 5, especially Rhonda Dunham and Sydney Herbst, as they contributed significantly to my learning and success.

TABLE OF CONTENTS

Acknowledgements.....ii

Chapter

1. INTRODUCTION TO THE STUDY.....1

Background
Statement of the Problem
Purpose of the Study
Potential implications for the Study
Hypothesis and Research Questions
Limitations
Assumptions
Definition of Key Terms

2. REVIEW OF RELATED LITERATURE.....17

Introduction
Missouri's Assessment History
Application of Summative Assessments
Application of Formative Assessments
Summary

3. RESEARCH AND METHODOLOGY.....39

Introduction
Problem and Purposes Overview
Research Hypotheses
Population and Sample
Data Collection and Instrumentation

4. ANALYSIS OF DATA.....55

Introduction
Organization of Data Analysis
Results on Correlation Analyses
Summary

5. DISCUSSION.....62

Summary of the Study
Findings
Conclusions
Implications
Future Research

Summary

6. REFERENCES.....	89
7. VITA.....	95

Chapter 1

INTRODUCTION TO THE STUDY

Background

Historically, the federal government of the United States gave the state governments power to oversee educational policy thus ensuring education would not be controlled by a select few (Marshall & Gerstl-Pepin, 2005). Following the publication of *A Nation at Risk* in 1983, the federal government began a push toward accountability of the education of America's youth. Efforts by both the George H.W. Bush and the Clinton administrations relied on voluntary cooperation from state and local officials (Peerson & West, 2003). However, beginning in 1994, the revised Elementary and Secondary Education Act (ESEA), renamed the Improving America's Schools Act (IASA) tied Title I funding to each state's development of content and performance standards (Peerson & West, 2003). It was during this reauthorization, the concept of adequate yearly progress began surfacing although the IASA set no deadlines for states or local schools to attain proficiency.

In 1999, President George W. Bush used education as a major issue for his initial term in office. While governor in Texas, Bush continued the push of his predecessors and annualized testing in the subjects of reading and math in

grades three through eight. Texas required that prospective graduates pass the Texas Assessment of Academic Skills (TAAS) and tied teachers' and administrators' effectiveness and their continued employment to student performance. With President Bush's urging, the ESEA was again reauthorized in 2001 and renamed the No Child Left Behind act (NCLB). Among the provisions of NCLB was a call for accountability. Accountability was defined by Congress as measured student performance on a standardized basis for the purpose of determination of adequate yearly progress (AYP) among schools with Title I funding tied to a district's ability to meet AYP (Peerson & West, 2003).

Therefore, while federal educational policy provides the directives for the operation of school districts through mandates such as No Child Left Behind (NCLB) and, beginning in 2009, Race To the Top (RTT) (U.S. Department of Education, 2009), states are given the latitude to interpret and implement federal guidelines. For example, while all states receiving federal monies are required to administer some type of summative assessments to students in specific grade levels in specific subject areas including communication arts (now called English Language Arts or ELA), mathematics, science and social studies, states have the latitude to expand testing into other grade

levels and other subjects. States are also given the latitude to develop their own state assessment. The assessment program the State of Missouri developed is known as the Missouri Assessment Program (MAP).

At the local level, states develop suggested curricular standards currently known as Grade Level Expectations (GLEs) as defined by the State of Missouri. However, local school districts have the latitude to revise, amend or expand the standards to develop a curriculum unique to that particular school district. Currently, other than the state-mandated summative assessments administered at the end of a school year, states also have the latitude to choose and administer their own types of common assessments to periodically assess benchmark thresholds. Depending upon the district, these common assessments may also be used as formative assessments with the purpose of guiding or changing instruction for a student or a group of students based upon results and analysis of the common assessments results.

While states had the latitude to select the format of the assessment, NCLB mandated that students in grades three through eight be tested each year in reading and math. Additionally, students were to be tested again during one year in high school in the subjects of reading and math.

Each state's tests were to report not only the total aggregate results but also disaggregate scores in the categories of race, ethnicity, economic status, and disability. States were also required to participate in the National Assessment of Educational Progress (NAEP) in two grades every other year.

Because the MAP was only given in grades 3-11 in the Spring with the primary function being a summative test of all content standards that should have been mastered by students in a particular grade level, the Cape Girardeau No. 63 District (also more commonly known as the Cape Girardeau Public School District) recognized the need for more frequent assessments. Beginning in 2002 the district initiated administration of the Developmental Reading Assessment (DRA) and Rigby Running Reading Records (a similar, but more informal, assessment of reading level) in grades K-8 to assist teachers in placing students into guided reading groups and to track reading progress across the year as part of the district's Balanced Literacy program. In the elementary grades (grades K-6), an assessment wall was maintained to allow reading coaches and teachers quick visual access to the progress of their students in the area of reading as measured by the DRA in the Fall and the Spring and Rigby Running Reading Records

administered as a midyear assessment. This wall listed all levels which could be assessed by the DRA and the Rigby Running Reading Records. Each grade level was assigned a color, and each student's name was written on a colored square of paper corresponding to his or her assigned grade. Squares were then placed under the level obtained by the student as measured by the most recent reading assessment administered.

In 2005, the district switched to the revised edition of the DRA, the DRA2, which continued to measure reading recognition and comprehension but added a fluency component across levels and a written response component beginning at level 30 which corresponds roughly to a third grade reading level.

Beginning in 2009, the district also put into place additional periodic assessments, referred to by the district as common formative assessments, so called because these assessments were common to all schools across designated grade levels and were to be used to guide instruction to improve student learning and, as a result, improve performance on the state summative assessments as, theoretically, all assessments were directly related to the GLEs. In the area of reading, the common assessments continued to be the DRA2 and Rigby Running Reading Records.

In addition to the standardized reading formative assessment, the district self-developed a common formative writing assessment (DWA) and a common formative mathematics assessments(DMA). Unlike the DRA2, neither assessment was a formal, standardized assessment. Writing prompts for grade levels and mathematic items were pulled from grade level materials, which were then either scored against a rubric in the case of the DWA or given a percentage score in the case of the DMA. These assessments were administered in grades K-8 beginning in 2006 and 2007 respectively. Overall results from the DRA, DWA, and DMA were reported to the Literacy Coaches and/or the building level principals who then analyzed the performance of students both as a total group and as disaggregated groups based upon race, free and reduced lunch status, and receipt of special services as indicated by having an Individual Education Plan (IEP). Results were periodically reported to the Cape Girardeau Board of Education by either the Assistant Superintendent for Curriculum and Instruction or the district Curriculum Coordinator. During the report, it was stressed to the Board of Education that district common formative assessments had several purposes:

1. Assessments provided "snapshots" of individual students and groups of students for that moment in

time to allow for adjustment in instruction to a class or intervention in instruction to an individual.

2. Assessments were "MAP-like" in design allowing students to experience items designed to be reflective of constructed response items and performance events as well as multiple-choice items.
3. Because assessments were intended to be "MAP-like" and also to reflect the GLEs taught for mastery at that grade level, such assessments were thought by the literacy specialists and district administration to have some correlation to performance on the MAP so that student results on the common assessments could be used to predict student success on the MAP.

Statement of the Problem

While NCLB mandated an end-of-year, summative assessment, many schools began to recognize that high-stakes testing, such as the MAP, did not provide timely or detailed information about an individual student or their particular class. Therefore, they supplemented state testing with benchmark testing (Heritage, 2007). Much research exists indicating the value of the use of formative assessments and/or benchmark assessments to give teachers information about student performance and assist

in making instructional decisions (Black & William, 1998; Stiggins & Chappuis, 2005; Brookhart, 2008; Clark, 2010; Dorn, 2011).

However, while the MAP has been mandated by the state of Missouri since 1998, and district common assessments have been administered by the Cape Girardeau Public School District for the purpose of formative assessment since 2005, no formal research has been conducted by the district to determine what correlation, if any, exists between performance on any of the district assessments and performance on the MAP. This is not unusual at the state level as little peer-reviewed research exists on the MAP itself. S.D. Whitney, Ph.D, Associate Professor of Educational Psychology with the University of Missouri, indicated that peer-reviewed research on specific state tests are rare (personal communication, December 10, 2012).

While the district administers formative assessments in both reading and mathematics, there is a difference in the two types of formative assessments used. The formative assessment for reading, The Developmental Reading Assessment - Second Edition (DRA2) is a commercial, standardized test which has been formally developed, standardized and normed by Pearson Education Incorporated. The assessments for mathematics, while formative in nature,

were an amalgamation of test items gathered by the district's teachers. Additionally, no item error studies were conducted for the formative mathematic assessments. Item variance was more controlled with the standardization of the DRA2. Therefore, for the purposes of this study, only the communication arts portion was study used since formative mathematic assessments were not constructed with the same rigor as the formal and standardized DRA2.

Purpose of the Study

The purpose of this study is to predict a student's performance on the MAP based upon that student's performance on the DRA2. Additionally, there was a desire of this researcher to focus on the psychometric practices of this district as it relates to both formative and summative assessments. For many years, the district has added common formative assessments first in reading, then mathematics, then in writing. Reading has consistently been assessed with a standardized test (the DRA, then, as the DRA was revised, the DRA2).

Finally, as Race To the Top (RTT) implies, teacher performance may be judged, not on the final score of the summative evaluation, but on the ability of a teacher's students to show progress over time on a given assessment.

Potential implications for the study

Comparison of the DRA2 to the MAP has potential implications across school districts in Missouri. The DRA2 is a frequently used assessment in determining students' reading levels and abilities. Determining if there is a correlation between performance on DRA2 and the Communications Arts portion of the MAP could possibly impact not only the Cape Girardeau Public School District's decision to continue use of the DRA2 but could also influence other schools' decisions across the state. Examining the extent to how the information gleaned from the DRA2 is used to form or change teachers' practices could influence future practices of this district as well as shape practices of other districts that would choose this same instrument.

Hypothesis and Research Questions

The null hypotheses of this research were:

H01^a There are no significant correlational relationships between third grade MAP communication arts scores and the third grade DRA2 scores.

H01^b: There are no significant correlational relationships between fourth grade MAP communication arts scores and the fourth grade DRA2 scores.

H01^c: There are no significant correlational relationships when comparing third grade DRA2 scores to fourth grade MAP communication arts scores.

Research questions answered in this research were:

1. How does the third grade DRA2 predict performance on the MAP communication arts portion?
2. How does the fourth grade DRA2 in predict performance on the MAP communication arts portion?
3. How does the third grade DRA2 predict performance on the MAP communication arts portion in fourth grade?

Limitations

Limitations of this study include generalizing results based upon the results of only one school district among many in the state of Missouri. A second limitation is that while the greater majority of the public schools in the state of Missouri use the MAP as a summative assessment as mandated by law, not all school districts use the DRA2 in the same manner as the school district under study. While this school district uses the DRA2 as one of its formative assessment, a different formative assessment could yield different results. As a primary example, the DRA2 is a standardized instrument the Cape Girardeau Public School district uses as a formative assessment. There are multiple types of formative assessments which could yield very

different results than the formative instrument being used in this study.

The researcher is an elementary principal with some knowledge of psychometric practice. However, personal experience was more with individualized psychometric assessments rather than group assessments.

Assumptions

One key assumption to this study was the Missouri Assessment Program had not been comparatively normed with nor correlated with the Diagnostic Reading Analysis. As stated previously, personal correspondence with personnel from the Department of Elementary and Secondary Education and from researchers from the University of Missouri who used the MAP within their own research indicated little if any correlational analysis had been done with this particular assessment. Going further, one researcher indicated it was not typical practice to conduct correlational analyses between state assessments and other normed achievement assessments.

A review of literature found no instances of any correlational studies involving the MAP. Therefore, it is assumed no such studies exist. This point is extremely important not only when determining the purpose for this

study but also as discussions of implications for further study are discussed later in this research.

Definition of Key Terms

For the purposes of this study, definitions of key terms will include the following:

Assessment. "...[Assessment] refer[s] to all those activities undertaken by teachers...that provide information to be used as feedback to modify teaching and learning activities." (Black, P. & William, D., 1998).

Formative assessment. Formative assessment takes place during instruction and is understood by the Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards (FAST SCASS) as being a process of assessment-based feedback to help teachers and students make adjustments that will improve student achievement (Popham, 2008). Shepard, Hammerness, Darling-Hammond and Fust offer a more succinct definition of formative assessment stating that formative assessment is assessment conducted in the flow of instructional process.

Summative assessment. Assessments given at the end of a unit or time period including at the end of an academic school year. These assessments are intentionally given after all instruction and student learning have ended with

the intent to support levels of proficiency. (Ainsworth & Viegut, 2006).

Benchmark tests. Also known as an interim test. These types of tests are administered periodically to assess mastery of important curricular outcomes. (Popham, 2008).

Summary

The Cape Girardeau Public School District, like schools around the nation, has participated in several different testing initiatives to comply with the summative testing required by the federal government as a result of requirements of the Elementary and Secondary Education Act (ESEA) and subsequent reauthorizations. The state of Missouri has developed the summative tests used by the district. The current test, the Missouri Assessment Program, meets the requirement of the most recent reauthorization of the ESEA, the No Child Left Behind act (NCLB).

The school district, in an effort to garner information about students in the area of communication arts and mathematics, implemented a series of common formative assessments. Common formative assessments were used by the district to gather more timely and detailed information about students and, it was believed, to

supplement or change instruction in order to positively influence scores on the summative evaluation, the MAP.

For communication arts, progress monitoring was done first with the Developmental Reading Assessment (DRA) and then with its newest revision, the Developmental Reading Assessment Second Edition (DRA2). The DRA and its revision, the DRA2, is a formal and standardized test of reading. Mathematics and writing were also assessed with common formative assessments. However, both subjects were tested with non-standardized tests developed within the district. Because this researcher wanted to control for item error variance, this study focuses on the only formal common assessment administered by the district, the DRA2, and compared it to the MAP to determine if any correlation exists between the two tests.

This purpose of this study was to determine if there was a correlation between the DRA2 (the common formative assessment) and the MAP (the large scale summative evaluation). Because the DRA2 is used not just locally but also throughout the state of Missouri, determining if there is a correlation between the two tests is significant to many school districts across the state. Also of significance is the fact that correlational analysis of common formative assessments and end-of-the-year summative

assessments is not commonly performed even though the practice of administering common formative assessments to guide instruction with the end result of positively influencing performance on summative assessments is routine throughout the state and the nation (Ainsworth & Viegut, 2006; Halverson, 2010; Heritage, 2007; Stiggins & DuFour, 2009). Instructional time and instructional methods could be affected by the results of this study.

Chapter 2

REVIEW OF RELATED LITERATURE

Introduction

The use of summative assessments is not unique to Missouri. End-of-the-year summative assessments with the purpose of determining adequate yearly progress (AYP) has been mandated by federal law since 2001 (Peerson & West, 2003). While Missouri has implemented end-of-the-year summative assessments since 1978 beginning with the Basic Essential Skills Test (BEST), transitioning to the Missouri Mastery and Achievement Test in 1987, and most currently administering the Missouri Assessment Program (MAP), there have been no mandates, state or federal, for more periodic assessments that would inform classroom instruction in a timely manner.

Researchers have noted that end-of-the-year assessments, while frequently analyzed for specific student performance, are difficult to use to inform classroom instruction as results at the end of the year are analyzed and projected on an entirely different group of students who may or may not benefit by the adjustments of instruction such analysis provides (Ainsworth & Viegut, 2006; Condermand & Hedint, 2012; Popham, 2008). Many researchers note that formative assessment is more useful

in making decisions that will positively impact student achievement (Ainsworth & Viegut, 2006; Black & William, 1998; Chappuis & Chappuis, 2006; Chappuis, Commodore & Stiggins, 2010; Marzano, 2006; Moss & Brookhart, 2012; Popham, 2008; Reeves, 2007).

Missouri's Assessment History

Missouri, like most states, developed its own state assessment in response to the reauthorizations of the ESEA. The Basic Essential Skills Test (BEST) was one of the first state-developed tests by Missouri and was given in the spring of each year beginning in 1978 (Freeman & Hatley, 1981). The Department of Elementary and Secondary Education of the state of Missouri (DESE) began to develop a different and more formalized assessment of skills beginning in 1985. The Missouri Mastery and Achievement Test (MMAT) was adopted in 1987 to comply with state mandates of the Excellence in Education Act of 1985. According to the *Missouri Mastery and Achievement Test Technical Manual*, "The MMAT battery...[was] devised to encourage and promote quality in Missouri schools" (Osterland & Merz, 1987, p. 6). Osterland & Merz reported that beginning on February 27, 1985, DESE formed four subcommittees made up of DESE personnel, teachers, and administrators and began the process of identifying Core

Competencies and Key Skills in the areas of communication arts (language arts/reading/English), mathematics, science, and social studies/civics for grades 2 through 10. These four committees were overseen by The Core Competencies and Key Skills Oversight Committee that was composed of educators, business and industry leaders, and state officials. The final Core Competencies and Key Skills identified by the four subcommittees and finalized by the Oversight Committee were made available to schools and the general public following publication of the *Core Competencies and Key Skills for Missouri Schools* in October, 1986.

In November, 1985 DESE contracted with the Center for Educational Assessment (CEA), a division of the University of Missouri-Columbia College of Education, to develop a criterion-referenced test based upon the Key Skills (Osterlind & Merz, 1987). Osterlind & Merz (1987) state, "The final version of the MMAT consist[ed] of thirty-four criterion-referenced tests, which assess[ed] student performance on Key Skills in grades 2 through 10." At that time, students in Grade 2 were assessed annually in the spring in the subject areas of language arts/reading and mathematics. Grades 3 through 10 were assessed annually in

the spring in the subject areas of reading/English/language arts, mathematics, science, and social studies/civics.

Development of the Missouri Assessment Program (MAP)

In 1993, the Missouri Senate passed Senate Bill 380, the *Outstanding Schools Act of 1993*. The State Board of Education directed DESE "to identify the knowledge, skills, and competencies that Missouri students should acquire by the time they complete high school and to assess student progress toward those academic standards" (Missouri Department of Elementary and Secondary Education, June, 1998, p. 4). DESE again worked with teachers, school administrators, parents, and business professionals to develop The Show-Me Standards. According to the Missouri Department of Elementary and Secondary Education, these standards departed from the Key Skills developed in 1986 by "requiring the student to not only...acquire knowledge but to be able to communicate and to apply their knowledge in a variety of settings" (Missouri Department of Elementary and Secondary Education, June, 1998, p. 4).

To assess the Show-Me Standards, DESE designed the Missouri Assessment Program (MAP). According to *The Assessment Standards for Missouri Public Schools*, the MAP was initially designed to assess mathematics, science, communication arts, and social studies at a statewide level

once at each of three grade levels, elementary, middle, and high school, and was to assess health/physical education in grades 5 and 9. Plans were to also phase in a Fine Arts assessment in the elementary, middle and high school levels. An alternate assessment was also to be developed for a selected group of special education students whose Individual Education Programs (IEPs) determined the MAP would be an inappropriate assessment of their skills (Missouri Department of Elementary and Secondary Education, June, 1998). This MAP was to consist of multiple-choice items, constructed-response items, and performance events. The TerraNova, a nationally norm-referenced test developed by CTB/McGraw Hill, was chosen as multiple-choice component of the MAP. Constructed-response and performance events items were developed by groups of professionals in the state in conjunction with CTB/McGraw Hill. The constructed-response items were designed not only to measure content knowledge but also to provide insight into how students derived their answers. Performance events, according to *The Assessment Standards for Missouri Public Schools* "...require[d] students to work through a complex problem or present a written argument.... Performance events generally allow for more than one approach to arriving at a

correct solution" (Missouri Department of Elementary and Secondary Education, June, 1998, p. 5).

The MAP became available in Spring, 1997, and was required in Spring, 1998. No assessments were required in grades K-2 or 12. The science and communication arts MAP assessments were required in grade 3. Mathematics and social studies MAP assessments were required in grade 4. The health/physical education MAP assessment was required in grades 5 and 9. There was no assessment required in grade 6. Science and communication arts MAP assessments were required in grade 7. Mathematics and social studies MAP assessments were required in grade 8. Mathematics and science MAP assessments were required in grade 10. Communication arts and social studies MAP assessments were required in grade 11.

Changes in the MAP have occurred over time. Currently, the health/physical education assessments are dropped completely. Social studies is not assessed in grades 3 through 8. Students in grades 3 through 8 are now assessed annually in communication arts and mathematics. Science is assessed in grades 5 and 8. Beginning in Fall, 2008, End-of-Course (EOCs) Assessments replaced the MAP in grades 9-11 and assessed the specific subjects of English I and II, Algebra I and II, Geometry, Biology, American History, and

Government (Missouri Department of Elementary and Secondary Education, 2012; Missouri Department of Elementary and Secondary Education, 2011.)

Another change in MAP was reflected in the initial reporting of scores. MAP was designed to not only report in standard scores but also in descriptive measures. The initial MAP descriptors fell into five categories: Step 1, Progressing, Nearing Proficiency, Proficient, and Advanced. In 2006, MAP descriptors changed into four categories: Below Basic, Basic, Proficient, and Advanced. The format of the MAP content stayed consistent in regard to the types of items contained within the test (multiple-choice/TerraNova items; constructed response items; and performance events items) up until 2009.

The method of scoring the MAP also stayed consistent from its inception until 2009 with some tests scored by CTB/McGraw Hill while other portions of the test (constructed response items and performance events items) being scored at assessment sites around the state by trained assessors consisting mainly of teachers from around the state who were trained using a rubric to hand-score these items. In 2009, this scoring method changed with all MAP tests being scored by CTB/McGraw Hill.

In summary, Missouri, like most states in the nation, developed a statewide, summative assessment system that was in keeping with federal law. The thinking by first the Bush Administration, then the Obama administration, was that testing was and is a way to monitor student performance while also serving as a vehicle of change of what and how tested subject matter is taught (Madaus & Russell, 2010/2011). The Missouri Department of Elementary and Secondary Education was quoted in a 1998 publication titled *Assessment Standards for Missouri Public Schools* that administrative use of the MAP would include,

...inform[ing] parents, the school board and the public about the effectiveness of schooling and ...revealing trends that will aid in setting priorities, planning the curriculum, revising instructional practices, and evaluating the progress of educational programs (Missouri Department of Elementary and Secondary Education, June, 1998, p. 2).

This same publication indicated instructional uses for teachers included, "...identifying strengths and weaknesses in classroom curriculum instruction" and "assisting[ing] teachers in monitoring individual student learning" (p.2). Missouri was and is not the only state governed by the

policies set forth by NCLB or RTT. As a result of the mindset developed by these policies, Heritage (2007) states many teachers equate assessment with end-of-the-year, high stakes testing. While the original intent of this type of end-of-year assessment appeared to be improving student achievement by obtaining information about an individual student's learning or a group of students' learning, Heritage goes on to note that in the current accountability environment, assessment is not regarded as a means of obtaining student information. It is, instead, seen as a way of categorizing students and schools.

Missouri is not alone in a lack of research or application as Dorn (2011) indicates that on a national level, little research exists on the use of structured formative-assessment implementation. However, Ainsworth and Viegut (2006) state common formative assessments should be collaboratively designed, periodically administered, and collaboratively scored to allow teachers to inform instruction. These authors suggest that if common formative assessments are aligned to the large scale assessments, they should then offer predictive value to the summative evaluation administered.

Application of Summative Assessments

Paul Black and Dylan Wiliam (1998) in writing their seminal article *Inside the Black Box: Raising Standards Through Classroom Assessment*, noted that "...[R]aising the standards of learning that are achieved through schooling is an important national priority" (p.1). They cited a specific assessment that was available at the time, the National Assessment of Educational Progress (NAEP), as an initiative developed to improve school planning and management. However, the NAEP, like many of the state assessments developed after the writing of this article, is given at the end of the year and is meant to assess the sum of the learning that has taken place over the course of a school year with the aim of judging individual, building and district proficiency levels. Popham (2008) notes that many, if not most, state accountability tests that were being used at the time are often amalgamations of student performance that made it hard to judge specific student performance since 1) student performance is "lumped" at a strand level, and 2) results occur *after* students leave a grade level with the end outcome of teachers being unable to adequately adjust instruction. Ainsworth and Viegut (2006) support this observation, noting that large-scale assessments are minimally useful to affect change in an

individual student's learning progress as the turn-around time is too long to provide specific and immediate feedback to the student. Condermand and Hedint (2012) assert that when assessing content knowledge, "Waiting to conduct assessments until after an instructional period misses opportunities for...teachers to reflect critically about their instruction...and students to adjust their thinking...processes...and have multiple opportunities to improve and demonstrate their learning" (p. 162).

Popham (2001) notes that large-scale assessments, when used alone, have minimal impact on an individual child's academic growth. Reeves (as quoted in Ainsworth & Viegut, 2006) writes in agreement noting,

The nation *is* over-tested, but we are under-assessed. The distinction is essential, as many schools continue to engage in summative testing—educational autopsies that seek to explain how the patient died but offer no insight into how to help the patient improve (p.ix).

Stiggins and DuFour (2009) concurred with Popham (2008) and Ainsworth and Viegut (2006) regarding time lapses between the assessments given and the results stating, "For effective communication, both teachers and students must

learn the results of assessments as early as reasonable”
(p. 641).

Ainsworth and Viegut (2006) observed that when educators are asked about the data about student achievement most currently used by educators, “[T]he most frequent answer is the data from large-scale assessments, even though the usefulness of that data for a classroom teacher is limited” (p.2). Madaus and Russell (2010/2011) note that high-stakes tests like the MAP typically confirm the achievement levels observed by teachers as students go into the test. What these tests do not do is to provide information about why students are struggling. In fact, Popham (2008) notes that “...there is no evidence that [either] district-developed or state-developed assessments boost student achievement” (p.10).

Therefore, while researchers currently acknowledge the legal mandates to administer summative evaluations for state accountability testing, these same researchers also recognize significant limitations of summative evaluations alone as a means of influencing student achievement. In summary, the intent of large-scale, summative assessments was to provide teachers with feedback about students’ mastery of achievement standards set forth by the district. In Missouri, the Grade Level Expectations (GLEs) were

assessed at the end of each academic year using the MAP. A number of reports were generated that provided information about each student's performance along with reports, which consolidated individual student reports into performance on each GLE or groupings of GLEs at the classroom, school, and district levels. These reports were provided to districts in August/September at the beginning of the next academic year. Therefore, the limitations noted by Popham (2008), Ainsworth and Viegut (2006), and Condermant and Hident (2012) were true for Missouri. Teachers did not get information in a timely fashion in order to make instructional adjustments that would impact the current group of students with whom they were working. While teachers did indeed receive a great deal of data about students, such data was received far after that student or group of students was under the influence of a particular teacher. Data was simply not available in real time to make real-time decisions to restructure teaching as researches would suggest (Ainsworth and Veigut, 2006).

Application of Formative Assessments

Many researchers in the field point to a different type of assessment, formative assessment, as being far more useful in making timely instructional decisions that impact student student achievement (Ainsworth & Viegut, 2006;

Black & William (1998); Chappuis & Chappuis, 2006; Chappuis, Commodore & Stiggins, 2010; Marzano, 2006; Moss & Brookhart, 2012; Popham, 2008; Reeves, 2007). Such assessment takes place with frequency and results are available to both teachers and students immediately. Teachers review data collected from formative assessments frequently. In fact, research indicates that schools with the largest achievement gains review student performance, teaching strategies, and leadership practices at least monthly. They also identify practices they have stopped doing as a result of frequent monitoring of the effectiveness of their programs or practices (Popham, 2008).

Many educators and test developers argue that summative assessments can also be used as formative assessments as results from end-of-the-instructional period assessment can, with analysis, be used to guide instructional practices. However, practitioners in the field note limitations when using summative assessments in a formative manner. Dirkson (2011) agrees that while summative assessments can be formative, limitations in timing and opportunity to revisit information frequently preclude summative assessments from this use.

Sherman Dorn observed in a 2011 article in *Exceptional Children* that NCLB should have achieved the result of expanding formative assessment as many districts either encourage or require tests that mimic or parallel the state summative tests several times throughout the school year (Dorn, 2011). Stiggins and Chappuis (2005) describe such assessments by terms such as common assessments, benchmark assessments, end-of-course examinations, and quarterly or monthly formative standardized assessments.

These types of assessments, especially those commonly recognized as benchmark assessments have come under criticism. Most benchmark systems attempt to provide appropriate data and avoid the lags involved in standardized testing (especially summative assessments) (Halverson, 2010). However, benchmark testing still is not usually judged as an adequate assessment system as they are typically administered only three to four times per year (Heritage, 2007). According to James Popham, "...[T]here is currently no research evidence supporting the hypothesis that [commercial benchmark testing] is educationally beneficial" (as quoted in Ainsworth & Viegut, 2006, p.10). Additionally, many teachers are fearful of benchmark testing (especially computerized benchmark testing) as they see administrators having such frequent, at-hand data, that

their effectiveness as teachers could be measured negatively without the ability to describe individual classroom characteristics (Halverson, 2010). Therefore, many teachers perceive that in using benchmark assessments it is they, not the students, who are being assessed when results of benchmark tests are reported to their building-level or district-level administrators.

Current research emphasizes that the most effective type of assessment impacting student performance on a regular basis in a classroom is formative assessment (Ainsworth & Viegut, 2006; Chappuis & Chappuis, 2006; Chappuis & Stiggins, 2010; Popham, 2008; Reeves, 2007; Stiggins & Chappuis, 2005; Stiggins & DuFour, 2009). Formative assessment is *not* (emphasis added) synonymous with benchmark testing (Popham, 2008). Referring back to Black and William (1998) while assessment would refer to all activities by teachers that provide information to be used as feedback, feedback becomes *formative assessment* "...when the evidence is actually used to adapt the teaching to meet student needs" (p. 140). Citing the research of Black and William, Popham found evidence that formative assessment raises standards concluding that formative assessment was an intervention leading to one of the largest gains in student learning ever reported (Popham, 1998). Well-

designed formative assessment, according to Dorn (2011), is extremely effective in assisting low achieving students in closing the achievement gap which was a primary purpose of the federally mandated testing required by NCLB.

Popham (2008) later spoke in an even more direct fashion about educators regarding formative assessment stating,

If you asked [educators] to explain [formative assessment}, they might tell you it involves testing students in the midst of an ongoing instructional sequence and then using the test results to improve instruction. By and large, this explanation is correct. But a "by and large correct" explanation just isn't good enough when it comes to formative assessment.

. . .[F]ormative assessment is a potentially transformative instruction tool that, if clearly understood and adroitly employed, can benefit both educators and their students (p.3).

Stiggins and DuFour (2009) also noted the need for expanded use of formative assessment in their article, *Maximizing the Power of Formative Assessments*. In discussing summative assessments, Stiggins and Dufour stated that such assessments were more widely used to make

changes at the institutional or policy-creation level. The authors noted that school-level assessments were also valuable whether they be summative or formative; however, these types of assessments gave information that best served the staff about overall classroom or schoolwide performance. Because of their work with Professional Learning Communities (PLCs), they noted that school-level assessments are used to identify curricular areas that are weak because many students are struggling to master curricular content. School-level assessments are also used to help teachers identify areas that he or she might need to strengthen within their own pedagogical skills. To affect change at the student level, Stiggins and DuFour recommended formative classroom assessments especially if the purpose is to identify students who might require more intensive interventions to master curricular content.

Ainsworth and Viegut (2006) make the analogy of going from a "snapshot" of student performance as assessed by a summative evaluation to a "photo album" when using a formative evaluation (p.2). Because formative assessment may come from various outputs, the emphasis is on the process of adapting instruction to meet students' needs and is dependent upon high-quality interactions between the teacher and the student (Clark, 2010). Popham (2008)

emphasizes that it is imperative for educators to understand that formative assessment is a process, not a test. Heritage (2007) also emphasizes that formative assessments are "...a systematic process to continuously gather evidence about learning" (p.141). Therefore, researchers emphasize the act of formative assessment is deliberate in focus and should be "...a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics" (Popham, 2008, p.6). In other words, formative assessments cannot be classified as "a test" because it is an integral part of a process of adjusting what teachers or students may be doing in a classroom (Popham, 2008).

Because formative assessment is a process rather than a time-bound assessment, formative assessments can be used before, during and after instruction. Formative assessments are usually informal (Conderman & Hedin, 2012; Bakula, 2010), clarify what students are supposed to be learning, improve instructional practices of teachers and allow for reteaching of concepts (Bakula, 2010; Clark, 2010). Because of the frequency and the informal nature of formative assessments, some researchers believe formative assessment can lose its meaning as a term as it can mean anything from

informal, haphazard judgment to frequent quantitative assessment of specific skills. When formative assessment is the former, such poorly designed formative assessments may negatively impact student performance (Dorn, 2011). When formative assessment is the latter, research is favorable that gaps in learning progressions can be identified in a timely manner that will impact instructional delivery in a positive direction and, as a result, will improve student achievement (Roskos & Neuman, 2012).

Researchers of assessments and experts in the area of assessments focus much attention on the specificity of what constitutes good formative assessment practices (Ainsworth & Viegut, 2006; Bakula, 2010; Brookhart, 2008; Chappuis & Chappuis, 2006; Clark, 2010; Conderman & Hedin, 2012; Dirkson, 2011; Dorn, 2010; Halverson, 2010; Heritage, 2007; Marzano, 2006; Popham, 2008; Reeves, 2007; Roskos & Newman, 2012; Stiggins & DuFour, 2009). Reeves (as quoted in Ainsworth & Viegut, 2006) states that, "Research over the past decade has presented an overwhelming case in favor of providing feedback to students that is frequent, specific, and accurate" (p. ix). This type of feedback comes from the administration of frequent formative assessment. Again, however, not all formative assessment is productive. Popham (2008) stresses it is imperative for educators to

understand that formative assessment is a process not just a test. Good formative assessment is based upon ongoing instruction and immediately correcting or adjusting the instruction. Experts including Marzano, Stiggins, Black, William, Popham and Reeves agree that frequent formative assessments that are short in nature and administered with frequency provide a "...better indication of a student's learning than one or two large assessments given in the middle and at the end of the grading period" (Ainsworth & Viegut, 2006, p. 27).

The purpose of assessment formative assessments helps set up a student-teacher partnership to either assist the student to become aware what he or she is expected to learn at the outset, to monitor progress, or to analyze his or her own learning (Stiggins & Chappuis, 2005). Roskos & Neuman (2012) refer to this as a "feedback loop" (p. 534). The authors stated that for such a feedback loop to be productive the student and the teacher have to have a clear concept of the learning outcome, compare what is happening in real time, and engage in intentional action to narrow the gap between the student's current and desired performance. Reeves is very specific in writing about such formative assessment stating that, "...[F]ormative assessment without formative analysis and instructional impact is not

formative assessment" (as quoted in Ainsworth & Viegut, 2006, p.x).

Summary

With the emphasis on high-stakes summative assessments, the use of formative assessments is limited, especially when the discussion turns to using formative assessments with fidelity and rigor (Heritage, 2007; Reeves, 2007). To effectively use formative assessment, teachers must know the concepts and skills students are to be taught throughout the year; must have the pedagogical knowledge that is appropriate to the student and to the purpose; and must have means to a measure of the students' previous knowledge that is current. Teachers must also know about the range of formative assessment strategies, how to align the assessment to the goals they are wanting to reach, and make quality inferences drawn from the assessments administered (Heritage, 2007). It is that type of analysis that should inform classroom instruction in a timely manner that will positively impact student learning. If formative assessments test skills students are learning throughout the entire school year, student performance should increase on end-of-the-year summative assessments.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

Introduction

Following the revision of the Elementary and Secondary Education Act (ESEA), renamed the No Child Left Behind Act of 2001 (NLCB), the U.S. Department of Education required that students in grades three through eight be tested each year in reading and math to assess the adequate yearly progress (AYP) of schools receiving federal funds for youth identified as disadvantaged under federal guidelines. Additionally, students were to be tested again during one year in high school in the subjects of reading and math. Each state's tests were to report not only the total aggregate results but also disaggregate scores in the categories of race, ethnicity, economic status, and disability. States were also required to participate in the National Assessment of Educational Progress (NAEP) in two grades every other year (National Assessment Governing Board, 2002).

In 1993, the Missouri Senate passed Senate Bill 380, the *Outstanding Schools Act of 1993*. The State Board of Education directed the Missouri Department of Elementary and Secondary Education (DESE) "...to identify the knowledge, skills, and competencies that Missouri students

should acquire by the time they complete high school and to assess student progress toward those academic standards"

(Missouri Department of Elementary and Secondary Education, June, 1998, p. 4). DESE worked with teachers, school administrators, parents and business professionals to develop The Show-Me Standards. According to DESE, these standards departed from a previous set of standards developed in 1986 and known as the Key Skills by "...requiring students not only to acquire knowledge but to be able to communicate and to apply their knowledge in a variety of settings" (Missouri Department of Elementary and Secondary Education, June, 1998, p. 4).

To assess the Show-Me Standards, DESE designed the Missouri Assessment Program. According to *The Assessment Standards for Missouri Public Schools*, the MAP was initially designed to assess mathematics, science, communication arts, and social studies at a statewide level once at each of three grade levels: elementary, middle, and high school and was to assess health/physical education in grades 5 and 9. Plans were also to phase in a Fine Arts assessment in the elementary, middle and high school levels. An alternate assessments was also to be developed for a selected group of special education students whose Individual Education Programs (IEPs) determined that the

MAP would be an inappropriate use of their schools (Missouri Department of Elementary and Secondary Education, June, 1998). While scores were calculated as standard scores, for the purposes of meeting AYP requirements, standard scores were grouped into one of five groups by as set forth by DESE: Level 1, Below Basic, Basic, Proficient and Advanced. To meet the requirements of AYP, a set percentage of students in each of the total and disaggregated groups had to fall within the proficient or advanced categories. The percentage was set by federal mandate and increased in increments with a target of 100% of all students falling within the advanced and proficient range by 2014.

The state of Missouri annually administered the MAP beginning in 1998 although various changes have been made both to the grade levels assessed as well as the subjects being assessed. In 1998, all students in grades 3 through 8 were assessed annually in communication arts and mathematics. At the secondary level, the MAP was replaced with End of Course exams (EOCs) in the subject areas of English I and II, Algebra I and II, Geometry, Biology, American History, and Government (Missouri Department of Elementary and Secondary Education, 2012; Missouri Department of Elementary and Secondary Education, 2011).

The groupings for meeting AYP requirements were dropped from five to four to be aligned with NAEP reporting: Below Basic, Basic, Proficient, and Advanced.

While the Cape Girardeau No. 63 School District, like all public school districts in the state, participated in the MAP, the school district continued to conduct concurrent achievement assessments. Up until the 1990's the district continued to administer the Science Research Associates (SRA) Achievement tests in grades 2-6. Budget reductions in 1995 limited the administration of these tests to grades 2 and 6. Further budget reductions in 1999 eliminated administration of the SRA.

Beginning in 2002, the district began administering the Developmental Reading Assessment (DRA) in grades K-6 as a component of its implementation of the Balanced Literacy approach to teaching communication arts. This was later expanded to include grades 7 and 8. The DRA2 continues to be administered in late August/early September in grades 1-8, in December/January in kindergarten and grades 5-8, and in mid-April/mid-May in grades K-8. A Running Reading Record is given to students in grades 1-4 in December/January (Cape Girardeau School District, August, 2010). The DRA2 is a norm-referenced achievement test which reports scores in DRA levels ranging from A-80 which

correspond to grade level equivalencies K-Grade 7-8. The DRA and Running Reading Records are then categorized into one of four categories: Above Benchmark, On Benchmark, Below Benchmark(1-2 grade levels below), and 3+ Levels Below Benchmark.

Because of the elimination of a standardized achievement test, the district recognized the need for some type of benchmark assessment to monitor student achievement in the areas of writing and mathematics. Beginning in 2006, common assessments in writing and mathematics were developed. According to the district's testing schedule, the math common assessments are given to grades K-8 in November, late February/early March, and early May. The writing common assessments are given to grades 1-8 in late August/early September, to grades K-6 in early December/mid-January, and in grades K-8 in mid-April/mid-May (Cape Girardeau School District, August, 2010).

These common assessments were developed by teachers in grade level meetings. Common writing assessments are graded according to a rubric which results in a numerical score as well as a categorical score corresponding to the four categorical levels of the MAP: Below Basic, Basic, Advanced, and Proficient. Common math assessments result in a percentage score. These percentage scores, too, are

grouped into five categories: 92-100% indicating the student is performing at an advanced level; 82-91% indicating the student is performing at a proficient level; 70-81% indicating the student is performing at a basic level; 60-69% indicating the student is performing at a below basic level; and 0-59% indicating the student is most likely performing below grade level by at least two or more levels.

Cape Girardeau Public School District, like most schools across the state and nation, has done little to no research to determine if some or all of their assessments have any correlation with performance on the MAP. Therefore, this study examines a cohort of students from third grade through the end of fourth grade to determine if there are any significant relationships between student performance on the third grade MAP and third grade DRA2 scores, third grade MAP and fourth grade MAP scores, third grade MAP and fourth grade DRA2 scores, third grade DRA2 scores and fourth grade MAP, third grade DRA2 scores and fourth grade DRA2 scores, and fourth grade MAP and fourth grade DRA2 scores. The research also examines if teachers of third and fourth grade students take the information from the DRA2 to inform and change instructional practices throughout the school year.

Problem and Purposes Overview

The initial premise of the district for administering regular common assessments in reading, writing, and mathematics was two-fold: 1) to periodically assess (or "benchmark") student achievement prior to the MAP and allow for instructional adjustments in communication arts and mathematics; and 2) to establish some predictability of individual student success on the MAP in the areas of communication arts and mathematics.

The purpose of this study is to investigate the effectiveness of the local assessment practices of the school district as it relates to the study of the communication arts portion of the district's curriculum. According to the district's Comprehensive School Improvement Plan (CSIP), the district is to conduct a formal analysis of scores to determine any correlational trends between the district's reading and writing assessments when compared to the MAP communication arts results (Cape Girardeau Public Schools, October 18, 2010).

Research Hypotheses

The null hypotheses of this research were:

H01^a There are no significant correlational relationships between third grade MAP communication arts scores and the third grade DRA2 scores.

H01^b: There are no significant correlational relationships between fourth grade MAP communication arts scores and the fourth grade DRA2 scores.

H01^c: There are no significant correlational relationships when comparing third grade DRA2 scores to fourth grade MAP communication arts scores.

Research questions answered in this research were:

1. How does the third grade DRA2 predict performance on the MAP communication arts portion?
2. How does the fourth grade DRA2 in predict performance on the MAP communication arts portion?
3. How does the third grade DRA2 predict performance on the MAP communication arts portion in fourth grade?

Population and Sample

To analyze the correlations between the MAP and the DRA2, scores from students in grades 3 and 4 in each of the five elementary schools of the district were utilized. To be included in the sample, students took both the DRA2 and the MAP in the Cape Girardeau Public School District during the 2009-10 school year and the 2010-11 school years. The selection of these particular school years allowed for a consistent comparison of both assessment instruments as the format of the MAP for those particular years included three elements: selected response items, constructed response

items and performance events. Subsequent years eliminated the use of the performance event task of the MAP. Students who took the MAP-Alternate test were not included in the sample.

The study was comprised of two cohorts of students. The total sample of Cohort 1 students (students in grades 3 and 4 during the 2009-10 school year) consisted of 202 students attending grade 3 and 4 in the Cape Girardeau Public School District. The sample included 83 males and 119 females. Racial demographics for the sample included 2 Asian students, 75 Black students, 6 Hispanic students, 1 Indian student, and 118 White students. The total sample of Cohort 2 students (students in grades 3 and 4 during the 2010-11 school year) consisted of 194 students attending grade 3 and 4 in the Cape Girardeau Public School District. The sample included 79 males and 115 females. Racial demographics for the sample included 7 Asian students, 64 Black students, 6 Hispanic students, 1 Indian student, and 116 White students. This population percentage for each racial demographic is consistent with the district's total racial distribution for the school years of the sample.

Data Collection and Instrumentation

The purpose of this study was to use a correlational analysis of the DRA2 (the common formative assessment used

by the Cape Girardeau Public School District to assess the area of communication arts) and the MAP (the large scale summative evaluation) to determine if the common formative aligned to the summative evaluation.

The research is an *ex post facto* design as the purpose of the study was to describe and interpret existing data to determine if correlations exist between two previously administered tests. Kerlinger (1986) states that this type of research is a systematic empirical inquiry in which the researcher does not have direct control of the variables because "their manifestations have already occurred or because they are inherently not manipulable" (p.520).

Research was collected using the district's data warehouse system termed PULSE. This system is computerized and allows the district to collect both district-created assessment data from each student enrolled in the Cape Girardeau Public School district as well as state- and federal-assessment data for each student enrolled in the district.

For the purposes of this study, results from the spring administration of the Developmental Reading Assessment - Second Edition (DRA2) were compared to results of the communication arts results of the Missouri Assessment Program (MAP).

The DRA2 is an individually administered test of reading that measures what the authors describe as "...three crucial components of reading: reading engagement, oral reading fluency, and comprehension" (Pearson, 2011, p. 11). Assessments are conducted during one-on-one sessions where students read selected assessment texts and respond to questions or prompts either in verbal or written form (Pearson, 2011). Students reading DRA2 texts Level 28 and above also write a one-page summary of what they have read in addition to writing responses to literal comprehension, interpretation, and reflections questions or prompts. Typically, testing sessions for third and fourth grade students last approximately 45 minutes to one hour. Test-Retest reliability for the DRA2 for Comprehension is .99 for the grade range 1-3 and .97 for the grade range 4-6. Test-Retest reliability for the DRA2 for Fluency is .97 for the grade range 1-3 and .93 for the grade range 4-6 (Pearson, 2011). Correlations between the DRA2 and three other well-known tests of reading comprehension and fluency have been conducted to establish criterion-related validity. Classifications ranging from large to very large in magnitude have been shown when comparing the DRA2 to the Gray's Oral Reading Test-fourth Edition (GORT-4), the DIBELS Oral Reading Fluency Test-6h Edition (DORF), and the

Gates MacGinitie Reading Test-fourth Edition with correlation coefficients ranging from .60 to .76. According to Pearson,

It should be noted that these correlations are especially impressive given the different tasks and scoring procedures involved among the various tests. Furthermore, the variation in the student sample means that this test demonstrates validity for a wide range of ages, ethnic backgrounds, and reading levels. In sum, the results from the concurrent validity study provide support for the validity of the DRA2 as a measure of reading comprehension and fluency (2011, p.57).

The Missouri Assessment Program (MAP) is a group-administered assessment. In the third and fourth grades, the MAP is administered in the areas of communication arts and mathematics. During this study only the communication arts portion was analyzed. The Communication arts portion of the MAP is administered over a four-day period in grade 3 and a three-day period in grade 4. Session 3 of the grade 3 communication arts portion of the MAP and Session 2 of the grade 4 communication arts portion of the MAP are strictly timed Terra Nova sessions. All remaining sessions have timing guidelines but are considered to be untimed as

long as a student or students are making conscientious effort to complete the test as judged by the test administrator. Test items include selected-response items, constructed-response items, and performance events (including writing prompts) (Missouri Department of Elementary and Secondary Education, 2010). The reliability on the MAP tests was evaluated using Cronbach's coefficient alpha. Reliability was .91 for grade 3 communication arts and .93 for grade 4 communication arts (Missouri Department of Elementary and Secondary Education, 2010) indicating acceptable reliability for this portion of the MAP. Construct validity was established using a subtype construct validity of divergent validity. According to the *Missouri Assessment Program: Grade-Level Assessments Technical Report 2010*, correlations were computed between the communication arts, mathematics, and Science scale scores. Correlation coefficients were .71 when comparing grade 3 communication arts to grade 3 Math and .73 when comparing grade 4 communication arts to grade 4 Math suggesting moderate correlations between the two subject areas (Missouri Department of Elementary and Secondary Education, 2010). The Department notes that while the tests are not perfectly related, suggesting different constructs are being "tapped," the tests do appear to be "tapping into

a similar knowledge base" (Missouri Department of Elementary and Secondary Education, 2010, p. 140).

Permission was obtained from the Cape Girardeau Public School District in 2010 to collect and analyze any data reported on PULSE. For the purposes of this study, it was determined that correlational research (or associational research) would be appropriate when analyzing collected data as there was no attempt to influence the variables being analyzed.

According to Fraenkel and Wallen (2003), correlational research is carried out to either explain human behaviors or predict likely outcomes. The purpose of this study is the latter: to predict likely outcomes. More specifically, the purpose of this study is to predict a student's performance on the MAP based upon that student's performance on the DRA2. Pearson product-moment correlation (or Pearson r) will be used to analyze correlations between the DRA2 and MAP at each of the two grades being studied. According to Gravetter and Wallnau (2004) a Pearson correlation "measures the degree and direction of linear relationship between two variables" (p. 527). A Pearson r can range from .00 meaning there is no relationship between the variables all the way up to 1.00. When there is a perfect linear relationship as the value of

X increases, there is a perfectly predictable increase in Y with the result of $r = 1.00$. Likewise the relationship can range from $.00$ to -1.00 meaning that as the value of X increases, there is a perfectly predictable decrease in Y with $r = -1.00$.

Strength of relationships may be interpreted as follows:

If $r = +.70$ or high Very strong positive relationship
 $+.40$ to $+.69$ Strong positive relationship
 $+.30$ to $+.39$ Moderate positive relationship
 $+.20$ to $+.29$ Weak positive relationship
 $+.01$ to $+.19$ No or negligible relationship
 $-.01$ to $-.19$ No or negligible relationship
 $-.20$ to $-.29$ Weak negative relationship
 $-.30$ to $-.39$ Moderate negative relationship
 $-.40$ to $-.69$ Strong negative relationship
 $-.70$ or higher Very Strong negative relationship
 (faculty.quinnipiac.edu/libarts/polsci/statistics.html, 2013).

In referencing the previously stated research questions, the combination of the DRA2 and communication arts MAP data from grade 3 will be analyzed for correlation between those two variables and performance on the communication arts MAP in grade 4. The combination of the

third grade DRA2, third MAP communication arts, and fourth grade DRA2 will also be analyzed for predicting performance on the MAP communication arts portion in grade 4.

Summary

The primary focus of this study is to predict a student's performance on the MAP based upon that student's performance on the DRA2. Both assessments are formal, standardized assessments. The DRA2 is a reading assessment that is used commonly across the United States to assess reading competency. The MAP is the summative assessment currently used by the state of Missouri to meet the guidelines of the No Child Left Behind act.

Although the MAP and the DRA2 are administered in the Cape Girardeau Public School district in grades 3-8, this study was limited to analysis of students in grades 3 and 4. Students included in the sample took both the DRA2 and the MAP during the 2009-10 school year and the 2010-11 school year utilizing a sample of 396 students. Data was analyzed using Pearson product-momentum correlation.

CHAPTER 4

ANALYSIS OF DATA

Introduction

The Cape Girardeau Public School District in compliance with the mandates of the No Child Left Behind (NCLB) Act assesses all students in grades 3-8 with an end-of-the-year assessment or summative assessment known as the Missouri Assessment Program (MAP). Additionally, throughout the year, the district administers a series of periodic assessments to help guide or form instruction throughout the year. These assessments, also known by the district as formative assessments, have been assumed to correlate with performance on the MAP.

The problem statement for this study focused on whether or not the instrument selected as a common formative assessment for the district in the area of communication arts (DRA2) correlated with the summative assessment (the MAP). The purpose of this study was to use a correlational analysis of the DRA2 (the common formative assessment used by the Cape Girardeau Public School District to assess the area of communication arts) and the MAP (the large scale summative evaluation) to determine if the common formative aligned to the summative evaluation. More specifically, the purpose of this study is to predict

a student's performance on the MAP based upon that student's performance on the DRA2.

Organization of Data Analysis

To analyze the correlations between the MAP and the DRA2, scores from students in grades 3 and 4 in each of the five elementary schools of the district were utilized. As the configuration of the Cape Girardeau Public School's elementary grade levels is K-4, these grades were chosen they are the only two grades that take both the MAP and the DRA2 as MAP is not administered across the state of Missouri until grade 3. Therefore, this is an *ex post facto* design in which there are no control schools in existence for comparison purposes.

To be included in the sample, students took both the DRA2 and the MAP in the Cape Girardeau Public School District during the 2009-10 school year and the 2010-11 school years. The selection of these particular school years allowed for a consistent comparison of both assessment instruments as the format of the MAP for those particular years included three elements: selected response items, constructed response items, and performance events. Subsequent years eliminated the use of the performance event task of the MAP. Students who took the MAP-Alternate test were not included in the sample.

The study was comprised of two cohorts of students. The total sample of Cohort 1 students (students in grades 3 and 4 during the 2009-10 school year) consisted of 202 students attending grade 3 and 4 in the Cape Girardeau Public School District. The sample included 83 males and 119 females. Racial demographics for the sample included 2 Asian students, 75 Black students, 6 Hispanic students, 1 Indian student, and 118 White students. The total sample of Cohort 2 students (students in grades 3 and 4 during the 2010-11 school year) consisted of 194 students attending grade 3 and 4 in the Cape Girardeau Public School District. The sample included 79 males and 115 females. Racial demographics for the sample included 7 Asian students, 64 Black students, 6 Hispanic students, 1 Indian student, and 116 White students. This population percentage for each racial demographic is consistent with the district's total racial distribution for the school years of the sample.

Pearson product-momentum correlation (or Pearson r) were performed to determine correlational relationships between third grade DRA2 scores and third grade MAP communication arts scores, fourth grade DRA2 scores and fourth grade MAP communication arts scores, and, finally, third grade DRA2 scores to fourth grade MAP communication arts. Test scores were converted to z-scores in order to

make comparisons between the DRA2 and MAP communication arts scores. The results of the calculations are located in Table 1.

Table 1

Pearson Correlations between DRA2 and MAP

Cohort 1		
	Third Grade MAP	Fourth Grade MAP
Third Grade DRA2	.76**	.65**
Fourth Grade DRA2		.70**
Cohort 2		
	Third Grade MAP	Fourth Grade MAP
Third Grade DRA2	.65**	.74**
Fourth Grade DRA2		.76**

**Correlation is significant at the .01 level (2-tailed)

Results on Correlation Analyses

Correlation coefficients were computed between the DRA2 scores and MAP scores for both third and fourth grades for both the 2009-10 (Cohort 1) and 2010-11 (Cohort 2) school years. The results of the correlational analyses presented in Table 1 show that 6 out of the 6 correlations were statistically significant and were greater than or equal to .01.

Using the descriptors of Quinnipiac University (2013) in interpreting Pearson r , when comparing the Cohort 1 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a very strong positive correlation between the scores at both grade levels. When comparing the Cohort 1 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a strong correlation between the scores.

When comparing the Cohort 2 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a strong positive correlation between the scores at the third grade level and a very strong positive correlation between the scores at the fourth grade level. When comparing the Cohort 2 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a very strong correlation between the scores.

In addressing Research Question 1 "How does the third grade DRA2 predict performance on the MAP communication arts portion?" results from Cohort 1 reflected a very strong positive correlation ($r = .76, p < .01$) while results from Cohort 2 reflected a strong positive correlation ($r = .65, p < .01$). Based on the results of these two cohorts, the DRA2 does appear to be a good predictor of

performance on the communication arts portion of the MAP for third grade students.

Research Question 2 asks, "How does the fourth grade DRA2 predict performance on the MAP communication arts portion?" Results from Cohort 1 found a very strong positive correlation ($r = .70$, $p < .01$) as did results from Cohort 2 ($r = .76$, $p < .01$). Based on the results of these two cohorts, the DRA2 does appear to be a good predictor of performance on the communication arts portion of the MAP for fourth grade students.

Finally, in answering Research Question 3, "How does the third grade DRA2 predict performance on the MAP communication arts portion in fourth grade?", there was a strong positive correlation between the DRA2 and MAP for Cohort 1 ($r = .65$, $p < .01$) and a very strong positive correlation between the DRA2 and MAP for Cohort 2 ($r = .74$, $p < .01$). Even with nearly a year gap between administrations of the two tests, the third grade DRA2 does appear to be a good predictor of performance on the communication arts portion of the MAP for fourth grade students.

Summary

All correlations between the DRA2 and MAP for both grade levels over both cohorts reflected strong or very

strong positive correlations. Students scoring lower on the DRA2 also score lower on the MAP both within their own grade level and when comparing between the third and fourth grade level. Likewise, students scoring higher on the DRA2 also score higher on the MAP both within their own grade level and when comparing between the third and fourth grade level. All correlations were significant at the .01 level.

The purpose of this study was to predict a student's performance on the MAP communication arts portion based upon that student's performance on the DRA2. Based upon results of this study using Pearson product-moment correlations (or Pearson r), there are significant positive correlations between a student's performance on the DRA2 and the MAP communication arts portion at the third and fourth grade levels. Therefore, a student's performance on the DRA2 does appear to be a strong to very strong predictor of performance on the MAP communication arts portion at both the third and fourth grade levels. A discussion of implications of this study and possible avenues for future research follows in Chapter Five.

Chapter 5

DISCUSSION

Summary of the study

Following the publication of *A Nation at Risk* in 1983, the federal government initiated efforts that would increase the accountability of the education of America's youth (Marshall & Gerstl-Pepin, 2005). Efforts initially relied on voluntary cooperation from state and local officials and tied Title I funding to each state's development of content and performance standards (Peerson & West, 2003). In 2001, another reauthorization of the Elementary and Secondary Education Act (ESEA) renamed the No Child Left Behind act (NCLB) required each state to assess student performance on a standardized basis for the purpose of determining adequate yearly progress (AYP) among schools with Title I funding tied to a district's ability to meet AYP (Peerson & West, 2003). While states had the latitude to select the format of the assessment, NCLB mandated that students in grades three through eight be tested each year in reading and math and again during one year in high school.

Although Missouri implemented end-of-the-year summative assessments since 1978, to address the requirements of NCLB, the state of Missouri developed an

assessment program known as the Missouri Assessment Program (MAP). The MAP is given in grades 3-11 in the Spring and serves as a summative assessment of the content standards that should have been mastered by students in a particular grade level (Missouri Department of Elementary and Secondary Education, June, 1998). The Missouri Department of Elementary and Secondary Education was quoted in a 1998 publication titled *Assessment Standards for Missouri Public Schools* that administrative use of the MAP would include,

...inform[ing] parents, the school board and the public about the effectiveness of schooling and ...revealing trends that will aid in setting priorities, planning the curriculum, revising instructional practices, and evaluating the progress of educational programs (Missouri Department of Elementary and Secondary Education, June, 1998, p. 2).

This same publication indicated instructional uses for teachers included, "...identifying strengths and weaknesses in classroom curriculum instruction" and "assisting[ing] teachers in monitoring individual student learning" (p.2).

Researches note that while end-of-the-year summative assessments such as the MAP are frequently analyzed for specific student performance, there are difficulties in

using these types of assessments to inform classroom instruction. Dirkson (2011) notes that while summative assessments can be formative, limitations in timing and opportunity to revisit information frequently preclude summative assessments from this use.

Popham (2008) observes that many, if not most, state accountability tests that used at the time are often amalgamations of student performance that made it hard to judge specific student performance since 1) student performance is "lumped" at a strand level, and 2) results occur *after* students leave a grade level with the end outcome of teachers being unable to adequately adjust instruction. Results of these end-of-the-year summative assessments are analyzed and projected on an entirely different group of students who may or may not benefit by the adjustments of instruction such analysis provides (Ainsworth & Viegut, 2006; Condermand & Hedint, 2012; Popham, 2008).

Many researchers indicate that a more useful instrument in making instructional decisions that impact student instruction and student achievement is the use of formative assessments (Ainsworth & Viegut, 2006; Black & William, 1998; Chappuis & Chappuis, 2006; Chappuis, Commodore & Stiggins, 2010; Marzano, 2006; Moss &

Brookhart, 2012; Popham, 2008; Reeves, 2007). Dorn (2011) observed that NCLB should have achieved the result of expanding formative assessments. Popham (1998) found evidence that formative assessment raises standards, concluding that formative assessment was an intervention leading to one of the largest gains in student learning ever reported. Stiggins and DuFour (2009) also noted the value of formative assessments observing that while national or state summative assessments were widely used to make changes at the institutional or policy-creation level, assessments given at the school-level best served the staff about overall classroom or school-wide performance and also helped teachers identify areas that the teacher might need to strengthen within his or her own pedagogical skill set.

As the MAP is given only once during the Spring of the school year, the Cape Girardeau No. 63 District (also referred to as the Cape Girardeau Public School District) recognized the need for more frequent assessments to gauge performance. The school district, therefore, implemented a series of common formative assessments. These assessments were to be used by the district to gather more timely and detailed information and to inform instruction with the end result that scores would be positively influenced on the year end summative assessment, the MAP.

Beginning in 2002, in the area of communication arts, the district chose the Developmental Reading Assessment (DRA) as a component of its total formative assessment program. The district currently administers the Developmental Reading Assessment, Second Edition (DRA2) to periodically analyze student performance in the area of communication arts. However, Cape Girardeau Public School District, like most schools across the state and nation, has done little to no research to determine correlation of the DRA2 or any of its other formative assessments with performance on the MAP.

To that end, the purpose of this study was to predict a student's performance on the MAP based upon that student's performance on the DRA2. A correlational analysis using Pearson product-moment correlations (Pearson r) was conducted on two cohorts of students. Cohort 1 was comprised of students in grades 3 and 4 attending one of the five Cape Girardeau Public Schools during the 2009-10 school year and who took both the Spring DRA2 and the MAP. Cohort 2 students was comprised of students in grades 3 and 4 attending one of the five Cape Girardeau Public Schools during the 2010-11 school year and who also took both the Spring DRA2 and the MAP.

Research questions answered in this research were:

1. How does the third grade DRA2 predict performance on the MAP communication arts portion?
2. How does the fourth grade DRA2 in predict performance on the MAP communication arts portion?
3. How does the third grade DRA2 predict performance on the MAP communication arts portion in fourth grade?

Research was collected using the district's data warehouse system termed PULSE. For the purposes of this study, results from the spring administration of the DRA2 were compared to results of the communication arts results of the Missouri Assessment Program (MAP). Permission was obtained from the Cape Girardeau Public School District in 2010 and given to this researcher to collect and analyze any data reported on PULSE.

For the purposes of this study, it was determined that correlational research (or associational research) would be appropriate when analyzing collected data as there was no attempt to influence the variables being analyzed.

According to Frankel and Wallen (2003), correlational research is carried out to either explain human behaviors or predict likely outcomes. Pearson product momentum correlations (or Pearson r) was used to analyze correlations between the DRA2 and MAP at each of the two

grades being studied. The combination of the DRA2 and communication arts MAP data from grade 3 was analyzed for correlation between those two variables and performance on the communication arts MAP in grade 4. The combination of the third grade DRA2, third MAP communication arts, and fourth grade DRA2 was also analyzed for predicting performance on the MAP communication arts portion in grade 4.

Findings

The study was comprised of two cohorts of students. The total sample of Cohort 1 students (students in grades 3 and 4 during the 2009-10 school year) consisted of 202 students attending grade 3 and 4 in the Cape Girardeau Public School District. The sample included 83 males and 119 females. Racial demographics for the sample included 2 Asian students, 75 Black students, 6 Hispanic students, 1 Indian student, and 118 White students. The total sample of Cohort 2 students (students in grades 3 and 4 during the 2010-11 school year) consisted of 194 students attending grade 3 and 4 in the Cape Girardeau Public School District. The sample included 79 males and 115 females. Racial demographics for the sample included 7 Asian students, 64 Black students, 6 Hispanic students, 1 Indian student, and 116 White students.

Correlation coefficients were computed between the DRA2 scores and MAP scores for both third and fourth grades for both the 2009-10 (Cohort 1) and 2010-11 (Cohort 2) school years. The results of the correlational analyses presented in Table 1 show that 6 out of the 6 correlations were statistically significant and were greater than or equal to .01.

Using the descriptors from Quinnipiac University (2013) when comparing the Cohort 1 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a very strong positive correlation between the scores at both grade levels. When comparing the Cohort 1 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a strong correlation between the scores.

When comparing the Cohort 2 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a strong positive correlation between the scores at the third grade level and a very strong positive correlation between the scores at the fourth grade level. When comparing the Cohort 2 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a very strong correlation between the scores.

Conclusions

In addressing Research Question 1 "How does the third grade DRA2 predict performance on the MAP communication arts portion?" results from Cohort 1 reflected a very strong positive correlation ($r = .76$, $p < .01$) while results from Cohort 2 reflected a strong positive correlation ($r = .65$, $p < .01$). Based on the results of these two cohorts, the DRA2 does appear to be a good predictor of performance on the communication arts portion of the MAP for third grade students.

Research Question 2 asks, "How does the fourth grade DRA2 predict performance on the MAP communication arts portion?" Results from Cohort 1 found a very strong positive correlation ($r = .70$, $p < .01$) as did results from Cohort 2 ($r = .76$, $p < .01$). Based on the results of these two cohorts, the DRA2 does appear to be a good predictor of performance on the communication arts portion of the MAP for fourth grade students.

Finally, in answering Research Question 3, "How does the third grade DRA2 predict performance on the MAP communication arts portion in fourth grade?", there was a strong positive correlation between the DRA2 and MAP for Cohort 1 ($r = .65$, $p < .01$) and a very strong positive correlation between the DRA2 and MAP for Cohort 2 ($r = .74$,

$p < .01$). Even with nearly a year gap between administrations of the two tests, the third grade DRA2 does appear to be a good predictor of performance on the communication arts portion of the MAP for fourth grade students.

Using the descriptors from Quinnipiac University (2013) when comparing the Cohort 1 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a very strong positive correlation between the scores at both grade levels. When comparing the Cohort 1 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a strong correlation between the scores.

When comparing the Cohort 2 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a strong positive correlation between the scores at the third grade level and a very strong positive correlation between the scores at the fourth grade level. When comparing the Cohort 2 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a very strong correlation between the scores.

As referenced earlier, while the Cape Girardeau Public School district has given formative assessments since 2005,

no formal research has been conducted by the district to determine a correlation until this study was conducted. As also referenced, this is not unusual at the state or national level as little peer-reviewed research exists either on the MAP as interviews indicated that peer-reviewed research on specific state tests are rare (personal communication, December 10, 2012) making comparison of the results of this study to prior research difficult.

This particular body of research appears to be unique not just to this district and state but to other states as well. While formative assessments are certainly recommended for administration with the aim of making instructional decisions that impact student achievement (Ainsworth & Viegut, 2006; Black & William, 1998; Chappuis & Chappuis, 2006; Chappuis, Commodore & Stiggins, 2010; Marzano, 2006; Moss & Brookhart, 2012; Popham, 2008; Reeves, 2007), few, if any, states conduct formal research to determine correlations between any formative assessments administered during the year and results of end-of-the-year, high stakes summative testing. Therefore, to disagree with or dispute research conducted on this subject of this research study is challenging since the greatest body of research is done in the area of formative assessment.

Taking into account research, it does appear that districts are well-served to conduct frequent, formative assessments. Ainsworth and Viegut (2006) have stated that if common formative assessments are aligned to the large scale assessments, they would offer predictive value to the summative evaluation administered. In fact, Ainsworth and Viegut (2006) specifically state, "When intentionally aligned..., each level of assessment results provides educators with 'predictive value' as to how students are likely to do on the next level of assessment" (p. 19). It does appear, therefore, that continuing to formally analyze formative assessments for correlation to the summative assessment does serve specific instructional value.

Implications

As high-stakes summative testing continues to be mandated, school districts would certainly want to ensure best assessment practices are followed. Research certainly indicates formative assessments are being useful in making timely instructional decisions (Ainsworthy & Viegut, Moss & Brookhart, 2012; Popham, 2008; Reeves, 2007). There also does not appear to be anything that would preclude districts from ensuring that formative assessments correlate with end-of-year high-stakes summative testing by conducting formal analysis of correlation.

Because high-stakes summative testing in the United States is mandated by federal law (U.S. Department of Education, 2004), states and districts wishing to continue to receive federal funding must participate and meet the accountability standards set forth by federal law. While educators and test developers argue that summative assessments can be used as formative assessments, practitioners in the field note limitations in timing and opportunity when using summative assessments in a formative manner (Dirkson, 2011).

Madaus & Russell (2010/2011) state, "High-stakes tests provide information that confirms what most teachers already know. For students whose level of achievement is relatively low, however, current [high-stakes] tests fail to provide diagnostic information about *why* they are struggling" (p.21). Dorn (2011) notes that NCLB should have achieved the result of expanding formative assessments that mimic or parallel the state summative tests and suggest that these assessments should be given several times throughout the school year. Clark (2010) states that formative and summative assessment practices can be aligned. Most current researchers in the field of educational assessment overwhelmingly recommend the use of formative assessment to analyze student performance and

guide instructional decision making (Ainsworth & Viegut, 2006; Conderman & Hedin, 2012; Dorn, 2011; Heritage, 2007; Popham, 2008; Reeves, 2009; Stiggins & Chappuis, 2005). Therefore, assimilating the thoughts of these researchers, implications for application of research is clear: schools district must actively work to align what is mandated with what is best practice for students (i.e. expanded use of formative assessments).

Ainsworth and Viegut (2006) suggest an approach that begins with data. These authors are very specific in their steps for data-driven instruction decision making, outlining five steps in this process:

1. The charting of student performance data;
2. Analyzing the data;
3. Setting a goal for improvement;
4. Selecting specific teaching strategies to meet that goal;
5. Determining results indicates to gauge the effectiveness of the selected teaching strategies (p.13)

Based upon this data-driven instruction decision making process, the authors describe three approaches to standards-based instruction and assessment. One of the three approaches has direct implications for districts

choosing to begin with analysis of state assessment data.

This approach consists of the following steps:

- a. Analyze state assessment data to identify areas where students are scoring low and to determine the weight of state test requirements of standards.
- b. Identify the standards representing areas of need.
- c. Unwrap the prioritized standards and determine Big Ideas and Essential Questions (p.16).

Ainsworth and Viegut describe the process as circular (i.e. once you complete the final step, you begin the process again).

Predicated upon identification of prioritized standards, is the selection and/or development of school-based formative pre- and post-assessments with data teams analyzing data and developing effective teaching strategies based upon results of the assessments and adjusting assessments as necessary to finesse the alignment of the formative assessments to the prioritized standards once again creating a circular process.

Following administration of school-based common formative post-assessment, Ainsworth and Viegut suggest that district formative or summative assessments be

administered prior to administration of the annual state assessment. Again, data is analyzed and adjustments made to the school-based post-assessments and district formative and summative assessments.

The Ainsworth and Viegut approach is supported by research of other educational experts. Rick Stiggins and Rick DuFour (2009) describe a similar process containing three levels of assessments: classroom assessments, school level assessments, and institutional-level assessments.

In referencing the practice of the Cape Girardeau Public School district, the district has selected a formative instrument, the DRA2, which has a strong to very strong correlation with the summative instrument used by the state of Missouri, the MAP. The district has knowledge that few districts within the nation appear to have: a formal analysis of the correlation of a formative assessment to an end-of-year summative assessment. However, to make the most effective use of this knowledge, the district must implement the cyclical data analysis system suggested by researches and use formative assessment in the flow of the instructional process (Ainsworth & Viegut, 2006; Shepard, L., Hammerness, K., Darling-Hammond, L. & Fust, F., 2005; Stiggins & DuFour, 2009). Therefore, while correlation has been established which confirms the

assessments are aligned, the district can take its data analysis of the DRA2 one step further: it can begin to look at the skills within the DRA2 to make instructional decisions regarding students. It can also analyze the DRA2 for information that it does not give and develop school-based assessments that will help fill in the gaps of information. In other words, teachers should be answering the questions Stiggins and DuFour suggest should be answered for any assessment administered:

1. What are the instruction decisions to be made?
2. Who will be making the decisions?
3. What information will help them make good decisions? (p. 640).

Using these questions as guides, it is suggested that while the district has confirmed knowledge that the selected formative assessment it is using does positively correlate with the end-of-year summative assessments, there are some additional steps to be taken to make the best use of the information provided by the DRA2. Educators would do well to examine the DRA2 to determine the types of questions asked and review the test for error analysis both at the classroom level and the student level. Halverson refers to this process as actuation and states,

"Schools need structured occasions to burn assessment information into actionable knowledge. Actuation refers to the process through which faculty and staff come to understand and act upon the effects of their interventions on student learning" (p.131).

Research indicates that the teacher should be the "who" in the "Who will make the decisions?" (Ainsworth & Viegut; Bakula, 2010; Clark, 2010; Conderman & Hedin, 2012; Popham, 2008; Stiggins & DuFour, 2009), when examining the DRA2. In addressing what instructional decisions are to be made and what information will help them make good decisions, the process of actuation suggests that teachers must have time and opportunity to delve into the assessments they give beyond looking at the scores students obtain as the primary means of assessing mastery of concepts. Research on formative assessment suggests that the questions asked and how they are answered gives insight into student performance and allows the teacher to give timely and individualized feedback to students. The function of such feedback is to assist teachers in making adjustments to their instruction, whether it be to the whole class or to an individual student, that will improve student achievement (Popham, 2008).

As it relates to this specific research, teachers must have time to review the DRA2 to first analyze what the test asks and how it is asked. They must then analyze results of the test to see how students responded to specific questions and look for patterns of responses at both the classroom level and the individual student level. This information must both affect instruction and be used when giving feedback to students or the information from assessment loses its full instructional impact. This study indicates the results of the DRA2 results can be expanded now that the relationship between the DRA2 and MAP has been established. Analyzing not just the final score but the specific errors carries with it the possibility of the opportunity to make specific changes to instructional delivery in communication arts based upon data analysis of the DRA2.

In conclusion, significant positive correlation has been established between the DRA2 and the MAP in grades 3 and 4 for this school district. This information confirms the district has selected an appropriate instrument (the DRA2) that can predict final performance on the end-of-year summative assessment mandated by the state (the MAP). However, research indicates that the district can make even more productive use of the formative instrument now that

has ascertained correlation exists. A wealth of information exists on how school districts can analyze formative assessments to impact instruction. Research suggests that the district would make significant gains in student achievement should it take the next steps in application of formative assessment.

Future Research

Continued research into the correlation of the formative assessments and other summative assessments is certainly warranted. While most public schools across the nation do give their state's version of an end-of-year summative assessment, no research was found that established correlation between any end-of-year summative assessments and other formative assessments. Correlational studies between the two would appear to benefit both teachers and students who, according to Doug Reeves, live in a nation that is over-tested but under-assessed. Giving tests which do not inform instruction while hoping to improve student performance on high-stakes, summative assessment appears to be an unnecessary waste of time, effort, and monetary resources.

Missouri, like most states in the United States, will soon be switching to one of the two summative assessments that are aligned to the new Common Core Standards. With the

shift of emphasis from the state-designed tests to these new summatives, the Smarter Balanced Assessment Consortium (SBAC) and the Partnership for Assessment of Readiness for College and Careers (PARCC), it would appear that the need for research into formative assessments that significantly correlate with these summatives would be useful and make teaching time productive, especially as a new authorization of the ESEA emphasizes an individual student growth model over the current use of meeting set Annual Yearly Progress (AYP) goals. This is a particularly salient point when one is aware that formative assessments yield particularly good results with low achievers (Black & Wiliam, 1998).

Another area not covered in this research was the particular way the DRA2 was used by the teachers of the Cape Girardeau Public School District. Other than the knowledge that the test was administered, little is known about how teachers translated the knowledge they gained from administration of the DRA2 into their educational practices. If the DRA2 is truly a formative assessment, teachers should be analyzing and applying information that is skill specific to their instruction. This body of research did not explore either analysis or application.

Finally, great limitations are apparent when attempting to generalize this research to other entities.

This study was limited to grades 3 and 4 in five elementary schools in one school district in Southeast Missouri. The demographic of the district was predominantly white. Generalization of the results of this study to other districts should be done with great caution as results could be very different at grade levels above 3 and 4. Results could also be different based upon a different demographic make-up of a district. However, even this aspect would be of interest for further research as correlations could be different based upon grade level, ethnicity, and even social-economic status of the student population.

Summary

Missouri, like most states in the nation, developed a statewide, summative assessment system that was in keeping with federal law. Also, in keeping with federal law, this summative assessment was administered only one time during the year. In Missouri, administration was conducted in the Spring at the end of the school year. Beginning in 2009, the Cape Girardeau Public School District put additional periodic assessments in the place which were referred to by the district as common formative assessments. One of these assessments, the DRA2 was used as a formative assessment in the area of communication arts. Assessments were intended

to reflect the GLEs taught for mastery at the specific grade level administered and were believed to have some correlation to performance on the end-of-the-year summative assessment, the MAP.

The purpose of this study was to predict a student's performance on the MAP based upon that student's performance on the DRA2 as no formal analysis of scores had been conducted by the district even though correlation between the two tests was assumed. To analyze the correlations between the MAP and the DRA2, scores from students in grades 3 and 4 in each of the five elementary schools of the Cape Girardeau Public School district were utilized. The configuration of the elementary buildings of the district is grades K-4. Therefore, only grades 3 and 4 were analyzed for correlation as only grades 3 and 4 are administered both the DRA2 and the MAP in these buildings. To be included in the sample, students must have taken both the DRA2 and the MAP in the Cape Girardeau Public School District during the 2009-10 and 2010-11 school years. The selection of these particular school years allowed for a consistent comparison of both assessment instruments as the format of the MAP for the years selected were consistent.

Cohort 1 students (students in grades 3 and 4 during the 2009-10 school year) consisted of 202 students. Cohort

2 students (students in grades 3 and 4 during the 2010-11 school year) consisted of 194 students. The population percentage for each racial demographic in the student sample was consistent with the district's total racial distribution for the school years of the sample.

Research questions answered in this research were:

1. How does the third grade DRA2 predict performance on the MAP communication arts portion?
2. How does the fourth grade DRA2 in predict performance on the MAP communication arts portion?
3. How does the third grade DRA2 predict performance on the MAP communication arts portion in fourth grade?

Research was collected using the district's data warehouse system termed PULSE. For the purposes of this study, it was determined that correlational research would be appropriate when analyzing collected data. Person product-momentum correlation (or Pearson r) was used to analyze correlations between the DRA2 and MAP at each of the two grades studied.

Correlation coefficients were computed between the DRA2 scores and MAP scores for both third and fourth grades for both the 2009-10 (Cohort 1) and 2010-11 (Cohort 2) school years. Using the descriptors of Quinnipiac University (2013) in interpreting Pearson r , when comparing

the Cohort 1 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a very strong positive correlation between the scores at both grade levels. When comparing the Cohort 1 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a strong correlation between the scores.

When comparing the Cohort 2 Third and Fourth Grade DRA2 scores to the Third Grade and Fourth Grade communication arts scores, there was a strong positive correlation between the scores at the third grade level and a very strong positive correlation between the scores at the fourth grade level. When comparing the Cohort 2 Third Grade DRA2 scores to the Fourth Grade communication arts scores, there was a very strong correlation between the score.

In addressing Research Question 1 "How does the third grade DRA2 predict performance on the MAP communication arts portion?" results from Cohort 1 reflected a very strong positive correlation ($r = .76$, $p < .01$) while results from Cohort 2 reflected a strong positive correlation ($r = .65$, $p < .01$). Based on the results of these two cohorts, the DRA2 does appear to be a good

predictor of performance on the communication arts portion of the MAP for third grade students.

Research Question 2 asks, "How does the fourth grade DRA2 predict performance on the MAP communication arts portion?" Results from Cohort 1 found a very strong positive correlation ($r = .70$, $p < .01$) as did results from Cohort 2 ($r = .76$, $p < .01$). Based on the results of these two cohorts, the DRA2 does appear to be a good predictor of performance on the communication arts portion of the MAP for fourth grade students.

Finally, in answering Research Question 3, "How does the third grade DRA2 predict performance on the MAP communication arts portion in fourth grade?", there was a strong positive correlation between the DRA2 and MAP for Cohort 1 ($r = .65$, $p < .01$) and a very strong positive correlation between the DRA2 and MAP for Cohort 2 ($r = .74$, $p < .01$). Even with nearly a year gap between administrations of the two tests, the third grade DRA2 does appear to be a good predictor of performance on the communication arts portion of the MAP for fourth grade students.

Based upon results of this study using Pearson product-momentum correlations (or Pearson r), there are significant positive correlations between a student's

performance on the DRA2 and the MAP communication arts portion at the third and fourth grade levels. Therefore, a student's performance on the DRA2 does appear to be a strong to very strong predictor of performance on the MAP communication arts portion at both the third and fourth grade levels.

REFERENCES

- Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Thousand Oaks, CA: Corwin Press.
- Bakula, N. (2010). The benefits of formative assessments for teaching and learning. *Science Scope*, 34(1).
- Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2).
- Brookhart, S. (2008). *How to give effective feedback to your students*. Alexandria, VA: ASCD.
- Burke, G. & Want, Y. (2010). Methods and uses of classroom assessments employed in teaching grades three through five in five school districts in the Mississippi Delta. *Education*, 130(4), 657-665.
- Chappuis, J. & Chappuis, S. (2006). *Understanding school assessment: A parent and community guide to helping students learn*. Portland, OR: Educational Testing Service.
- Chappuis, S. Commodore, C, & Stiggins, R. (2010). *Assessment balance and quality: An action guide for school leaders* (third Ed.). Portland, OR: Pearson Assessment Training Institute.

- Clark, I. (2010). Formative assessment: There is nothing so practical as good theory. *Australian Journal of Education*, 54(3).
- Conderman, G. & Hedin, L. (2012). Classroom assessments that inform instruction. *Kappa Delta Pi Records*, 48(4), 162-168.
- Coughlan, D., & Brannick, T. (2005). *Doing action research in your organization* (2nd ed.). Thousand Oaks, CA: Sage.
- Dirkson, D. (2011). Hitting the reset button: Using formative assessment to guide instruction. *Phi Delta Kappan*, 92(7).
- Dorn, S. (2010). The political dilemmas of formative assessment. *Exceptional Children*, 76(3).
- Fraenkel, J., & Wallen, N. (2003). *How to design and evaluate research in education* (5th ed.). San Francisco: McGraw-Hill Higher Education.
- Freeman, V. & Hatley, R. (1981). *Relationships Between School District Level Inputs and the Output Performance of Students on the Missouri Basic Essential Skills Test*. Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).

- Gravetter, F. & Wallnau, L (2004). *Statistics for the Behavioral Sciences* (6th ed.). Belmont, CA: Thomson Wadsworth.
- Halverson, R. (2010). School formative feedback system. *Peabody Journal of Education* , 85(2), 130-46.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2).
- Kerlinger, F.N. *Foundations of behavioral research* (third ed.). New York: Holt, Rinehart & Winston.
- Madaus, G., & Russell, M. (2010/2011). Paradoxes of high-stakes testing. *Journal of Education*, 190(1/2).
- Marshall, C., & Gerstl-Pepin, C. (2005). *Re-Framing educational politics for social justice*. New York: Allyn & Bacon.
- Marzano, R. (2006). *Classroom assessment & grading that work*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Missouri Department of Elementary and Secondary Education (June, 1998). *Assessment Standards for Missouri Public Schools* (Word Document Available via the Web at dese.mo.gov/divimprove/assessment/assessmentstandards.pdf).

Missouri Department of Elementary and Secondary Education
(2012). *Missouri Assessment Program Grade-Level
Assessments: Guide to Interpreting Results*. Monterey,
CA: CTB/McGraw-Hill LLC.

Missouri Department of Elementary and Secondary Education
(2010). *Missouri Assessment Program: Grade Level
Assessments Technical Report, 2010*. Monterey, CA:
CTB/McGraw-Hill LLC.

Missouri Department of Elementary and Secondary Education
(2011). *Online End-of Course Assessments: Guide to
Interpreting Results 2011-12*. Brewster, NY: Questart
Assessment, Inc.

Moss, C.M. & Brookhart, S.M. (2012). *Learning targets:
Helping students aim for understanding in today's
lesson*. Alexandria, VA: Association for Supervision
and Curriculum Development.

National Assessment Governing Board (2002). The NAEP Law.
In *NAEP: Nation's Report Card*. Retrieved June 6, 2013,
from www.nagb.org/naep/naep-law.html

Patton, M. (1997). *Utilization-focused evaluation: The new
century text*. Thousand Oaks, CA: Sage.

Pearson. (2011). *K-8 Technical manual (Second Edition)*.
Saddle Ridge, NJ: Pearson Education, Inc.

- Pearson's r Correlation* (2013). Retrieved September 21, 2013 from *faculty.quinnpiac.edu/libarts/polsci/statistics.html*.
- Peerson, P.E. & West, M.R. (Eds.) (2003). *No Child Left Behind? The politics and practice of school accountability*. Washington, DC: Brookings Institution Press.
- Popham, W.J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, W.J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Preskill, H., & Torres, R. (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.
- Reeves, D. (Ed.) 2007. *Ahead of the curve: The power of assessment to transform teaching and learning*. Bloomington, IN: Solution Tree Press.
- Roskos, K. & Neuman, S. (2012) Formative assessment: Simply, no additives. *Reading Teacher*, 65(8), 534-538.

Shepard, L., Hammerness, K., Darling-Hammond, L., & Fust, F.

(2005). Assessment. In L. Darling-Hammond &

J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do.*

San Francisco: Jossey-Bass.

Stiggins, R. & Chappuis, S. (2005). Putting testing in perspective: It's for learning. *PL*, 17-20.

Stiggins, R. & DuFour, R. (2009). Maximizing the power of formative assessments. *Phi Delta Kappan*, 90(9).

U.S. Department of Education. (2004). *No child left behind: A toolkit for teachers.* Jessup, MD: Education Publications Center.

U.S. Department of Education, (2009). *Race to the top program.* Washington, D.C.: Author.

VITA

Ruth Ann Orr was born November 14, 1960 in Springfield, Missouri. She and her family moved to Puxico, Missouri, where she attended elementary and secondary schools until 1978. Dr. Orr completed her GED in July, 1978, and enrolled in Southeast Missouri State University the following August where she majored in Elementary Education, receiving a Bachelor of Science in Education in December, 1981.

Dr. Orr began her work career as an elementary teacher at Alma Schrader Elementary School (Cape Girardeau Public Schools) in 1983. She continued working at Alma Schrader while completing her Master of Arts degree from Southeast Missouri State University in Guidance and Counseling. She transferred jobs within the Cape Girardeau Public School district and worked as a counselor/psychological examiner in various elementary schools within the Cape Girardeau Public School district from 1988-1999. She again transferred jobs within the school district, beginning work as the Special Services Process Coordinator for the Cape Girardeau Public School Central Administrative Office from 1999-2002.

Dr. Orr completed requirements for her Administrative Certificate in 2001. In 2002, Dr. Orr was hired as the principal of Alma Schrader Elementary School where she is

currently employed. While working as the principal of Alma Schrader, Dr. Orr enrolled in the Educational Leadership and Policy Analysis cooperative program between Southeast Missouri State University and the University of Missouri. She completed all requirements for her doctorate in Educational Leadership and Policy Analysis in October, 2013.

Dr. Orr is married to Kevin Orr. They have three children: Heath, Meghan, and Elizabeth, and one grandchild, Aiden Galloway. She currently resides in Cape Girardeau, Missouri, and is the principal at Alma Schrader Elementary School.