Predicting performance on the Kansas Computerized Assessments from performance on the Measures of Academic Progress

By

Allison Kaye Folger

Submitted to the graduate degree program in the Department of Educational Research and Psychology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for degree of Master of Science in Education.

Chairperson Dr. Bruce Frey

Dr. Neal Kingston

Dr. William Skorupski

Date Defended: December 9, 2011

The Thesis Committee for Allison Kaye Folger

certifies that this is the approved version of the following thesis:

Predicting performance on the Kansas Computerized Assessments from performance on the Measures of Academic Progress

Chairperson Dr. Bruce Frey

Date approved: December 9, 2011

Abstract

As schools prepare their students for state standardized assessments, more and more schools are turning to other standardized assessments to help predict performance on state assessments. One such predictor assessment is the Measures of Academic Progress. This study looked at the predictive relationship existing between the Measures of Academic Progress and the Kansas Computerized Assessments and found that there is a statistically significant relationship between the two tests, but the strength of that relationship, as well as the validity of the two tests, raise some questions for further study.

Table of Contents

Chapter One

Introduction

Problem Statement

How will students perform on state standardized assessments? This is one of the most important questions teachers and school districts face as they move closer to the year 2014 when, according to federal No Child Left Behind legislation, all students should be proficient in the areas of reading and math. In 2002 when the No Child Left Behind (NCLB) legislation passed, education as most states and school districts new it changed. Before this time, states, school districts, and even teachers experienced a fair amount of autonomy in their teaching, and few states had standardized achievement tests to measure their students' achievement (U.S. Department of Education, 2008). As the Nation's Report Card continued to show a decline in the abilities of America's students, especially in the areas of math and reading, it was clear that changes in the public education sector needed to be made. NCLB established yearly benchmarks to measure student growth through the observation of standardized tests scores.

Since the passage of NCLB, many standards in education have changed. Now, all states have accountability plans and measure the academic progress of their students, and all states participate in the Nation's Report Card (U.S. Department of Education, 2008). In Kansas, the Kansas Computerized Assessments (KCA) were developed to measure the progress of elementary, middle, and high school students in the content areas. The effects of these changes are evident. Now that states have established consistent standards of what students should know and are testing students on the mastery of those standards, educators are beginning to see the achievement gap close. From 2003 to 2007, the achievement gap between white and Hispanic eighth grade students and between white and African-American eighth graders narrowed by three points in math, and all three demographics made gains. By 2007, reading and math scores for fourth graders rose, and math scores for eighth graders rose (U.S. Department of Education, 2008).

The final requirement of NCLB proscribes that all students should be performing at grade level in the content areas of math and reading by 2014. As that date quickly approaches, all school districts need to know that their students will be ready to achieve such high standards and are trying to determine how best to help their students attain them.

Schools use a variety of methods to determine where their students are performing before they take high stakes state assessments. One such method used is the Measures of Academic Progress (MAP) test. The MAP was published in 2003 and developed to “provide educators the information they need to improve teaching and learning in the areas of reading, language usage, mathematics, and science” (Cizek et. al., 2004). Because the MAP is aligned with the KCA, teachers should be able to use the results of the MAP to predict performance on the KCA and use the MAP results to guide instruction to improve performance on the KCA.

Because schools face pressure to meet the NCLB requirements, it is necessary to know if predictor assessments such as the MAP are truly reliable forecasters of performance on state standardized assessments. As educators, it is necessary to know what students do and do not understand in order to help them be successful test takers and students. If the MAP is strongly correlated with the Kansas Computerized Assessments and can reliably predict performance on the KCA then it is a useful tool that should continue to be used in the district, but if the MAP is not strongly correlated with the KCA, then perhaps district funds should be used to find a more useful and accurate predictor.

Currently, there is no research to show the relationship between the MAP and the KCA beyond that conducted by the Northwest Evaluation Association, the creator of the MAP. Research is needed to show if the MAP is a valid predictor and if school districts should continue using it.

Research questions

The following questions guided this research study:

Primary research question:

Does a predictive relationship exist between the seventh and eighth grade reading and math scores on the fall administration of the Measures of Academic Progress (MAP) and the seventh and eighth grade scores on the spring administration of the Kansas Computerized Assessments (KCA) in the content areas of reading and math during the 2008-2009, 2009-2010, and 2010-2011 school years?

Secondary research question:

Are the Measures of Academic Progress (MAP) for math, the Measures of Academic Progress (MAP) for reading, the Kansas Computerized Assessments (KCA) for math, and the Kansas Computerized Assessments (KCA) for reading for the 2008-2009, 2009-2010, and 2010-2011 school years valid assessments measuring what they are intended to measure?

Hypothesis

In regards to the primary research question, it was hypothesized that the Measures of Academic Progress would have a predictive relationship with the Kansas state reading and math assessments scores for seventh and eighth grade students during the specified years. It was also believed that the seventh and eighth graders' scores on the MAP test would be strongly correlated with scores on the Kansas reading and math assessments at the .05 alpha level because the MAP test and the Kansas assessments were aligned in 2007.

For the secondary research question, Campbell and Fiske's multi-trait, multi-method matrix was used (Campbell & Fiske, 1959), adding in the extra component of multi-year. In the multi-trait, multi-method, multi-year matrix, it would be expected that the relationship between the same method measuring the same trait from year to year (the 2009 KCA for math as compared to the 2010 KCA for math, for example) would be the strongest. Comparisons of the same trait using different methods (such as the 2008 MAP for math and the 2009 KCA for math) would be expected to have a medium relationship, while the divergent validity measures of different traits using the same method (comparing the math MAP scores to the reading MAP scores), as well as different traits using different methods (the comparison of math MAP to reading KCA scores and math KCA scores to reading MAP scores) would be expected to have a low correlation. Conversely, it would be expected that the different traits using the same method would have a slightly stronger relationship than the comparison of different traits using different methods since different traits measured by different methods should have very little connection.

Predicted results for the secondary research question are represented in the following table:

Figure 1

Multi-trait, multi-method, multi-year matrix of predicted correlations for MAP and KCA scores

			2008-2009				2009-2010				2010-2011			
			Math		Reading		Math		Reading		Math		Reading	
			MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA
2011	Reading	KCA	low	low	medium	high	low	low	medium	high	low	low	medium	X
2011	Reading	MAP	low	low	high	medium	low	low	high	medium	low	low	X	
2010	Math	KCA	medium	high	low	low	medium	high	low	low	medium	X		
2010	Math	MAP	high	medium	low	low	high	medium	low	low	X			
2010	Reading	KCA	low	low	medium	high	low	low	medium	X				
2010	Reading	MAP	low	low	high	medium	low	low	X					
2009	Math	KCA	medium	high	low	low	medium	X						
2009	Math	MAP	high	medium	low	low	X							
2009	Reading	KCA	low	low	medium	X								
2009	Reading	MAP	low	low	X									
2008	Math	KCA	medium	X										
2008	Math	MAP	X											

Summary

Because the No Child Left Behind legislation demands that all students reach proficiency in the areas of math and reading by the year 2014, schools have begun to use standardized assessments such as the Measures of Academic Progress to predict performance on state standardized assessments like the Kansas Computerized Assessments and to guide instruction to prepare students for such standardized assessments. Teachers and schools need to know that the MAP assessment is a reliable predictor of performance on the KCA. Study of the correlation between and predictive value of the MAP test on the KCA is necessary.

Chapter Two

Review of Literature

Kansas Computerized Assessments (KCA)

In the spring of 2003, Kansas began using the Kansas Computerized Assessments (KCA) to replace previous standardized testing practices and to begin to meet the requirements of the No Child Left Behind (NCLB) legislation. The test was then revamped in 2006 to reflect its current design and administration. The Kansas Department of Education (KSDE) contracted with Wested to create the testing items based on KSDE qualifications and the Center for Educational Testing and Evaluation at the University of Kansas for all other components of the test design (Irwin, Poggio, et al., 2007). The KCA are given in reading, math, science, and social studies.

The reading and math KCA are designed for all students to take in grades three through eight and grade 10 in math and 11 in reading. An early reading diagnostic assessment for kindergarten through second grade is also available (Kansas Department of Education, 2010c). Separate, modified assessments for students with disabilities also exist. The Kansas Assessment of Modified Measures (KAMM) can be given to students with moderate disabilities to measure specific outcomes. The Kansas Alternative Assessment (KAA) is an assessment based on the Individualized Educational Program (IEP) goals of students with severe disabilities and set Kansas standards (Irwin, Poggio, et. al., 2007). A Spanish version of the math assessment is also available for non-native English speakers, and a static, paper-pencil version of the assessments also exists for use as an accommodation (Kansas Department of Education, 2010b). Students can be exempted from taking the tests if they move into a district within certain time periods prior to the testing window, but other than those exemptions, all students are expected to take some form of the Kansas assessments.

Questions for the KCA were created by Wested and were derived from the Kansas Curricular Standards of what students in Kansas should know in the content areas at certain grade levels (Irwin, Poggio, et. al., 2007). Test items are divided into benchmarks and indicators. On the seventh grade math assessment there are 84 questions involving 15 different indicators with four to eight questions per indicator. The eighth grade math assessment includes 86 questions covering 15 different indicators (Kansas Department of Education, 2010b). The reading assessment ranges from 58 questions at the third grade level to 84 questions at the seventh grade level with four to six questions covering each of the eleven to sixteen indicators per grade level (Kansas Department of Education, 2010c). Those questions are arranged in a multiple choice format with four response options per question with one correct answer. Five parallel forms of the math assessment exist at each grade level, and three to four parallel forms of the reading assessment exist at each grade level. Reading text selections exist in the areas of narrative, expository, technical, and persuasive and are designed for the appropriate grade level (Irwin, Poggio, et. al., 2007).

The reading and math assessments are administered in several testing sessions, usually over several days, within the testing window. The reading and math KCA are both given in three untimed testing sessions (about 45 minutes each for reading (Kansas Department of Education, 2010c) and 45-60 minutes each for math). Calculators can be used on two sections of the math assessment (Kansas Department of Education, 2010b), scratch paper and pencils, and a number of other math manipulatives can also be used on the math assessments.

Scores for the KCAs are reported in a number of different ways. Student performance level scores are reported in five categories, or performance levels. Those performance levels are: Academic Warning, Approaches Standard, Meets Standard, Exceeds Standard, and Exemplary. These levels are determined by cut scores. Cut scores are the percentage of correct answers on each assessment; they are designed as part of NCLB and are not set by districts, nor are they comparable to letter grades (A,B, C, etc) (Kansas Department of Education, n.d.). The cut scores for Kansas, as approved by the Kansas Board of Education, for the seventh grade reading KCA are: 0-49% correct equals Academic Warning, 50-62% correct equals Approaches Standard, 63-76% correct equals Meets Standard, 77-86% correct equals Exceeds Standard, and 87-100% equals Exemplary. For the eighth grade reading assessment, those cut scores are: 0-49% correct means Academic Warning, 50-63% means Approaches Standard, 64-78% means Meets Standard, 79-88% means Exceeds Standard, and 89-100% means Exemplary (Irwin, Poggio, et. al., 2007).

For math, the same standards were used to create cut scores. The seventh grade math cut scores are: 0-43% correct equals Academic Warning, 44-55% correct equals Approaches Standard, 56-70% correct equals Meets Standard, 71-83% correct equals Exceeds Standard, and 84-100% correct equals Exemplary. Eighth grade assessment cut scores are slightly different: 0-44% correct means Academic Warning, 45-57% means Approaches Standard, 58-72% means Meets Standard, 73-85% Exceeds Standard, and 86-100% means Exemplary (Irwin, Poggio, et. al., 2007).

In addition to performance level scores and percent correct scores (cut scores), KCA results are also reported by indicator. Students getting correct answers for each indicator are reported on the individual student level, the building level, district level, and state level (Kansas Department of Education, 2010b).

There are several intended purposes of the Kansas Computerized Assessments. One is to help with determining school accreditation. The results of the test are also used to verify if a school makes Adequate Yearly Progress (AYP) as part of the NCLB legislation. AYP is determined through the use of performance levels and cut scores. To meet the NCLB legislation mandates, Kansas determined certain percentages of students within subgroups, schools, and districts that must meet the performance level of Meets Standard on the reading and math KCA in order to ensure that 100% of students perform at that level by 2014. For 2010, the AYP target for elementary schools in reading was 83.7% of students meeting the standard and 81.3% of students for the district. In math, the 2010 AYP targets were 82.3% for schools and 76.4% for districts. (Kansas Department of Education, 2010a). For 2011, 87.8% (Kansas Department of Education, 2010c) and 86.7% of students in grades three through eight needed to perform at the Meets Standard performance level in reading and math, respectively, in order to meet AYP (Kansas Department of Education, 2010b). In 2006-2007, 89% of public schools in Kansas made AYP, which was up from 84% in 2005-2006 (Denny, 2007).

The revised KCA went through several evaluations prior to and after its Spring 2006 administration to ensure its validity and reliability. All testing items went through content and fairness review committees who determined if each item met the Kansas Curricular Standards and other content specifications. Questions were also examined to ensure that all content was appropriate, not offensive or insensitive, and that no groups were unfairly advantaged or disadvantaged due to item wording. For math, items were grouped by benchmark and indicator tested and ranked, based on difficulty, from low to high. Reading test passages were examined for readability, word count, and indicator tested. Differential Item Functioning (DIF) using the Mantel-Haenszel technique was conducted to determine how different genders and ethnic groups responded to each item. An equity review committee composed of minority teachers in Kansas was also formed to look at any bias the KCA might pose to members of various ethnic groups (Irwin, Poggio, et. al., 2007).

The reliability of the KCA was also reviewed in the summer of 2006. The Cronbach coefficient values for reading across grade levels and forms ranged from .88 to .95, while ranging from .91 to .95 for math. The overall standard error of the measurement on percent correct score scale ranged from 3.65 to 4.70 for reading and from 3.95 to 4.60 for math. Classification consistency ("the extent to which the classifications agree on the basis of two independent administrations of the test" (Irwin, Poggio, et. al., 2007)) and classification accuracy ("the extent to which the actual classifications that are based on observed cut scores approximate those that are based on 'true' cut scores" (Irwin, Poggio, et. al., 2007) were also determined to help judge the reliability of the KCA. For math, classification consistency values ranged from .59 for fifth grade to .72 for tenth grade, while classification accuracy was consistently higher ranging from .69 for fifth grade to .80 for tenth grade. Classification consistency for reading ranged from .54 for fourth grade to .74 for eleventh grade, while classification accuracy ranged higher than consistency with the lowest coefficient value being .87 (Irwin, Poggio, et. al., 2007).

It was also necessary to ensure that the test was actually measuring what the students knew and that the results could be used to make appropriate inferences, generalizations and decisions. To test the validity of the KCA, test structure was evaluated to verify the dimensionality of the test and identify the constructs measured by the items. This was performed using TESTFACT 4.0. The inter-item correlation for math ranged from .311 to .451, increasing as grade level increased. In reading, the inter-item correlation ranged from .336 to .385 and did not demonstrate evidence of increasing with grade level. Both the reading and math tests demonstrated strong evidence of unidemensionality (Irwin, Poggio, et. al., 2007).

Validity was also tested by comparing KCA formative assessment results with those of the Spring 2008 summative KCA results. For math, correlations for all forms of the assessment across all grades ranged from .71 to .87. Correlations for all forms across all grades for reading ranged from .76 to .83. The coefficient values demonstrated a high predictive value of the formative assessment on the summative assessment. Teacher ratings of student performance were also used to ensure that the KCA was indeed a good representation of what students knew. The median coefficient value across grade levels of .67 for reading and .62 for math demonstrated that there was a strong positive relationship between teacher ratings and students' actual performance on the KCA (Irwin, Poggio, et. al., 2007).

There are several benefits to using the KCA system for standardized testing in Kansas. Because they are online, they are scored quickly and provide teachers and administrators with immediate feedback. The fact that the tests are no longer taken with paper and pencil cuts down on shipping costs since it is no longer necessary to mail in tests, and it improves test security because tests do not have to be kept at the school. Security on the tests is also improved by the fact that there is a testing window in which all exams must be completed, and each student has a testing ticket that gives him/her access to only the assessment being taken at that time (Irwin, Poggio, et. al., 2007). There are also load managing software and redundant backup systems to help with the security of the data. The reactivation options that allow student to start testing, stop testing, and resume testing at a later time are also positives of the online administration of the test (Peyton, 2008a). The accuracy of the scoring of the tests has also been improved. The use of the online formative tests available in each of the tested areas is also beneficial for teachers and students (Irwin, Poggio, et. al., 2007). Teachers can use the results of the formative assessments to guide instruction throughout the year in order to improve students' skills in areas of need, and students can use the formative assessments to acclimate themselves to the test design and mechanics. Students are also potentially more motivated to take the assessments online as opposed to the traditional paper-pencil format (Peyton, 2008a).

In a survey published in 2008 of 166 teachers representing 50 different Kansas school districts, at least 61% of respondents rated the KCA as being of high quality. No participants rated the KCA as being of poor quality. About 70% of respondents believed that students preferred to be tested online. Over half of survey participants found the KCA to have many positives that outweighed any of its negatives. Ninety percent of survey participants thought the feedback from the formative and summative tests was useful, and 85% said that they actually did alter their instruction after looking at the results of the assessment (Peyton, 2008b).

Some of the concerns with the KCA online format were voiced in a study conducted in 2006 of 292 Kansas standard setting participants. A majority of participants (78%) expressed that the use of computers for testing did influence their instructional time. Ninety percent of respondents believed that the KCA had a moderate to significant impact on their building. Drawbacks of the KCA in the opinions of survey participants were that the number of students often was larger than the number of computers available causing limited availability of computers for students, and that the increased use of computers for formative and summative assessment "diminished the bond between teacher and student" (Peyton, 2008a). A fear that students would begin to associate computers with testing was also expressed by teachers (Peyton, 2008a).

Despite any of the perceived negatives of the KCA, it is a valid and reliable testing system that provides teachers in the state of Kansas with immediate feedback as to how to best help their students reach the goal of meeting the standard of proficiency set by the NCLB legislation in math and reading by 2014.

Measures of Academic Progress (MAP)

The Measures of Academic Progress (MAP) is a computerized, adaptive standardized assessment developed by the Northwest Evaluation Association (NWEA). It is used by more than ten percent of K-12 school districts nationwide and one-third of districts in the Midwest to test the subject areas of reading, math, language usage, and science.

The test is intended for students in grades three through ten. There is also a diagnostic test that can be given at the second grade level (Cizek et. al, 2004). The MAP is designed to give teachers immediate feedback as to the skill levels of their students so that teachers can adapt instruction to help meet student needs. The test aligns with state standardized assessments so that results can also be used to predict performance on state tests. In five states the MAP has been determined to be a sufficient predictor of performance on state assessments (U.S. Department of Education, 2009). In 2007, NWEA correlated the MAP RIT scale with the scale used for the Kansas assessments so that the RIT scale could be used to determine a good chance for success on the Kansas assessments. Teachers can use this information to tailor instruction to meet the needs of students and provide extra assistance to those with lower RIT scores who will need extra help in preparing for the state assessments.

To administer the test, school districts must purchase a license for the test and then complete several training sessions. MAP training stretches across four days and covers a variety of topics. One day is devoted to instructing teachers and administrators how to administer the test. Another day revolves around understanding how to generate and interpret the outcome reports for the test. Learning how to use the data to set growth goals for students and evaluate the instructional programs and practices of the teachers comprises another day of training. The last day covers how to use those reports to differentiate instruction to meet the needs of all students. Schools may set up additional training days; conference calls and on-site visits may also be conducted for schools during their year of MAP implementation (U.S. Department of Education, 2009).

The MAP is divided into five subtests: reading, language usage, math, science, and general science that are tested using multiple choice questions. Those subtests are divided into subcategories. For reading, those subcategories are word meaning, literal comprehension, interpretive comprehension, and evaluative comprehension. Writing process, composition structure, grammar and usage, punctuation, and capitalization comprise the subcategories of the language usage test. The subcategories for the math subtest include operations and computation, equations and numerals, geometry, measurement, problem solving, statistics and probability, and applications. Concepts and processes make up the science subtest, while the general science subtest consists of life science, earth/space science, and physical science. Each subcategory has at least seven questions, making each subtest 40 to 50 questions long. Both the reading and language usage tests have four answer options. The math test has five answer options. Questions for the test are taken from a large pool of items (1500 questions for math, 1200 questions each for reading and language usage). Because there is such a large pool of questions to pull from, test items will not be administered to a student two years in a row (Cizek et. al., 2004).

Once districts are ready to administer the test, they must have a wired computer lab available as the test cannot be taken wirelessly (A static, paper/pencil version of the test called the Achievement Level Test (ALT) may also be used.). The MAP is untimed and adaptive so that each student receives testing items optimal for his or her ability level. The first question of the test is given at "five RITs below student's current ability as estimated from previously administered MAP in that subject area" (Cizek et. al., 2004). If no previous MAP data is available for a student, the test begins with a question that is five RITs below the mean. Performance on the first item provides an improved estimate of ability, and items are then chosen to fit student needs in order to find "as precise an estimate as possible of student ability levels" (Cizek et. al., 2004). If the test is completed in less than six minutes, its results are considered invalid (Cizek et. at., 2004). It is recommended that the test be taken three times a year (fall, winter, spring) with an optional summer test (US Department of Education, 2009).

MAP scores are conveyed in a number of different forms in order to make them meaningful for all stakeholders. Scores are reported in Rasch units (RIT scores). Rasch units use a continuous interval scale (US Department of Education, 2009) ranging from 140 to 300 and are the transformation of the Rasch ability estimates (Cizek et. al., 2004). Scores are based on normed data that was gathered from the spring of 2001 to the fall of 2004 and involved 2.3 million students from 5616 different schools in 794 districts in 32 states. Test results are reported in individual class and system reports. Each student's performance in each subcategory is summarized in standardized reference terms using scale scores, percentage correct scores, achievement categories (low, medium, high). Status norms (how the student performs compared to the group) and growth norms (student pre-test to post-test gains) are also reported for each subtest and subcategory (Cizek et. al., 2004).

Reliability and validity evidence can be used to support the utility of the MAP for its intended purpose. Internal reliability for total scores on the MAP range from .92 to .96 according to data estimates from varying numbers of students across grade levels collected from a 1999 MAP study. A study of the test-retest reliability of the MAP for fall-to-spring or spring-to-spring scores collected in 2002 produced stability estimates ranging from .77 to .94 and showed a general increase in stability across grade levels. According to Cizek et.al. (2004), "comprehensive test development and administration procedures and documentation support conclusions that MAP scores differentiate between students' level of ability in tested subjects." NWEA further demonstrates validity of the MAP assessment by correlating scores from the ALT reading, math, and language usage with scores on the Iowa Test of Basic Skills and the Stanford Achievement Test 9th edition. Correlations range from .77 to .84 across all grade levels. ALT MAP scores are also correlated with state level assessments in Wyoming, Colorado, Indiana, and Washington (Cizek et.al., 2004).

There are many advantages to using the MAP as a predictive standardized assessment. Teachers do not have access to test items, so they cannot teach to the test, thus it is a good snapshot of student ability levels. MAP test items align with state content standards (U.S. Department of Education, 2009), and scores are reported in a timely fashion (Shields, 2008), allowing teachers to make instant adaptations to their lessons to meet the needs of their students and prepare them for state standardized tests. Teachers can also use the scores to measure student growth over time such as from fall-to-spring or spring-to-spring to see what gains students have made. Because the test is reasonably short, it increases testing efficiency and only minimally interrupts valuable instructional time. Since the test is adaptive to student needs, it allows for more precise testing, helps students feel appropriately challenged, and keeps students from getting bored during testing time (Cizek et. al., 2004).

Despite these positives, there are some disadvantages to the MAP. Because the test is adaptive, students cannot skip questions, nor can they go back to review or change previous answers to questions. The test is computerized and untimed, which could cause some scheduling problems as different classes try to get into school computer labs to test, and students finish testing at different rates. Some schools also may not have enough testing stations to accommodate all students at one time. The MAP testing manual provides no suggestions for how to handle the scheduling of computers for test administration as it does not include any information pertaining to technology requirements. MAP subtests also have some inconsistencies. Some tests use letter options while others use number options. During the math test, the calculator pops up on screen for some questions where it is clearly unnecessary to use, but not for some questions where it seems like the calculator might be helpful (Cizek et. al., 2004). This could be confusing or distracting for some test takers.

There are also some problems with evaluating the test design and reliability and validity studies. Most of those problems revolve around the fact that very little information is given about how they were determined. For example, beyond the number of questions NWEA created for each subtest, "information is not provided on specific curricular guides, state content standards, professional association recommendations, or textbook representativeness of the item writers" (Cizek et.al., 2004). Additionally, minimal information regarding the demographics or other defining characteristics of the norm group is given, making it difficult for test administers to determine how appropriate the norm group is as a comparison group. For validity testing, the ALT MAP assessment was used, and while the ALT uses similar questions as the MAP, it is not the same test, and consequently should probably not be used to support the validity of the MAP. More descriptive information is also needed regarding the groups used to test reliability and validity since NWEA does not provide that information (Cizek et. al., 2004).

Another problem with the MAP is that studies investigating the effects of the MAP on student outcomes are actually scarce. There is plenty of research on formative assessments and student outcomes but very little on benchmark assessments, which is what the MAP test is. Because of this, formative assessment data is often cited to support the use of the MAP test despite the fact that formative assessment data is sometimes misleading. Many studies involving formative assessments "have design constraints, such as nonrandom samples and nonequivalent sample groups, that undermine their validity" (U.S. Department of Education, 2009). Most formative assessment practices that have been reviewed have focused on classroom-based assessment practices that are used more frequently than the MAP, and thus would not be effective for comparisons with the MAP assessment. The Regional Educational Lab Northeast and Islands conducted two different studies on the impact of benchmark assessments more similar to the MAP on student outcomes and found no significant differences in math achievement gains between schools that use quarterly benchmarks and those that do not (U.S. Department of Education, 2009). This would mean that, at best, further study of the MAP test, specifically on student achievement gains, is needed.

In conclusion, the MAP is a fairly easily administered assessment that can be used to predict how students will perform on state standardized assessments, guide instruction, and show growth over time. However, further study is necessary to help support the effectiveness of the MAP as a viable predictor of performance on higher order tests. NWEA should also provide better documentation of the demographics of norm groups and groups used for reliability testing, the processes used to write and select questions, and the reasons why the ALT is a worthy substitute for the MAP in validity tests.

Comparisons of MAP to state standardized assessments

The Northwest Evaluation Association (NWEA) spent time aligning the Kansas state assessments to the MAP test so that the MAP test could be used to predict performance on the KCA. For example, in Kansas 95% of seventh grade students who obtain a RIT score of 240 on the spring MAP for reading should pass the Kansas state reading assessment that same spring. Conversely, of those seventh grade students scoring a 165 on the spring MAP math test, only one percent should pass the Kansas state reading assessment that spring. For math, 99% of seventh graders should pass the spring Kansas math assessment if they attain a 270 RIT score that same spring. On the other hand, those with spring MAP RIT scores of 170 are not predicted to pass the Kansas math assessment (Northwest Evaluation Association, 2009). The two assessments were aligned in 2007 and realigned in 2009. To align the two tests, NWEA identified specific Rasch Units (RIT scores) scale scores from the MAP test and linked them to the proficiency levels for each state test for each grade level. They also estimated the probability that a student with the specific RIT score would achieve at the "proficiency" level or better on state assessments. NWEA employed an "Equipercentile Method," taking students from Kansas who had both MAP test and state assessment scores and using that data to generate cut score estimates and pass/fail predictions for the Kansas assessment. This was done to determine cut scores despite the fact that the two assessments are measured on two different scales (Northwest Evaluation Association, n.d.). While NWEA provides some information about the process their researchers used to align the MAP and the KCA, they do not provide any information about the number of or the demographics of the schools in Kansas who take the MAP and thus served as the testing population. That information is missing from the NWEA official website and was not provided to this researcher despite a request for detailed information about the testing population (personal communication, March 2011).

While no studies looking at the predictive value of the MAP on the KCA for specific schools or districts could be found, there are some studies linking the MAP to other state standardized assessments. One particular study looked at the MAP and the Tennessee standardized assessment system and found that there was a strong positive relationship between the MAP and the Tennessee Comprehensive Assessment Program (TCAP) for fourth and fifth graders in the areas of language arts and math. An $R=.727$ and $R= .669$ showed strong linear relationships between the MAP and the TCAP for language arts and math, respectively. For language arts, $r = .758$ also meant a strong correlation between predicted and observed values on the TCAP. An $R^2 = .541$ meant that 54% of the TCAP scores for language arts could be explained by MAP scores. In math, $r = .720$ showed a strong correlation between observed and predicted TCAP scores, and an R^2 of .448 meant that 45% of the variance in the TCAP scores for math could be explained by MAP math test scores (Nugent, 2009).

A study of the MAP and the Missouri Measures of Academic Performance (MoMAP) also found a strong correlation between the two tests for sixth, seventh, and eighth graders despite the fact that the MoMAP is a higher order test taken in a paper-pencil format that requires test takers to write out answers and provide reasoning (Shields, 2008).

Because there are no comparisons of MAP and KCA outside of the alignment process through NWEA, a more in depth look at the relationship between them is necessary.

Summary

The Kansas Computerized Assessments (KCA) are administered every spring in Kansas to determine how students perform in the areas of math and reading in order to meet the federal No Child Left Behind requirements. The Measures of Academic Progress (MAP) is designed to be administered several times in a school year to help teachers determine their students' strengths and weaknesses and predict performance on state assessments. The MAP and the KCA have been aligned so that teachers can use MAP scores to guide instruction and predict performance on the KCA. While there have been studies confirming the predictive value of the MAP on other state standardized assessments, very little evidence exists to support those claims in Kansas, and little information is provided as to how the MAP and KCA were aligned.

Chapter Three

Methods

Participants

Approval for this study was obtained from the Human Subjects Committee at the University of Kansas prior to the initiation of the research because the necessary data was collected from minors. This data was provided from the district and all students were deidentified.

A group of seventh and eighth grade students from one smaller sized school in a rural district in Kansas was used for this study. Seventh and eighth grade students, numbering approximately 175 between the two grade levels, composed the participants in this study. The grades were chosen together in order to have a large enough sample size. Students in the group took the MAP twice yearly, once in the fall and once in the spring, and took the math and reading KCA yearly in either March or April over a three year period. The study was approved by the building principal who provided necessary data for the study.

For the 2008-2009 and 2009-2010 school years, the building had an enrollment of approximately 340 students in a district of a little over 1,100 students. During the 2008-2009 school year, the building make-up was 94.4% white, 2.6% Hispanic, 2.3% other, and .6% African American. Males made up 51.3% of the school population and girls 48.7%. Forty-five percent of the population was considered economically disadvantaged, qualifying for free and reduced lunches. Students with disabilities comprised 23.8% of the student population (Kansas Department of Education, 2009). The district had a graduation rate of 93% in 2009, which was slightly above the state average of 89% (Kansas Department of Education, 2010d).

Between the 2008-2009 and 2009-2010 school years, the district went through a building reorganization changing from three kindergarten through eighth grade buildings to a kindergarten through second grade building, a third and fourth grade building, and a fifth through eighth grade building. While this consolidation did change the make-up of buildings, it did not change MAP or KCA testing procedures. The reorganization occurred at the beginning of the school year, well before the MAP or KCA were administered, and thus did not affect the data of this study.

During the 2009-2010 school year, 90.6% of students were white, with the rest of the demographic make-up of the school composed of African Americans (1.2%), Hispanics (4.1%), and other (4.1%). Males made up 54.4% of the population and girls 45.6%. Forty-two percent of the building population was considered economically disadvantaged. Students with disabilities comprised 14.3% of the student population, and .3 percent were ELL students. The building had an attendance rate that held steady from year to year at about 95% (Kansas Department of Education, 2010d).

One-hundred percent of the teachers in the building in which the study was conducted were considered fully licensed in the 2008-2009 school year, while only 90% of the teachers were considered fully licensed during the 2009-2010 school year. During the 2009-2010 school year, 100% of core classes in the selected school were taught by highly qualified teachers in the areas of elementary self-contained classes, English language and literature, fine and performing arts, mathematics, and social sciences and history (Kansas Department of Education, 2010d).

On the 2010 Reading KCA, 39.6% of the current seventh graders (who make up the eighth grade population of this study) performed at the Exemplary level and zero percent were at the Academic Warning level; 35.4% of those students performed at the Exemplary level on the Math KCA and one percent performed at the Academic Warning level. Exemplary performances of those seventh graders on both the reading and math KCA were higher than the state averages of 37% and 26.8%, respectively (Kansas Department of Education, 2010d). The previous year as sixth graders, 42.9% of them scored at the Exemplary level on both the Reading and Math KCA, and zero percent at the Academic Warning level. Again, those Exemplary performances were higher than the state averages of 33.3% for reading and 30.7% for math (Kansas Department of Education, 2009).

Instruments

The Measures of Academic Progress (MAP) and the Kansas Computerized Assessments (KCA) were selected as the instruments of this study. A detailed explanation of psychometric data for the MAP and KCA was presented earlier in the Review of Literature section. Both measures have been shown to be valid and reliable measures of student performance. The internal reliability for total scores on the MAP ranged from .92 to .96 according to data estimates from varying numbers of students across grade levels from a 1999 MAP study (Cizek, et. al., 2004). The reliability of the KCA was tested in 2006 with the Cronbach coefficient values for reading across grade levels and forms ranging from .88 to .95 and from .91 to .95 for math (Irwin, Poggio, et. al., 2007).

Data from the fall 2008, 2009, and 2010 administrations of the reading and math MAP tests was entered into the Statistical Package for the Social Sciences (SPSS) and compared to the spring 2009, 2010, and 2011 results of the reading and math KCA.

Data Analysis

For this study, existing correlations between the seventh and eighth grade reading and math scores on the fall administration of the Measures of Academic Progress (MAP) assessment and the seventh and eighth grade scores on the spring administration of the Kansas Computerized Assessments (KCA) in the content areas of reading and math during the 2008-2009, 2009-2010, and 2010-2011 school years were examined. To do this, a zero-order correlation was used to relate one predictor variable (MAP reading and math scores) with one criterion variable (KCA reading and math scores).

Power, according to Keith (2006), "refers to the ability [to] correctly reject a false null hypothesis. It is a function of the magnitude of the effect, the alpha, or probability level chosen for statistical significance; and the sample size used in the research" (pg, 202). In addition, Keith also states that "common values for power are .8 or .9, meaning that given a particular effect size, one would like to have an 80% or 90% chance of rejecting a false null hypothesis of no effect" (pg.202). G power was used to determine how large a sample size was needed for the relationship to be considered statistically significant. Given a two-tailed test, an effect size of .3, and an alpha level of .05, a sample of 134 would yield a power of .95. The actual sample size for this study ranged from 158 to 171.

For the secondary research question determining the validity of the tests, the multi-trait, multi-method, multi-year matrix was used in order to use some reliability and validity evidence to determine the overall validity of the scores from the two methods (MAP and KCA). As Campbell and Fiske stated (1959), "in order to examine discriminate validity, and in order to estimate the relative contributions of trait and method variance, more than one trait as well as more than one method must be employed in the validation process…. This can be accomplished through a multitrait-multimethod matrix" (Campbell & Fiske, 1959). Using this we can see the "intercorrelations resulting when each of several traits is measured by each of several methods" (Campbell & Fiske, 1959).

Positive Results

If the research showed a positive relationship between performance on the Measures of Academic Progress (MAP) and the Kansas Computerized Assessments (KCA), then the null hypothesis that there is no relationship between MAP and KCA scores would be rejected, and the hypothesis that a predictive relationship exists between seventh and eighth grade reading and math scores on the fall administration of MAP and the seventh and eighth grade scores on the spring administration of the KCA for reading and math during the 2008-2009, 2009-2010, and 2010-2011 school years would be accepted. It could then be concluded that the Measures of Academic Progress could be used as a good predictor of performance on the Kansas Computerized Assessments. The school district used in this study would then be better prepared to determine which students need more effort and resources devoted to them in order to prepare them for the KCA.

Negative Results

If the research suggested that there was a negative relationship between performance on the Measures of Academic Progress (MAP) and the Kansas Computerized Assessments (KCA), then the null hypothesis that there was no relationship between MAP and KCA scores would be accepted, and the hypothesis that a predictive relationship existed between seventh and eighth grade reading and math scores on the fall administration of MAP and the seventh and eighth grade scores on the spring administration of the KCA for reading and math during the 2008-2009, 2009-2010, and 2010-2011 school years would be rejected. In this case, the school district should then consider reallocating funds from the Measures of Academic Progress to something that is a better predictor of Kansas Computerized Assessment performance.

Summary

Test scores from a convenience sample of seventh and eighth grade students at a small, rural, Midwestern school were used to determine the correlation between the Measures of Academic Progress and Kansas Computerized Assessments testing instruments. Any positive results of this study would help this particular school and district determine if there is a relationship between performance on the Measures of Academic Progress and the Kansas Computerized Assessments for their students.

Chapter Four

Results

The data from the fall 2008, fall 2009, and fall 2010 administrations of the Measures of Academic Progress (MAP) and the spring 2009, spring 2010, and spring 2011 administrations of the Kansas Computerized Assessments (KCA) were entered into the Statistical Package for the Social Sciences (SPSS) and appropriate analyses were conducted. Results from those analyses are presented in this section.

Presented in Table 2 are the descriptive statistics from the 2008, 2009, and 2010 fall administrations of the Measures of Academic Progress (MAP) for math and reading for all seventh and eighth grade students, and the 2009, 2010, and 2011 spring administrations of the Kansas Computerized Assessments (KCA) for math and reading. The total number of students completing the assessment, the range of scores, the minimum and maximum scores obtained on the measure, as well as the standard deviation are detailed. Scores for the Measures of Academic Progress are reported in Rasch Units (RIT scores) that can range from 140 to 300. Scores for the Kansas Computerized Assessment are reported in cut scores (the percentage of correct answers on each assessment) and can range from 0 to 100.

Table 2

Descriptive Statistics for MAP and KCA

Test	N	Range	Minimum	Maximum	Mean	Standard Deviation
MAP Math 2008	159	74	184	258	219.77	12.25
MAP Math 2009	160	68	191	259	226.96	12.58
MAP Math 2010	172	73	196	269	232.01	13.43
MAP Reading 2008	159	59	179	238	213.99	10.27
MAP Reading 2009	160	70	178	248	217.67	11.92
MAP Reading 2010	171	70	179	249	221.47	12.67
KCA Math 2009	160	54	45	99	82.16	12.40
KCA Math 2010	164	75	24	99	77.41	13.13
KCA Math 2011	178	64	36	100	72.16	15.10
KCA Reading 2009	159	52	46	98	83.01	9.73
KCA Reading 2010	164	47	50	97	81.83	9.24
KCA Reading 2011	176	68	31	99	80.30	11.42

Results for Research Questions

The following questions guided this research:

1) Does a predictive relationship exist between the seventh and eighth grade reading and math scores on the fall administration of the Measures of Academic Progress (MAP) assessment and the seventh and eighth grade scores on the spring administration of the Kansas Computerized Assessments (KCA) in the content areas of reading and math during the 2008-2009, 2009-2010, and 2010-2011 school years?
2) Are the Measures of Academic Progress (MAP) for math, the Measures of Academic Progress (MAP) for reading, the Kansas Computerized Assessments (KCA) for math, and the Kansas Computerized Assessments (KCA) for reading for the 2008-2009, 2009-2010, and 2010-2011 school years valid assessments measuring what they are intended to measure?

Research question #1

To determine if the Measures of Progress was a good predictor of performance on the Kansas Computerized Assessments for the 2008-2009, 2009-2010, and 2010-2011 school years, zero-order correlations were conducted for the two tests for the appropriate school years. Tables 3 through 8 present the correlational data for either the math or reading Measures of Academic Progress (MAP) and the corresponding Kansas Computerized Assessments (KCA) for the specified school year. Presented are the number of students completing both assessments, the correlation between the MAP and KCA using the Pearson Correlation Coefficient (r), the proportion of variance the KCA shares with the predictor test (the MAP) (R^2), the significance level, the unstandardized coefficient (b), and the standardized coefficient (β). Following each table, Figures 1 through 6 depict the scatterplots of test scores for the Measures of Academic Progress and the Kansas Computerized Assessment for the specified school years. The scatterplots depict how the scores on the two assessments fit around the regression line.

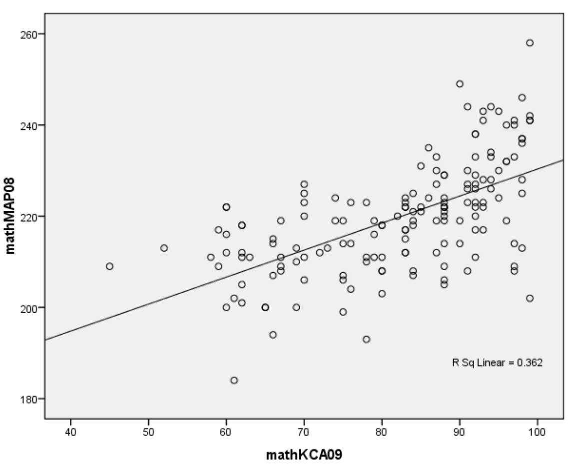
Table 3

Correlation between 2008 math MAP and 2009 math KCA

Tests	N	r	R^2	Sig.	b	β
MAP Math 2008	159	.60	.36	.00	.61	.60
KCA MAP 2009						

Figure 1

Scatterplot of MAP and KCA scores for math for 2008-2009



Results indicate a moderate correlation between the two assessments at .60 ($r=.60$, $p<0.001$). This demonstrates that the 2008 MAP for math was a moderate predictor of performance on the 2009 KCA for math. The R^2 value of .36 means that the overall model explains about 36% of the variance in student performance on the 2009 math KCA. The unstandardized coefficient of .61 means that for every point a student's score increased on the 2008 math MAP, his score on the 2009 math KCA should have increased by .61 points. The standardized coefficient of .60 shows that a student's score on the KCA should increase by .60 standard deviations if the student's score on the MAP increased by one standard deviation. Figure 1 also supports that the MAP was only a moderate predictor of performance on the KCA because there are many points not lying along the regression line.

Table 4

Correlation between 2008 reading MAP and 2009 reading KCA

Tests	N	r	R^2	Sig.	b	β
MAP Reading 2008	158	.54	.29	.00	.53	.54
KCA Reading 2009						

Figure 2

Scatterplot of MAP and KCA scores for reading for 2008-2009

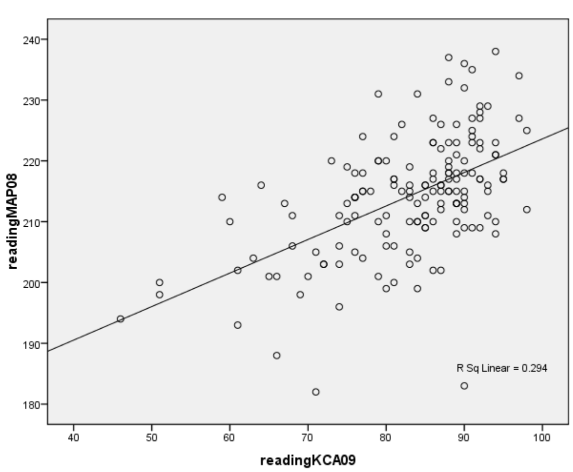


Table 4 indicates a moderate correlation between the two assessments at .54 ($r=.54$, $p<0.001$). This demonstrates that the 2008 reading MAP was a moderate predictor of performance on the 2009 reading KCA. The R^2 value of .29 means that the overall model explains about 29% of the variance in student performance on the 2009 reading KCA. The unstandardized coefficient of .53 means that for every point a student's score increased on the 2008 reading MAP, his score on the 2009 reading KCA should have increased by .53 points. The standardized coefficient of .54 shows that a student's score on the KCA should increase by .54 standard deviations if the student's score on the MAP increased by one standard deviation. Because scores in Figure 2 are not concentrated around the regression line, it is also visually evident that the MAP for reading for 2008 was not a strong predictor of performance on the KCA for reading in 2009.

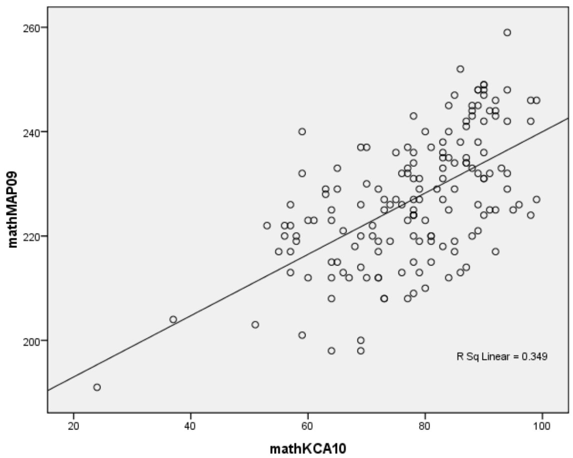
Table 5

Correlation between 2009 math MAP and 2010 math KCA

Tests	N	R	R^2	Sig.	b	β
MAP Math 2009	160	.59	.35	.00	.60	.59
KCA Math 2010						

Figure 3

Scatterplot of MAP and KCA scores for math for 2009-2010



A moderate correlation between the two assessments at .59 ($r=.59$, $p<0.001$) is indicated in Table 5. This demonstrates that the 2009 math MAP was a moderate predictor of performance on the 2010 math KCA. The R^2 value of .35 means that the overall model explains about 35% of the variance in student performance on the 2010 math KCA. The unstandardized coefficient of .60 means that for every point a student's score increased on the 2009 math MAP, his score on the 2010 math KCA should have increased by .60 points. The standardized coefficient of .59 shows that a student's score on the KCA should increase by .59 standard deviations if the student's score on the MAP increased by one standard deviation. Figure 3 also demonstrates that there is not an abundance of scores concentrated around the regression line, so the math MAP for 2009 was not a strong predictor of performance on the math KCA for 2010.

Table 6

Correlation between 2009 reading MAP and 2010 reading KCA

Tests	N	r	R^2	Sig.	b	β
MAP Reading 2009	160	.68	.47	.00	.53	.68
KCA Reading 2010						

Figure 4

Scatterplot of MAP and KCA scores for reading for 2009-2010

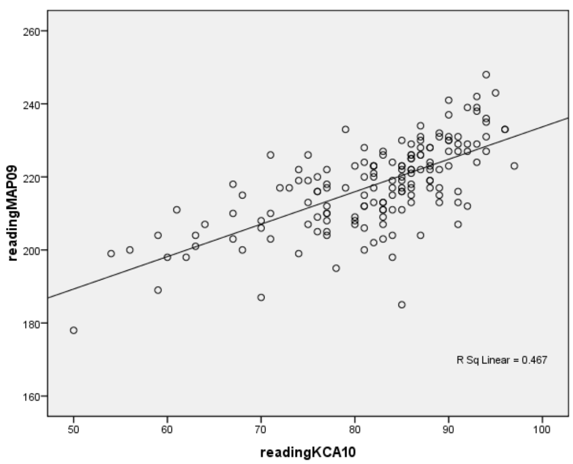


Table 6 shows a moderate relationship that is approaching a strong correlation between the two assessments at .68 ($r=.68$, $p<0.001$) . This demonstrates that the 2009 reading MAP was a moderate to strong predictor of performance on the 2010 reading KCA. The R^2 value of .47 means that the overall model explains about 47% of the variance in student performance on the 2010 math KCA. The unstandardized coefficient of .53 means that for every point a student's score increased on the 2009 reading MAP, his score on the 2010 reading KCA should have increased by .53 points. The standardized coefficient of .68 shows that a student's score on the

KCA should increase by .68 standard deviations if the student's score on the MAP increased by one standard deviation. Figure 4 supports the idea that the relationship is moderate to strong because there is a higher concentration of scores located along the regression line.

Table 7

Correlation between 2010 math MAP and 2011 math KCA

Tests	N	r	R^2	Sig.	b	β
MAP Math 2010	171	.72	.52	.00	.81	.72
KCA Math 2011						

Figure 5

Scatterplot of MAP and KCA scores for math for 2010-2011

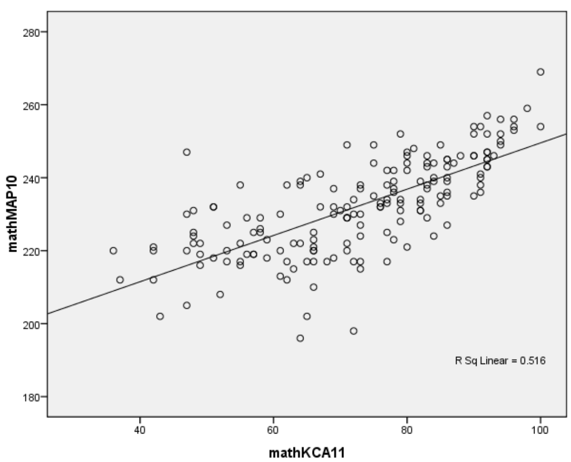


Table 7 shows a strong relationship between the two assessments at .72 (r=.72, p<0.001). This demonstrates that the 2010 math MAP was a strong predictor of performance on the 2011 math KCA. The R^2 value of .52 means that the overall model explains about 52% of the variance in student performance on the 2011 math KCA. The unstandardized coefficient of .81

means that for every point a student's score increased on the 2010 math MAP, his score on the 2011 math KCA should have increased by .81 points. The standardized coefficient of .72 shows that a student's score on the KCA should increase by .72 standard deviations if the student's score on the MAP increased by one standard deviation. Figure 5 shows that more scores are located along the regression line, illustrating that the math MAP for 2010 was a slightly more accurate predictor of performance on the math KCA for 2011.

Table 8

Correlation between 2010 reading MAP and 2011 reading KCA

Tests	N	r	R^2	Sig.	b	β
MAP Reading 2010	168	.70	.48	.00	.68	.70
KCA Reading 2011						

Figure 6

Scatterplot of MAP and KCA scores for reading for 2010-2011

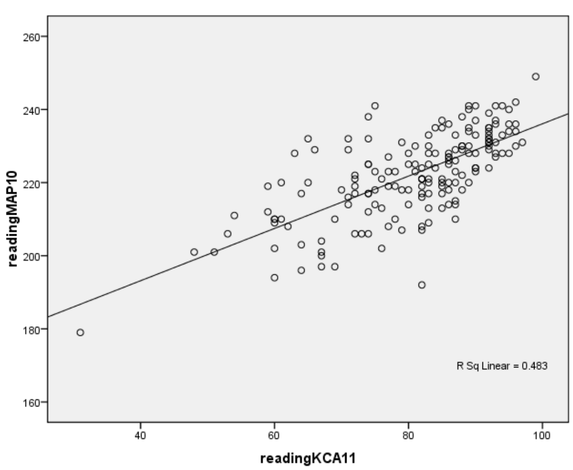


Table 8 shows a strong relationship between the two assessments at .70 ($r=.70$, $p<0.001$). This demonstrates that the 2010 reading MAP was a strong predictor of performance on the 2011 reading KCA. The R^2 value of .48 means that the overall model explains about 48% of the variance in student performance on the 2011 reading KCA. The unstandardized coefficient of .68 means that for every point a student's score increased on the 2010 reading MAP, his score on the 2011 reading KCA should have increased by .68 points. The standardized coefficient of .70 shows that a student's score on the KCA should increase by .70 standard deviations if the student's score on the MAP increased by one standard deviation. Figure 6 helps support the idea of a strong relationship between the reading MAP for 2010 and the reading KCA for 2011 because there are fewer scores falling off of the regression line and many scores concentrated on or very near the line.

Research question #2

To determine if expected patterns existed between math and reading Measures of Academic Progress and the math and reading Kansas Computerized Assessments for the 2008-2009, 2009-2010, and 2010-2011 school years, a multi-trait, multi-method, multi-year matrix was employed. This matrix should demonstrate if the assessments were valid measurement instruments. The expected relationships between the tests are again presented in Figure 7.

Figure 7

Multi-trait, multi-method, multi-year matrix of predicted correlations for MAP and KCA scores

			2008-2009				2009-2010				2010-2011			
			Math		Reading		Math		Reading		Math		Reading	
			MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA
2011	Reading	KCA	low	low	medium	high	low	low	medium	high	low	low	medium	X
		MAP	low	low	high	medium	low	low	high	medium	low	low	X	
2010	Math	KCA	medium	high	low	low	medium	high	low	low	medium	X		
		MAP	high	medium	low	low	high	medium	low	low	X			
2010	Reading	KCA	low	low	medium	high	low	low	medium	X				
		MAP	low	low	high	medium	low	low	X					
2009	Math	KCA	medium	high	low	low	medium	X						
		MAP	high	medium	low	low	X							
2009	Reading	KCA	low	low	medium	X								
		MAP	low	low	X									
2008	Math	KCA	medium	X										
		MAP	X											

Presented in Figure 8 are the actual correlational values for each test after the Pearson Product Moment Correlations for each test were conducted using SPSS.

Figure 8

Multi-trait, multi-method, multi-year matrix of correlations between MAP and KCA scores

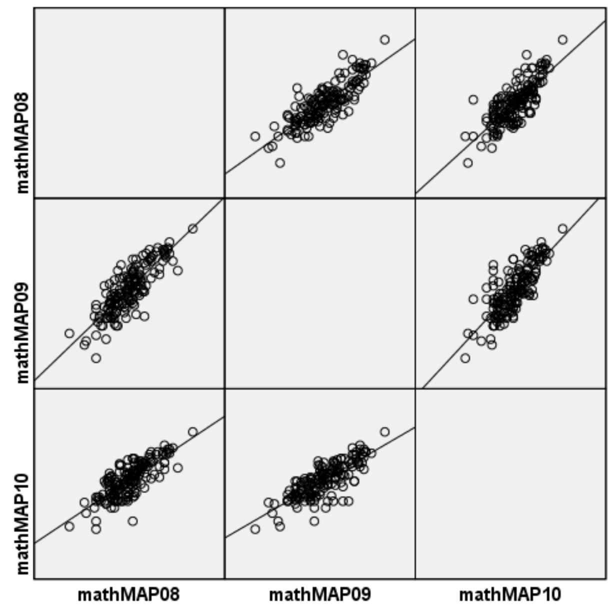
			2008-2009				2009-2010				2010-2011			
			Math		Reading		Math		Reading		Math		Reading	
			MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA	MAP	KCA
2011	Reading	KCA	medium 0.46	medium 0.53	medium 0.57	medium 0.66	medium 0.45	medium 0.58	medium 0.65	medium 0.70	medium 0.60	medium 0.66	medium 0.70	X
		MAP	medium 0.46	medium 0.52	high 0.71	medium 0.66	medium 0.60	medium 0.61	high 0.74	high 0.76	high 0.74	medium 0.52	X	
2010	Math	KCA	medium 0.55	high 0.72	medium 0.40	medium 0.50	medium 0.56	high 0.74	medium 0.48	medium 0.53	high 0.72	X		
		MAP	high 0.78	high 0.74	medium 0.62	medium 0.48	high 0.79	high 0.79	medium 0.66	medium .66	X			
2010	Reading	KCA	medium 0.64	medium 0.55	medium 0.70	medium 0.64	medium 0.57	medium 0.69	medium 0.68	X				
		MAP	medium 0.63	medium 0.48	medium 0.70	medium 0.57	medium 0.67	medium 0.50	X					
2009	Math	KCA	medium 0.61	medium 0.69	medium 0.54	medium 0.48	medium 0.59	X						
		MAP	high 0.83	medium 0.56	medium 0.60	medium 0.39	X							
2009	Reading	KCA	medium 0.45	medium 0.52	medium 0.54	X								
		MAP	medium 0.67	medium 0.45	X									
2008	Math	KCA	medium 0.60	X										
		MAP	X											

Figure 8 shows that the following tests were highly correlated: 2008 reading MAP and the 2010 reading MAP (r=.71), the 2009 reading MAP and the 2010 reading MAP (r=.74), the the 2010 reading KCA and the 2010 reading MAP (r=.76), the 2010 reading MAP and the 2010 math MAP (r=.74), the 2009 math KCA and the 2011 math KCA (.72), the 2010 math KCA and the 2011 math KCA (r=.74), the 2010 math MAP and the 2011 math KCA (r=.72), the 2008 math MAP and the 2010 math MAP (r=.78), the 2009 math KCA and the 2010 math MAP (r=.74), the 2009 math MAP and the 2010 math MAP (.79), the 2010 math KCA and the 2010 math MAP (.79), and the 2008 math MAP and the 2009 math MAP (.83). All other tests showed a medium level of correlation. No two tests had weak or no correlation.

Every math MAP was highly correlated with every other year of the math MAP with the strongest relationship being between the 2008 and 2009 assessments (r=.83), and the weakest relationship being not much behind that at r=.78 for the 2008 and 2010 assessments. Figure 9 shows the scatterplot matrix for all math MAP assessments which helps to demonstrate the high correlations between the test scores.

Figure 9

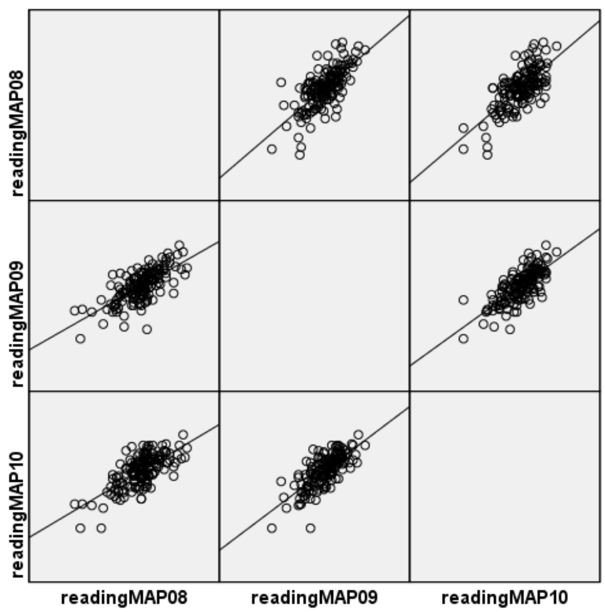
Scatterplot matrix of correlations between math MAP scores



Conversely, the same correlations do not exist between all reading MAP scores. Only the 2010 reading MAP is highly correlated with the other two years of the MAP (2008 reading MAP r=.71, 2009 reading MAP r=.74). Figure 10, while still showing a medium relationship between all of the reading MAP scores, also shows that there are many scores falling off of the regression line.

Figure 10

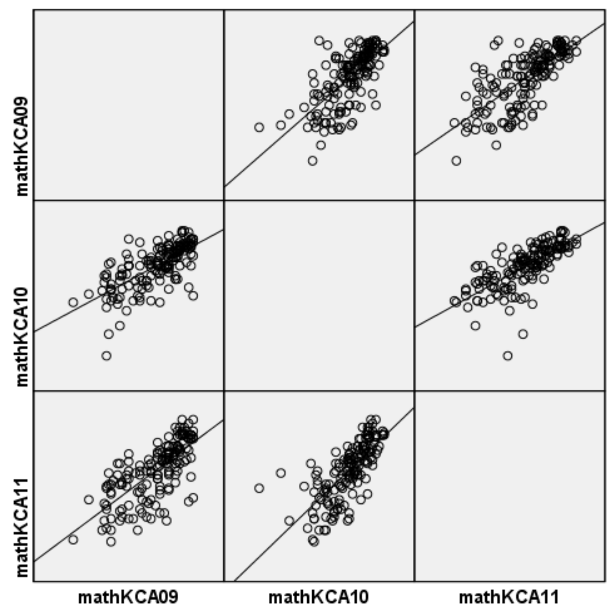
Scatterplot matrix of correlations between reading MAP scores.



For the math Kansas Computerized Assessments, the 2011 test was correlated at a high level with the previous two years' scores (2009 math KCA r=.72, 2010 math KCA r=.74), but only a medium correlation existed between the scores for the 2009 and 2010 assessments. Figure 11 of the scatterplots of tests scores further illustrates those relationships.

Figure 11

Scatterplot matrix of correlations between math KCA scores



No two reading KCA were highly correlated with one another. All reading scores for the KCA were moderately correlated.

Between assessments, the math MAP and KCA scores showed stronger correlations than the reading MAP and KCA scores. The only reading KCA scores to be correlated at a high level with any other test happened between the spring 2010 reading KCA and the following fall's reading MAP scores (r=.76). However, the 2010 math MAP and the 2011 math KCA were highly correlated (.72), the 2009 KCA and the following fall's MAP were highly correlated (.74), and the 2010 KCA and the following fall's MAP scores were correlated at a high level (.79).

The only strong correlation between subject areas occurred between the 2010 math MAP and the 2010 reading MAP which were strongly correlated at r=.74. All other cross subject matter tests were correlated with one another at a medium level.

Summary

Test scores from the Measures of Academic Progress and Kansas Computerized Assessments for math and reading for the 2008-2009, 2009-2010, and 2010-2011 school years were entered into the Statistical Package for the Social Sciences and descriptive statistics, correlational data, and scatterplots were produced to answer the research questions presented in this study. Regarding Research Question #1 and the predictive value of the Measures of Academic Progress on the Kansas Computerized Assessments, it was determined that the Measures of Academic Progress was at least moderately correlated with performance on the Kansas Computerized Assessments for the specified school years, and therefore, at least a moderate predictor of performance on the Kansas Computerized Assessments. Results were somewhat different than expected for Research Question #2 involving the correlations existing between and within all administrations of the Measures of Academic Progress and the Kansas Computerized Assessments for the specified school years. Every test was at least moderately correlated with every other test, and strong relationships existed, most notably, between yearly administrations of the Measures of Academic Progress for math.

Chapter Five

Discussion

With the passage of the No Child Left Behind legislation and an increased focus on state standardized test scores in schools, it is understandable that school districts would want to do all that is in their power to help predict student performance on those state tests and provide students with the individualized instruction necessary to improve performance on state tests. The Northwest Evaluation Association provides the Measures of Academic Progress as one such predictor of performance. The predictive value of the Measures of Academic Progress on the Kansas Computerized Assessments for one smaller sized rural district in Kansas guided this research study. As state funding for schools continues to diminish, it was important for this district to know if spending funding on the Measures of Academic Progress was a wise investment or if money should be allocated to more valuable predictors.

In regards to Research Question #1, results demonstrate that the Measures of Academic Progress is a predictor of performance on the Kansas Computerized Assessments, but the strength of that prediction is lower than expected. The predictive relationship of the MAP on the KCA was at its strongest for both math and reading during the most current administrations, but even those correlations showed relatively weak relationships. The r^2 for the relationship between the 2010 Math MAP and the 2011 Math KCA was .516, and the standard error of the estimate was 9.256, while the r^2 for the relationship between the 2010 Reading MAP and the 2011 Reading KCA was .483, and the standard error of the estimate was 8.543. The bigger the r^2 is and the smaller the standard error of the estimate, the better the model is. Since these two values are not incredibly strong in either regards, the MAP cannot be considered a strong predictor of performance on the KCA.

Because the Measures of Academic Progress is only a moderate predictor of performance on the Kansas Computerized Assessments, it cannot be considered the most valuable tool the school could use to help prepare students for the state standardized assessments. There might be more valuable and cost effective predictors of performance out there, and the district should consider looking into those.

Overall, findings for Research Question #2 were different than the expected patterns presented in Figure 7. Actual patterns were represented in Figure 12. The tests that were expected to have a high correlation generally do follow the expected pattern, but because everything else is correlated at a medium level, none of the low relationships are present. The MAP for math showed a high level of correlation from one year to the next suggesting that it is measuring what it set out to measure. The 2011 administration of the KCA for math and the 2010 administration of the MAP for reading were both strongly correlated with all of the previous administrations of those respective tests, also suggesting that those tests are valid assessments. The same was not the case for the KCA for reading; no two tests were highly correlated. This might lead one to question how valid the reading measure is and how well it depicts actual reading performance. This brings up an interesting question since students at the school used in this study tend to perform better on the reading KCA than on the math KCA. If the math KCA is a more valid and consistent measure, does it provide a more accurate representation of student skills than the reading KCA does? This would be something to consider in further studies.

Because every test was correlated with every other test at at least a medium level, the MAP tests cannot be considered very valid measures of student performance. The fact that the reading assessments were correlated with other reading assessments almost as strongly as with math assessments means that the reading tests can predict performance on math tests almost equally as well as they predict performance on reading tests, which should not be the intended purpose of the tests. This suggests that more than reading and math abilities, a students' general aptitude or testing ability is what can be best ascertained from looking at performance on these assessments.

There are a few factors that might have impacted this validity test. After the administration of the MAP test, teachers might adapt their instruction to better prepare students for the KCA. The MAP is aligned to the KCA, so teachers probably use the results of the MAP to see in what areas their students need additional instruction and teach accordingly. This could impact scores on the KCA and affect the results of this study since students who score low on the MAP might improve their KCA scores after more focused instruction.

Maturation might also have affected the validity of this study. Students probably change both positively and negatively between the fall administration of the MAP and the spring administration of the KCA. This could have both positive and negative effects on the KCA scores, which in turn, could have impacted the results of this study.

The practice effect might also have affected the testing validity of this study. Students take the MAP test every fall and spring and the KCA every spring. The repetition of this process might just make students better test takers over time. This might mean that changes in performance over time on both the MAP test and KCA are more dependent on attitude or being comfortable with testing, or other factors not considered in this study.

Some students may have performed on the KCA at a level lower than what their MAP test scores predicted. This could occur because of regression to the mean. Some students who do really well on the MAP test may be over-achieving and their "under-achieving" performance on the KCA might have been that they were just returning to average performances. However, since this study occurred over a three year time period, the effects of regressing to the mean were hopefully countered.

Limitations

Threats to external validity relate to the ability to make generalizations from the results of this study ("Chapter 8: Internal and External Validity", n.d.). With this particular study, population generalizations are difficult. Because a population mostly lacking in diversity that was not randomly selected but instead considered a convenience sample was used, it is difficult to generalize the results of this study to any students beyond this particular school.

Further questions of study

Because this study only performed zero-order correlations looking at the relationship between the Measures of Academic Progress and the Kansas Computerized Assessments, it would be interesting to determine how other variables such as grades in math and reading, overall grade point average, time spent on homework, and socio-economic status would have affected the predictive relationship between the Measures of Academic Progress and the Kansas Computerized Assessments. It is possible that any or all of these factors as well as other factors could have as strong as or a stronger predictive value than the Measures of Academic Progress.

It also would be interesting to determine why the math MAP and KCA scores were more strongly correlated than the reading scores. Generally students in this particular school perform better on the reading KCA than the math KCA, so it could be interesting to see what factors might influence why math scores were more strongly correlated with other math scores and why the same did not hold true for reading scores. Another area of interest might be looking to see if the same correlational patterns hold true for other grades beyond seventh and eighth. Continuing the study for the 2011-2012 and 2012-2013 school years might also provide compelling data to determine if a stronger correlational relationship between the Measures of Academic Progress and the Kansas Computerized Assessment is developing. Because the most recent administrations of the assessment provided the strongest correlations, continuing the study would help determine if this was by chance or if there is a pattern of performance developing. If the pattern of correlation continues, this could lead to other studies relating to the causes for increased correlation.

References

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matriz. *Psychology Bulletin*, *56*(2),

Cizek, G.J., Gierl, M.J., & Alves, C.B. Measures of Academic Progress. (2004). *Mental Measures Yearbook*, 18. (18013571)

Cronin, J. (2007, March). *A study of the alignment of the NWEA RIT scale with the Kansas assessment system*. Retrieved from http://docs.abileneschools.org/~deniseguy/Assessment/Site/Assessment_files/Kansas%20Alignment%20Report%204.18.07.pdf

Denny, K. Kansas Department of Education, (2007). *News: eighty-nine percent of schools make ayp in 2007* Topeka, KS: Retrieved from http://www.ksde.org/LinkClick.aspx?fileticket+f1dE2oauBF0%3D&tabid=403

Irwin, P.M., & Poggio, A., et.al. Kansas Department of Education, (2007). *Kansas assessments in reading and math 2006 testing manual* Lawrence, KS: Center for Educational Testing and Evaluation. Retrieved from http://www.cete.us/research/reports/pdfs/irwin2007_math.pdf

Kansas Department of Education, (2010a). *Adequate yearly progress (ayp) 2009-2010 fact sheet* Topeka, KS: Retrieved from www.ksde.org/LinkClick.aspx?fileticket=5deCgi3lOqI%3D&tabid=4398

Kansas Department of Education, (n.d.). *assessmentpldguidelines.* Topeka, KS: Kansas Department of Education. Retrieved from http://ksde.org

Kansas Department of Education, (2010b). *Kansas math assessment fact sheet 2010-11* Topeka, KS: Retrieved from www.ksde.org/LinkClick.aspx?fileticket=-9o7JER9Hpk%3D

Kansas Department of Education, (2010c). *Kansas reading assessment fact sheet 2010-11* Topeka, KS: Retrieved from http://www.ksde.org/LinkClick.aspx?fileticket=q-jU6m9nVJ0%3D

Kansas Department of Education, (2009). *Report card 2008-209* Topeka, KS: Retrieved from http://svapp15586.ksde.org/rcard/summary/fy2009/D04346440.pdf

Kansas Department of Education, (2010d). *Report card 2009-2010* Topeka, KS: Retrieved from http://svapp15586.ksde.org/rcard/summary/D04346440.pdf

Keith, T.Z. (2006). *Multiple regression and beyond*. Boston: Pearson Education Inc.

Keppel, G, & Wickens, T.D. (2004). *Design and analysis: a researcher's handbook*. Upper Saddle River, New Jersey: Pearson Prentice Hall.

McIntire, S.A., & Miller, L.A. (2007). *Foundations of psychological testing: a practical approach*. Thousand Oaks, Calif: Sage Publications.

Moore, R., personal communication, March 2011.

Northwest Evaluation Association, The Kingsbury Center. (n.d.). *Linking map to state tests: proficiency cut score estimation procedures* Retrieved from http://www.nwea.org/sites/www.nwea.org/files/NWEA%20State%20Standards%20Alignment%20Study%20Methods%20_3_.pdf

Northwest Evaluation Association. (2009). *A study of the alignment of the NWEA RIT Scale with the Kansas assessment system.* Retrieved from http://www.nwea.org/sites/www.nwea.org/files/resources/Kansas%20Alignment%20Report%20110709.pdf

Nugent, C. (2009). *Predicting standardized state achievement scores from Measures of Academic Progress.* Retrieved from http://faculty.ksu.edu.sa/Magda/My%20Research%20Library/Evaluation/Measures%20of%20academic%20progress.pdf

Peyton, V. (2008a). *Consequential validity of the Kansas computerized reading and mathematics assessments: educator perspectives*. Center for Educational Testing and Evaluation, University of Kansas, Lawrence, KS. Retrieved from http://www.cete.us/research/reports/pdfs/peyton2008_validity.pdf

Peyton, V. (2008b). *Quality and utility of the Kansas computerized assessment system from the perspective of the Kansas educator*. Center for Educational Testing and Evaluation, University of Kansas, Lawrence, KS. Retrieved from http://www.cete.us/research/reports/pdfs/peyton2008_utility.pdf

Price, I. (2000). What α-level. In *Research methods and statistics PESS202 lecture and commentary notes* (Chapter 5 Part II). Retrieved from http://www.une.edu.au/WebStat/unit_materials/c5_inferential_statistics/what_alpha_level.html

Shields, J. (2008, May). *A comparison of the NWEA Measures of Academic Progress and the Missouri Assessment Program.* Retrieved from https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/5589/research.pdf?sequence=3

Shadish, W., Cook, T.D., & Campbell, D.T. (2002) *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin Company.

Simon, S. (2005, August 18). *What is correlation? (pearson correlation)*. Retrieved from http://www.childrens-mercy.org/stats/definitions/correlation.htm

U.S. Department of Education, National Center for Education Statistics. (2009). *The impact of the Measures of Academic Progress on differentiated instruction and student achievement*. Washington, DC: common core of data.

U.S. Department of Education, (2007). *Nation's report card shows record gains* Washington, DC: Retrieved from http://www2.ed.gov/nclb/accountability/achieve/report-card2007.html

U.S. Department of Education, (2008). *Progress by our schools and the U.S. Department of Education* Washington, DC: Retrieved from http://www2.ed.gov/nclb/accountability/results/trends/progress.html