5-1-2012

A Study of the Relationship between Nonverbal Kindergarten Ability and Third-grade Reading Achievement

Aaron John Wills
University of Missouri-St. Louis, aaronwills16@gmail.com

A Study of the Relationship between Kindergarten Nonverbal Ability and Third-grade Reading Achievement

Aaron J. Wills
M.S. Educational Administration, Southwest Baptist University, 2001
M.A. Elementary Education, Truman State University, 1995
B.A. Business Administration, Truman State University, 1993

A Dissertation submitted to The Graduate School of the University of Missouri–St. Louis
in partial fulfillment of the requirements for the degree
Educational Doctorate in Educational Leadership

April 2012

Advisory Committee

Carole Murphy, Ph.D.
Chairperson

Kathleen Brown, Ph.D.

Cody Ding, Ph.D.

Margaret Dolan, Ph.D.

Table of Contents

Page

Appendix

List of Tables

Table Page

List of Figures

Abstract

Increased scrutiny of educational proficiency targets has intensified the urgency for educators to identify measurements that indicate students' likelihood of eventual achievement in reading. This regression analysis explored the relationship between nonverbal ability in kindergarten as measured by the Naglieri Nonverbal Ability Test (NNAT) and eventual third-grade achievement in reading and writing as measured by the Missouri Assessment Program (MAP). Naglieri and Ronning (2000) found a range of correlational *r* values from .49 to .56 when comparing NNAT results to results from the Stanford Achievement Test 9 (SAT-9) in reading when tested concurrently at various grade levels. The present research examined data from two cohorts of students (*n* = 794, 795) and produced correlational *r* values of .50 and .44 with a four-year span between assessment administrations. These *r* values are similar to those found in other research comparing ability with reading achievement in the early childhood years. Furthermore, this study examined the results of multiple regression analysis between seven student demographic subgroup categories and identified the NNAT's ability to predict MAP achievement for each group.

Chapter 1

Introduction

Background

Since 1998, Missouri has evaluated the achievement of its students, and consequently its school systems, through the Missouri Assessment Program (MAP). Teachers and administrators focus tightly on preparing students for these benchmark examinations that will tell the tale of whether or not they are seen as successful or deficient. So much is riding on the results—accreditation, adequate yearly progress ratings, regional status, and more. The MAP has caused an increased sense of urgency for raising achievement among Missouri's children and the anticipated implementation of the Common Core Standards beckons even more attention toward gaining measurable progress for each child. Like never before, educators are implementing innovative interventions and employing intense strategies to directionally enhance the young minds that enter their classrooms. Children are targeted at an early age—in fact, the enduring race toward proficient academic achievement begins for each child on the first day he or she walks through the kindergarten classroom door.

Those young minds, though, enter this race carrying vastly different experiences, abilities, and levels of mental development and training. As they arrive at school, they will be grouped into *classes*, but attempting to *classify* them by anything other than their chronological ages may be difficult at best. Various socioeconomic levels, family structures, religious orientations, physiological chemistries, language exposures, social experiences, and innate mental abilities have molded each child into a unique being that will begin this race at a different place than the child sitting across the table.

While there is limited uniformity in the starting gates, the finish line is the same for all. At the end of the third-grade year, the first MAP assessment awaits all Missouri public school children—measuring their acquired skills in communication arts (reading and writing) and mathematics. For most students, it will be less than four years between the first day of kindergarten and the first day of MAP testing. The urgency and pressures felt by educators is certainly justified. Four short years to proficiency!

But again, each child enters the race with different skills and abilities, some of which are quickly measured in kindergarten. Can these kindergarten ability assessments offer clues to predicting a child's eventual success on the high-stakes MAP test?

Researchers have shown an interest in studying the relationships between early childhood assessments and success in reading achievement tests. Carver (1990), and Naglieri and Ronning (2000b) have specifically studied the correlation of nonverbal ability tests and reading achievement. Hayes (1999), Bracken and McCallum (1998), and Naglieri and Ronning (2000a) found that nonverbal ability measurements have proven to be less influenced by language skills, and therefore may be more appropriate for students with limited English proficiency than measurements with verbal components. It would seem plausible then to assume that nonverbal ability tests may be of particular interest to primary-level educators since even the youngest students can participate (whether or not they are readers or even speakers of the English language) if in fact the nonverbal assessments can reveal inferential correlations to eventual reading achievement.

Purpose of the Study

This study examined the relationship between scores attained on kindergarten nonverbal general ability measurements and scores attained on third-grade MAP achievement measurements in reading. More specifically, the goal was to determine to what extent MAP scores can be inferred by the kindergarten NAI scores. The Naglieri Nonverbal Ability Test (NNAT) is a figural matrix assessment intended to measure general ability. The assessment is completely nonverbal making it appropriate for kindergarten students of all backgrounds as neither reading, writing, nor speaking skills are required for test completion (Naglieri 1997).

Naglieri and Ronning (2000b) studied correlations between Nonverbal Ability Index (NAI) scores and Total Reading scores on the Stanford Achievement Test Ninth Edition (SAT-9). The results revealed *r* values of .56, .52, .49, and .54 in kindergarten, first grade, second grade, and third grade respectively. Subjects in their study were administered both the NNAT and the SAT-9 test during the same testing window in the same year. Prior to that, during the norming process of the NNAT in 1995, Naglieri (1997) had compared NAI scores with SAT-9 Total Reading scores and calculated *r* values of .59, .54, .48, and .53 in kindergarten, first grade, second grade, and third grade respectively.

The purpose of this study was to examine the relationships between the results of the NNAT test given in kindergarten with the same students' results of the communication arts MAP achievement test administered toward the end of the third-grade year. The essential question was: Will the relationship that Naglieri and Ronning (2000b) found between nonverbal ability and reading achievement when tested simultaneously remain consistent when the assessments are given nearly four years apart? The data that was compared included NAI scores and MAP scaled scores in communication arts. This study also investigated how the relationships compare between seven demographic subgroups of students.

Hypothesis

Based on the findings of Naglieri and Ronning (2000a) it was hypothesized that the kindergarten NAI scores would correlate with the MAP scaled scores with an *r* value between .49 and .56. This *r* value range was the same that the researchers found to be true for correlation between NAI scores and SAT-9 scores. In addition, Carver (1990) studied correlations between the Raven Progressive Matrices test (another nonverbal ability test) and the National Reading Standards test. He found an average correlation *r* =.50, similar to that of Naglieri and Ronning (2000b). The hypothesis of this study was further supported by several other research studies (presented in the following chapter) which indicated that eventual reading achievement can positively correlate with measurements conducted during the early childhood years.

Delimitations

If the NNAT results do in fact correlate with the third-grade communication arts scores as hypothesized, then the results would verify that Missouri schools with similar demographics as those in the study's population could rely on nonverbal ability tests with young students to help infer the eventual success on the MAP, just as Naglieri and Ronning (2000b) found to be true with the Total Reading component of the SAT-9.

This study was not able to evaluate particular instructional programs or interventions. Gredler (1997) states that one problem with identifying strong predictor tests is that the instrument cannot accurately assess the impact of successful teaching. The student participants in the study are likely to have been exposed to several different reading programs, teaching styles, class sizes, out-of-school learning experiences, attendance rates, and other factors that contribute to learning in school (Jansky 1978).

Because this study explicitly examined the scores near the onset of kindergarten and toward the end of the third-grade year, the particulars of what happens in between were not directly evaluated. Likewise, the results of the study did not promote nor suggest effective instructional methods, interventions, or programs. It was assumed that the child's school employed the best strategies and interventions available at the time of instruction and students were given typical instruction for elementary school. Furthermore, the study did not determine the causes for the variance in NAI scores. The results of the study do not indicate why some students enter kindergarten at higher or lower levels of nonverbal ability than others, nor does it pinpoint the causes of achievement in reading and writing by the end of third grade. The results determine the level of correlation between kindergarten NAI scores and third-grade MAP scores and how the relationship varies between different demographic subgroups.

Limitations

The most significant limitations of the study may in fact come through the interpretation of the strength of the observed correlation coefficients. Previous research by Naglieri and Ronning (2000b) suggests that early elementary (kindergarten through third grade) NAI scores and SAT-9 Total Reading achievement scores correlate between $r = .49$ and $r = .56$. While the authors refer to this size of correlation as large, r values in this range certainly must be cautiously interpreted when considering practical educational applications. Naglieri and Ronning (2000b) based their interpretation on the criteria of Cohen (1977) who refers to an *r* value of .10 as small, .30 as medium, and .50 as large. But correlation coefficients must be interpreted with discretion. Hinkle, Wiersma, and Jurs (2002) caution against interpreting correlations without considering the meaning of the variables, but as a rule of thumb, they refer to *r* value correlations from .00 to .30 as having little if any correlation, .30 to .50 as low correlation, .50 to .70 as moderate correlation, .70 to .90 as high correlation, and .90 to 1.0 as very high correlation. Clearly, the criteria for categorizing the *r* value significantly impact the interpretation of the findings of this study and is a point of discussion in the later chapters.

Furthermore, Naglieri (1997) recognizes what is perhaps the most significant limitation of educational measurements of any kind in that an individual child's score could be influenced by factors beyond the control of the testing environment and the standardized test administration procedures. Illness, sleepiness, anxieties, or countless other disruptions could occur in a child's life causing inaccurate measurements of ability and/or achievement on any given day.

The motivation of the child must also be considered when defining the limits of the study. Young children participating in the NNAT measurement may have little or no motivation to do well. There are no incentives for the child and since the testing for this study takes place in kindergarten, there is likely to be little understanding of the purpose of the test. At the end of third grade on the other hand, students have possibly been given incentives to perform well on the MAP either by their parents, teachers, or school system. The MAP has significant implications for schools and several motivational strategies may be used. Students may be rewarded for working hard, taking their time, attendance and punctuality, etc. By third grade, students may also have developed a sense of academic pride—an eagerness to score well on the test to show how much they have learned. The individual motivators at this level may be both intrinsic and external. However, even at the third-grade level, there still may be some students who do not put forth their best efforts.

Other limitations include generalizability with respect to different demographic populations and geographic locations from the sample population. The study only considered scores from a large, suburban, mostly middle class Missouri public school district. Furthermore, of the sample population, only students with both scores (NNAT and MAP) were included in the study. Students with one or more missing scores due to absence, transiency, or other causes were not represented in the data.

Definitions

Figural Matrix Assessment: A nonverbal assessment in which test takers view and select matrices of geometrical shapes and lines arranged in related patterns.

Progressive Matrix Assessment: An assessment comprised of figural matrixes arranged so that they progressively increase in complexity from one test item to the next.

Communication Arts: A term used by Missouri educators to describe the combined subjects of reading and writing.

Significance

As educators, legislators, and policymakers continuously strive to increase student achievement in Missouri, it is more and more desirable to identify any data with correlational relationships to academic success. If in fact a child's kindergarten nonverbal ability scores can predict his or her achievement level at the end of third grade, schools could immediately identify those who may need special attention and could provide extra learning interventions in the earliest years of school.

Nonverbal ability tests correlate similarly across ethnic groups, socioeconomic statuses, geographic regions, and across public and private schools (Naglieri and Ronning 2000b) and thus may have broader interest to Missouri's educators than ability tests which include a verbal component. The Missouri Census Data Center (2009) indicates that Missouri's population is becoming more ethnically diverse as the percentage of African American, Hispanic, and Asian/Pacific Islander populations have increased over the last fifteen years (and are projected to continue to increase) while the percentage of the White/non-Hispanic population has decreased. If immigration from other non-English speaking countries is causing some of this population shift, it would seem plausible that a nonverbal ability assessment may become a more valuable assessment tool than one with a verbal component.

Chapter 2

Review of Related Literature

The passing of the federal No Child Left Behind Act (NCLB) in 2001 set into motion a critical quest to identify, diagnose, and remediate students who are or may become poor readers. Though reading achievement has always been a priority in schools, the NCLB mandate beckoned immediacy for schools to lead students to a level of reading proficiency. In fact, NCLB was designed to eventually require *all* students to be proficient readers by the end of the third-grade year. Reading proficiency measured by state assessments has become a non-negotiable goal of public school systems across the nation (No Child Left Behind [NCLB], 2001) According to NCLB, students *must* become proficient readers by the end of third grade. In Missouri, this means students *must* perform well in the Missouri Assessment Program (MAP) in the area of communication arts.

In order to successfully meet the stringent challenges set forth in NCLB, educators are seeking the most accurate measurements to predict and diagnose the students needing interventions in reading at the earliest stages of literacy development. Researchers agree that identifying poor readers at the earliest stages of development increase the likelihood of educators' abilities to prevent subsequent reading problems (Kennedy, Birman, & Demaline, 1986; Berninger, Thalberg, DeBruyn, & Smith, 1987; Taylor, Short, Frye, & Shearer, 1992; Clay 1993; Blachman 1994; Share & Stanovich 1995; VanDerHeyden, Witt, Naquin, & Noell, 2001). The urgency to identify and intervene with poor readers as early as possible is supported by multiple longitudinal studies which unfortunately suggest that a child who has low reading skills by the end of first grade will almost certainly be a low-level reader with below grade-level skills by the end of the elementary school years (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Torgesen & Burgess 1998; Torgesen 2002). Similarly, a U. S. Department of Education (2001) study revealed that 37% of our nation's fourth-grade students could not read at the level necessary to successfully complete their classwork.

Mastery of early reading skills is essential for school districts to meet state and federal guidelines; however it is much more crucial for students to be strong readers in order to ensure their chances of being successful learners in the later years of their schooling. Countless studies have demonstrated that a student's likelihood of success in school is relative to his or her reading skills gained in the primary grades of elementary school. Juel (1988) found that children who were poor readers at the end of first grade were still likely (.88) to be poor readers as they finished fourth grade. Similarly, Scarborough (1998) reported the likelihood (.77) that poor readers at the end of second grade would also be poor readers at the end of eighth grade. Cunningham and Stanovich (1998) found that vocabulary growth is negatively affected by delays in the development of early reading skills. Furthermore, vocabulary delays in young children correlated with low motivation and a poor attitude toward reading in later years (Oka and Paris 1986) and may lead to missed opportunities to develop essential reading comprehension skills (Brown, Palincsar, & Purcell, 1986).

The study of early reading interventions seemed to heighten in the early 1980s, shortly after the U.S. Congress passed the Individuals with Disabilities Educational Act (IDEA; Obrzut, Jones, Bolocofsky, & Heath 1981). IDEA mandated early identification of young children with special needs including learning disabilities. With the passing of the IDEA legislation, school districts, medical agencies, and others were newly accountable for locating and serving young students with developmental delays. This prompted a necessary interest in screening children for such delays before entering schools.

This study though has relevance beyond assisting Missouri school districts in targeting specific students for extra interventions. For years, the educational research community has been interested in correlative studies comparing early childhood skills, abilities, and developmental markers to later success in reading. Scarbourough (1998) analyzed hundreds of predictive reading studies from the mid-1970's through the late 1990's. Since then, researchers have continued to examine the relationship between pre-reading indicators and eventual success in reading.

The course of research attempting to identify predictive correlations in early childhood education has followed several paths with many specific aims. Some studies analyze the effectiveness of standardized testing versus teacher rating scales with the goal of identifying the most reliable preassessment data inputs. After analyzing several previous studies, Teisl, Mazzocco, and Meyers (2001) concluded that sometimes tests were better predictors and sometimes teachers' ratings were more accurate. Specifically, teachers' ratings were more likely to accurately predict low students than high. The predictive ability of teachers' ratings greatly improved when combined with data gained from traditional screening instruments. But as the accountability standards toward student achievement have grown in recent years, schools may now be interested in intervening before or during the kindergarten year. Time is of the essence, thus a year gone by without intervention may have more negative consequences than the positive effects of waiting a year for a teacher's rating.

Other studies attempted to predict preschool children with learning disabilities that would not otherwise be identified until later years after valuable intervention time had been lost. Scott, Delgado, Tu, and Fletcher (2005) found certain subsets of a kindergarten screening battery to predict 91% of third-grade students in special education classes and 85% of students enrolled in regular education classes. Although the amount of false positives (students predicted to be potential special education students that were not identified as special education students by third grade) was high at 57%, most of these children were performing poorly in their regular educational programs. Predictive items and methods such as this would be invaluable to educators aspiring to have all students at a proficient level by the end of third grade.

Even preschool tests are finding significant correlations with reading tests at the end of first grade. For example, Mann (1993) found the Phoneme Segmentation Test (PST) significantly correlates with later reading ability, as did a preschool test for invented spelling. Researchers such as Scott et al. (2005) analyzed the various items on pre-kindergarten tests to find the key tasks that may unlock the mystery of who will be successful in school and who will struggle. In addition, researchers have studied the validity of the tests themselves. As referenced earlier, Gredler (1997) states that one problem with identifying strong predictor tests is that the instrument cannot accurately assess the impact of successful teaching. Many interventions will impact the child along the journey to the post assessment. A child's likelihood of being an at-risk reader is also affected by a child's attendance rate, the quality of the teacher, the instructional strategies employed, the student's attitude toward learning, and other factors impacting student achievement in the early years (Jansky 1978).

Blachman (1983) questions whether our overwhelming interest and desire to find an accurate predictor of early learning exceeds our capacity to truly identify those at risk for future failure. But nonetheless, researchers continue to seek measures which will provide predictors for academic success, and general ability as a measurement of such continues to be a variable of interest. In fact, in a meta-analysis of 34 studies, Hammill and McNutt (1981) found that the median correlation between reading achievement and general ability as measured by the Wechsler Intelligence Scale for Children (WISC) Full Scale score was .44. Similarly, they found that in 33 different studies, a median .46 correlation existed when reading was compared with Stanford-Binet scores. Stanovich, Cunningham, and Freeman (1984) examined the literature and summarized the results of 60 studies between first and third grades. The correlations between various ability and achievement tests ranged from .19 to .76. They then conducted their own study on students in first grade (n = 56), third grade (n = 18), and fifth grade (n = 20) with an assortment of measurements including general intelligence, phonological awareness, decoding speed, and listening comprehension. Stanovich et al. (1984) compared those scores to reading comprehension at the end of the same school year. They reported a correlation of .33 between first graders' general ability and reading achievement scores and slightly higher correlations, .42 and .56, in third and fifth grades respectively. Other researchers such as Carver (1990) and Naglieri (2000b) later expressed doubt in the Stanovich et al. (1984) results due to the small sample sizes in the study.

Most of the studies listed above measured general ability with assessments containing more components than strictly nonverbal figural matrix items. However, Naglieri and Ronning (2000b) studied the relationship between a child's general ability measured with a progressive matrix nonverbal intelligence test and reading achievement. They studied the relationship between the NNAT and the reading component of the Stanford Achievement Test Ninth Edition (SAT-9) over a sample size of approximately 22,000 children. The study indicated that the overall correlation with Total Reading was large (.56). Carver (1990) also found a correlation averaging about .50 between nonverbal ability scores attained on the Raven Progressive Matricies test and the National Reading Standards test (*n* = 486 in grades 2-12).

The cultural neutrality of the figural progressive matrices nonverbal ability tests provides a measurement which may be able to reach a wide population of students. Sattler (1988) found that for quite some time, researchers have shown a keen interest in accurately assessing the intelligence of people from diverse backgrounds. Naglieri and Ronning (2000a) found that nonverbal ability scores attained with the NNAT assessment rendered consistent scores across various cultural groups. Their results suggest that the NNAT scores can be used as a fair assessment of White and minority students.

The current study was based on three major assumptions: 1) Missouri's minority population in increasing and projected to continue to increase (2009), 2) the strong and consistent evidence of correlations between nonverbal ability and reading achievement in the literature, and 3) the urgency in which educators are trying to increase reading scores to meet state and federal mandates.

Chapter 3

Research Design and Methodology

This study was designed to find the correlation between nonverbal kindergarten ability scores measured by the Naglieri Nonverbal Ability Test and third-grade reading achievement scores measured by the Missouri Assessment Program (MAP) assessment. Previous studies have indicated that a correlation exists between nonverbal ability and reading achievement, but this study determined whether the correlation held true when the assessments were given with more than a three year gap between the tests.

Research Design

The method employed in this research was quantitative correlation and regression analysis. The study revealed a Pearson *r* coefficient to describe the relationship between Naglieri Nonverbal Ability Test (NNAT) scores and MAP achievement scores in reading. Regression analysis was also conducted for a deeper understanding of the relationship and the statistical impact of various demographic variables. This method was selected to satisfy the purpose of determining to what extent a child's eventual reading achievement level near the end of third grade can be inferred by his or her measured nonverbal abilities in kindergarten. The data generated by the NNAT preassessment and the MAP postassessment was strictly numerical and was tested statistically to discover what relationships existed through the analysis of correlation coefficients and regression statistics.

Population and Sample

The research involved the evaluation of existing data gathered from a sample of 1589 children attending public elementary schools in a large suburban school district in

Eastern Missouri. The data included kindergarten NNAT nonverbal ability scores obtained in the fall semesters of 2004 and 2005 as well as third-grade MAP achievement data obtained from the same students in the spring semesters of 2008 and 2009.

The school district being sampled has a total K-12 student enrollment of approximately 18,000. The geographic area is predominately suburban middle class, but the district also accepts students from a nearby urban city through a voluntary student transfer program. Approximately 11% of the total student enrollment is composed of urban-dwelling voluntary transfer students. The school district is within 20 miles of a major U.S. city offering work, shopping, recreational, and cultural opportunities.

Most students who attended school in the targeted district participated in both assessments and will be considered in the study so long as they meet the following selection criteria. First, students must have been enrolled and present in the school district at the time of the assessments in both kindergarten and third grade. Transient students who missed either of the two assessments in the district are excluded. Second, students must have successfully obtained a score in each of the areas of both tests. For example, students who were absent and unable to make up the MAP test will not have a valid score to be tested. Likewise, a student who was not administered either test due to IEP restrictions will not be included in the study. For example, students with more profound special needs are sometimes given an alternative MAP test called MAP-A. Due to the differentiation in scoring and item tasks, these students will also be excluded from the study.

The NNAT test is administered to nearly all kindergarten students in the district during a fall testing window which is open during the first two months of the school year.

The results are calculated based on norming procedures provided by the test authors and explained in the assessment's technical manual (Naglieri 1997). The teachers and staff members who administer the test to kindergarten students are appropriately trained by the district. Scoring of the tests is automated offsite by the test publisher. Results are sent electronically to the school district at which time they are entered into a comprehensive student database.

Similarly, the MAP test is administered to students during a testing window in the spring semester of the third-grade year. The test is required by the state and there are no exemptions. Teachers and staff members are trained in the test administration procedures and scoring is completed offsite by the test publisher. When the data is returned to the district, it is entered into the student information database.

Instrumentation

Two published instruments were used in this study to measure the students' nonverbal abilities and achievement, respectively, the Naglieri Nonverbal Ability Test (NNAT) and the Missouri Assessment Program (MAP) test in communication arts. Both tests are widely used throughout elementary education and extensive statistical norms and specifications are available.

Naglieri Nonverbal Ability Test. The Multilevel Technical Manual for the NNAT (Naglieri 1997) describes the assessment as a progressive figural matrix test intended to measure a child's school ability. The author also states that the scores on the test can predict academic achievement. The test is completely nonverbal and seven levels of the test are available so that children in grades kindergarten through 12 can participate. This study considered scores obtained on the Level A form appropriate for kindergarten students. The progressive matrices are figural test items in which the test subjects use reasoning skills to complete geometric patterns. Because neither reading, writing, mathematical, speaking, nor listening skills are required, nearly all children are able to participate regardless of culture, schooling, or content knowledge. The NNAT is an extension and revision of the *Matrix Analogies Test* developed by Naglieri in the mid 1980s (Naglieri 1985). Figural matrix tests have been used in many assessment protocols throughout history and have been widely researched (Naglieri 1997).

Test items. The test items are constructed so that children observe a matrix with colors and geometric figures, and then select from the multiple choice responses the figure that shares the appropriate relationship with the missing area in the matrix. An example of a hypothetical figural matrix item is presented in Figure 3.1.

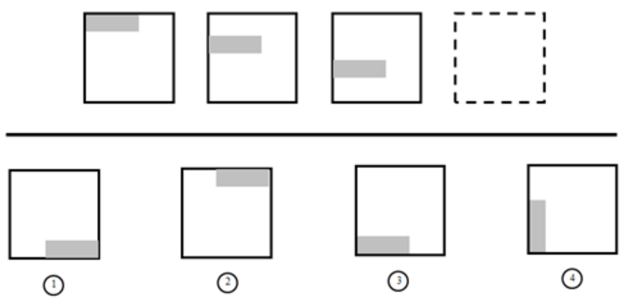


Figure 3.1. Hypothetical figural matrix item.

The NNAT is composed of four clusters each with a slightly different item type. The pattern completion cluster requires a child to observe a large design with a missing part. The child selects the response depicting the missing part. An example of a hypothetical pattern completion item is presented in Figure 3.2.

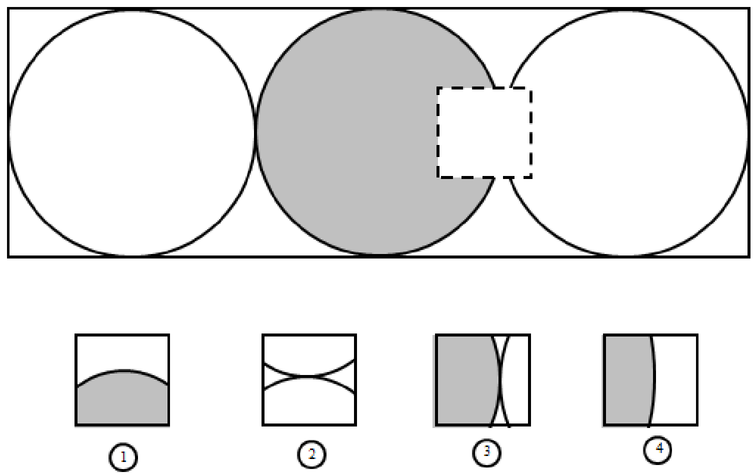


Figure 3.2. Hypothetical pattern completion item.

The other clusters of the NNAT include reasoning by analogy items, serial reasoning items, and spatial visualization items. Each cluster features figural matrices, but presents them in different ways so that the child must determine the relationships within each matrix. Each cluster is scored and reported separately, but the most reliable predictor of school ability is the combination of all clusters resulting in the nonverbal ability index (NAI) score (Naglieri 1997). Naglieri (1997) stated that several educational applications for the NAI score exist. It is a measure of general ability, a predictor of academic achievement, an indicator of students who may suffer from academic problems in school, and an indicator of gifted students exhibiting high general abilities.

Standardization. The complete NNAT was standardized in the fall of 1995 and the spring of 1996 with a K-12 sample population of nearly 90,000 students. Age-based total scores from the fall version of the kindergarten assessment were tested for reliability using the Kuder-Richardson Formula #20. The 38 items on the Level A Kindergarten assessment (n = 997) produced statistics for 5-year-olds and 6-year-olds respectively:

mean = 21.2, 23.6; standard deviation = 6.6, 6.5; standard error of measurement = 2.5, 2.4; and an r^2 value of 0.86, 0.86 (Naglieri 1997).

Validity. The author of the NNAT, Naglieri (1997), concedes that content validity is subject to the judgment of the users of the test. In other words, it is left up to the discretion of the user to determine if the types of items on the NNAT test adequately relate to school ability. Criterion-related validity was studied by correlating the NNAT to the Stanford Achievement Test (SAT) and the Aprenda2 achievement test. During the fall standardization of the NNAT, Naglieri (1997) found that the Level A Kindergarten assessment correlated at $r = .59$ in reading and $r = .66$ in math. The correlations were lower when compared with the Aprenda2 assessment during the spring standardization ($n = 719$) with $r = .30$ and $r = .39$ in reading and math respectively.

Scoring. The NNAT can be hand scored or machine scored. The scores obtained for this study were machine scored by the test publisher. The multiple choice test items are judged either correct or incorrect resulting in raw scores for each student. The raw scores are converted to scaled scores and both raw and scaled scores are sent electronically to the school district. During the standardization process of the NNAT, Item Response Theory procedures, specifically the Rasch model, were applied to the raw scores to determine the one-to-one scaled scores. The scaled scores (NAI) obtained by this method are particularly useful as they have a mean score of 100, are easily computed when hand scoring, and can be interpreted consistently among the various levels of the test across ages and grade levels (Naglieri 1997).

Missouri Assessment Program. The Missouri Assessment Program Technical Report (2008) indicates that the development of the MAP began in 1996 as a component of the Outstanding Schools Act and was designed to measure students' progress toward the state's academic standards known as the Show-Me Standards. Missouri students take their first MAP tests in the spring of the third-grade year in both mathematics and communication arts.

The tests measure content standards, process standards, and grade-level expectations (GLEs). In 2006, Missouri revised the MAP to include benchmark tests in mathematics and communication arts at each grade level beginning at third grade. This major revision was in response to the federally mandated No Child Left Behind Act passed in 2001. In 2008, a science test was introduced at the fifth- and eighth-grade levels. The MAP is continually revised to ensure that the measurement accurately reflects the most current learning standards (2008).

Test items. The third-grade communication arts MAP test is composed of multiple choice items, constructed response (short answer) items, and performance event items. The 2008 version of the test included a total of 58 items which were administered during four testing sessions. The third-grade communication arts test contains one major writing performance event in which students are asked to compose an original piece of writing using the entire writing process. Multiple choice items are each worth one raw point. Constructed response items and performance event items can earn up to four raw points depending on the skill being measured and the tasks required of the students.

After completing the tests, districts return the assessments to the state for scoring. Automatic scoring machines are used to check multiple choice items while constructed response items and performance event items are scored blindly by trained evaluators. In 2008, the third-grade communication arts assessment had a total of 68 raw points. Raw scores are translated to scaled scores. The lowest obtainable scaled score is 455 and the highest obtainable scaled score is 790.

Some multiple choice items on the MAP are retrieved from the nationally-normed TerraNova reading achievement test. In addition to factoring into the overall MAP scaled score, the results from these designated reading items are calculated separately to also provide a TerraNova national reading percentile score. Higher scores indicate higher achievement, lower scores indicate lower achievement.

The overall scaled MAP scores classify students into one of four achievement level categories: below basic, basic, proficient, or advanced. The cut scores for each level were predetermined by a committee of Missouri educators, policymakers, and citizens.

Though the program is under consideration for revision at the federal level, the original goal of No Child Left Behind is that all students will achieve at the proficient or advanced levels by 2014. Scores are reported and disseminated to districts electronically prior to the beginning of the following school year.

Validity. The MAP Technical Report (2008) states that the validity of the test is relative to the interpretation of the scores. The Missouri Department of Elementary and Secondary Education (DESE) have utilized several teams of educators and stakeholders over the years to create items that measure the academic standards valued by the state. The test items were designed to generate scores which could be used for the following purposes as stated in the Missouri Assessment Program Technical Report (2008).

- Identifying students' strengths and weaknesses on Missouri's Grade-Level Expectations

- Communicating expectations for all students
- Evaluating school-, district-, and/or state-level programs
- Informing stakeholders (teachers, school administrators, district administrators, DESE staff, parents, and the public) on the status of the progress toward meeting academic achievement standards of the state
- Meeting the requirements to measure Adequate Yearly Progress by NCLB
- Meeting the requirements of the state's accountability program, Missouri School Improvement Program (MSIP) (p. 10)

Furthermore, DESE continues to encourage program evaluation studies to support the uses of the MAP scores and is currently working with the Assessment Resource Center to conduct a study on consequential validity.

Reliability. Reliability studies with the MAP tend to focus on the hand scoring process for constructed response items. When human scorers are employed, undeniably, at least some level of subjectivity is present. MAP scorers though are carefully selected and trained. Scorers are also monitored for accuracy as approximately 5% of items are scored by a second reader to establish inter-rater reliability. According to the 2008 MAP Technical Report, the third-grade communication arts assessment had 11 constructed response items scored by hand. When second readers score an item, the rate of a perfect scoring match varies from 68% to 99% (dependent on the item number) with a mean of 86.3%. At least 99% of the time, items were scored with adjacent accuracy—that is, if one rater judged an item response with a score of four, the second rater scored the item response with either a three or a four (adjacent scores).

The test developers also conducted analyses on operational data. The analysis included classical item statistics as well as item response theory to help ensure that MAP items are accurately discriminating between levels of achievement and that overall test results are generalizable for the intended purposes (2008).

Data Collection Procedures

Data was not sought directly from students for the purpose of this study. Rather, it was existing data already gathered for other educational purposes. The first step though in retrieving the existing data components for this study was to obtain permission to access the school district's student database to obtain all relevant data for the two cohort populations (Appendix A). Once permission had been granted, the data was retrieved by district assessment personnel using district-created data retrieval software. When accessing and testing the data, students were identified only by a student number in order to protect anonymity. Next, a computer-driven data sort was conducted to determine which students have both eligible NAI scores and eligible MAP scores. Students with only zero or one, but not two eligible scores were removed from the sample population. From Cohort 1, 564 students were removed for this reason and in Cohort 2, 508 students were determined to be ineligible for not having both scores present. After the eligible students were identified, statistical computer software was used to test the data as presented below.

Data Analysis

Two populations were studied, consisting of two cohorts of students as described previously. The first cohort included 794 students who took the NNAT as kindergartners in 2004 and the MAP test as third graders in 2008. The second cohort was very similar being made up of 795 students who took the NNAT as kindergartners in 2005 and the MAP test as third graders in 2009. Each of these cohorts was treated as a population from which statistical relationships could be identified. Since items on the MAP test can vary from year to year, the data from the two cohorts were considered separately. The first step taken was to calculate descriptive statistics regarding the populations to characterize the data. The data was segmented by demographic variables including gender, race, and income level. Gender and race indications were supplied by parents upon registering their children for school. With respect to race, students were classified in one of the following six racial/ethnic categories: Asian/Pacific Islander, Black, Hispanic, American Indian, White, or Other/non-response. Income level was designated by the criteria of a student being eligible for free or reduced price school lunch. Low income students will be those who qualified for free or reduced price lunches at the time of taking the MAP test in third grade. Invariably, this criterion is not completely accurate as a descriptor of low income, but it is the only such indicator available. Qualifying for the free and reduced price lunch program only takes place if students' parents choose to complete the required paperwork at the school. Thus, some students who live in low-income homes may not be accurately designated as low-income students in the data pool. Also, it is possible that a family's income status could change just prior to MAP testing in third grade. These children would then be tagged as low income even though they would not have been considered low-income students during their first four years of schooling. Conversely, a family could have been in a low-income situation throughout the first nine years of a child's life, but lost the free or reduced lunch status just prior to MAP testing due to changing family circumstances. Clearly, there are several considerations that must be made when interpreting the data from this subgroup and generalizations should be made with caution.

For each population, the mean, standard deviation, skewness, and kurtosis of the NNAT and MAP scores was calculated. These statistics along with a visual inspection of scatterplots and histograms of both scores were used to characterize the distribution of the data. The mean and standard deviation was also calculated for all demographic subgroups to understand the differences between them and control for those differences in the later analysis.

Once the data was characterized with descriptive statistics, it was evaluated through correlation and regression analysis to establish statistical relationships between the variables of interest. The first and most important relationship analyzed was between the NAI kindergarten score and the third-grade MAP scores for each population studied. Because the goal was to assess whether the kindergarten NAI scores can be used to infer the later MAP scores, a Pearson correlation analysis will be used. It was not necessary to determine whether the NAI score causes the later result. If the correlation existed, the relationship was identified and the inferential capability of the NAI score was validated. For each population, a correlation coefficient (*r*) was calculated. An *r* value below zero indicated that the anticipated correlation does not exist. Subject to judgment is the interpretation of a positive *r* value. Hinkle et al. (2002) acknowledge that no definitive rules exist for interpreting the *r* value, but provide general rules of thumb for interpreting *r* in the behavioral sciences (defined in Chapter 1). Naglieri (2000b) however used Cohen's (1977) interpretive guidelines which assumed more significant correlations with smaller *r* values. The discrepancy over the significance levels of the *r* value interpretation is addressed in the discussion section of the study, but to maintain consistency with Naglieri's study, an *r* value below 0.3 will indicate that a weak correlation exists that may not be sufficient for inferential application. Under this condition, the NAI score will not be considered sufficient information on which to infer a future MAP score. An *r* value above 0.3 will be considered a sufficiently strong correlation to draw inferences about the MAP test score from the NAI score. Anything above 0.5 is a strong correlation from which reliable inferences can be made. Both cohorts' *r* values will be calculated independently. To reach a conclusion that a correlation is sufficient for drawing inferences, both populations' *r*-values should exceed the 0.3 threshold level. The square of the *r*-value (r^2) was also calculated and is interpreted as the percent of variation in students' third-grade MAP scores that can be explained by variation in the kindergarten NAI score.

After the population level conclusions are drawn, analysis was conducted to ascertain the consistency of the relationships across demographic subgroups including: gender, income, and ethnicity. For each subclass under consideration, a separate regression analysis was performed. For example, for the gender variable, the following regression construction was employed: MAP = $b_0 + b_1$*NAI + b_2*Male + b_3*MaleNAI + e, where Male is a variable that takes a value of 1 if the student is male and 0 if the student is female, and MaleNAI is the interaction variable—that is, the product of the NAI score and the Male variable. For purposes of the regression analysis, homoskedasticity of the data was assumed. That is, it was assumed that across all sub-classes of the population and levels of NAI score, the variance of MAP scores is similar. This is a reasonable expectation that was validated by a cursory review of the standard deviations of the subclasses to screen out any obvious outliers for whom the assumption clearly did not hold.

The F-statistic of the regression equation was first evaluated to determine whether the regression was statistically significant at the 95% confidence level. If that threshold was met, then the individual variables' *t*-statistics were evaluated at the 95% confidence level as well to determine whether the interaction variable (MaleNAI) was statistically significant. If the interaction variable was not significant, it was removed from the regression and Model 2 was executed. Model 2 had only the NAI variable and the subgroup variable (in this case, Male). The equation for Model 2 was MAP = b_0 + b_1*NAI + b_2Male + e. If in this regression the Male variable was significant, it was left in the equation to control for base differential in expected MAP achievement levels of males from females, where the value of the coefficient was interpreted as the average differential in score between groups (gender in this case). If neither the Male variable in Model 2 nor the MaleNAI variable in Model 1 was significant, the results were disregarded.

If the interaction variable (MaleNAI) was significant in Model 1, the coefficients were analyzed to determine the slope of the regression lines which indicated how much variation could be expected with changes in NAI. If the variable*NAI did not have a statistically significant ($p < .05$) coefficient, then the power of the NAI score to draw inferences about future MAP scores was no different for the subgroup variable (e.g. Male) than the opposite group (Female). If the interaction variable was not significant in Model 1, but the subgroup variable (e.g. Male) was significant in regression Model 2, then the coefficients and slopes of the parallel regression lines were analyzed.

The r^2 statistic of each regression equation was also evaluated and reported. This indicates the percent of variation in students' MAP scores that can be expected to be explained by the NAI scores and the demographic information about the students in question.

For each demographic variable considered, a similar regression to that described for males was run and comparable analyses were performed. Tables of results were compiled that provided quantitative evidence of expected performance differentials between different demographic groups as well as any information regarding the relative power of kindergarten NAI score in inferring third-grade MAP scores for those same groups.

Chapter 4

Data Analysis

The presentation of the results begins with a characterization of the data in which the two cohort samples are described statistically. The overall NNAT and MAP statistics are presented which are then followed by the regression analysis. The first section of the regression analysis describes the overall correlational relationship between NAI and achievement at which time the cohorts' r and r^2 values are revealed. Then, the regression statistics for each subgroup are presented in tables with indication of statistical significance. When the subgroup results are statistically significant, line graphs are presented with predicted MAP scores given an NAI score based on the coefficients produced by the regression calculations. For each of the seven subgroups, the regression results from both cohorts are presented side by side for comparison.

Characterization of the Data

The study included two cohorts with a combined total of 1589 students. The students were 51.5% male and 48.5% female. The racial makeup of the students was 8.6% African American, 8.7% Asian, 2.4% Hispanic, <1% Native American, 77.4% White, and 2.9% were not identified by race. Thirteen percent of the students qualified for free or reduced-price lunch (FRL), 1.1% spoke English as a second language (ESL) and 18.9% had Individualized Education Plans (IEPs) to accommodate identified disabilities. The two cohorts were of similar size and demographic makeup as depicted in Table 4.1 below.

Table 4.1

Demographic Characteristics of the Sample

	Cohort 1	Cohort 2
	Third-grade cohort: 2008	Third-grade cohort: 2009
Total students	794	795
Male	416	402
Female	378	393
African American	66	70
Asian	60	78
Hispanic	15	23
Native American	1	0
White	627	603
No race indicated	25	21
Free/reduced	107	102
ESL	8	10
IEP	134	166

Nonverbal Ability Index. Initial calculations on the data in Cohort 1 revealed a Nonverbal Ability Index (NAI) mean score of 108.4 with a standard deviation of 20.7. The scores had a skewness of -0.29 and kurtosis of -0.81. Similarly, the mean score in Cohort 2 was 107.0 with a standard deviation of 21.3. Skewness was -0.15 and kurtosis was -0.96. Specific statistics for the subgroups in both cohorts are presented in Table 4.2.

A visual inspection of the NAI histogram for Cohort 1 revealed a distribution of data with higher frequencies from 110 to 130 which is above the normed average (100) for this assessment. Similarly, in Cohort 2, the distribution of scores was also more heavily weighted above 100. The histograms do not present any other significant abnormalities. The NAI distributions of data for both cohorts are presented in Figures 4.1 and 4.2.

Table 4.2

Nonverbal Ability Index Statistics

	Cohort 1				Cohort 2			
	M	Max	Min	*SD*	*M*	Max	Min	*SD*
Total	108.4	150	57	20.6	107.0	150	54	21.3
Males	107.0	149	57	21.0	107.7	150	54	20.7
Females	109.8	150	65	20.1	106.3	150	61	21.8
FRL	94.3	147	61	20.6	93.1	136	54	19.4
Non-FRL	110.5	150	57	19.7	109.1	150	58	20.8
White	110.2	150	57	19.6	108.6	150	61	20.3
Black	88.4	124	62	17.0	89.8	148	54	20.5
Hispanic	94.0	126	72	17.2	99.4	133	61	21.4
Asian	113.5	145	66	20.6	112.7	148	61	21.8
Native Am.	95.0	95	0	0.0	0.0	0	0	0.0
IEP	97.6	149	57	22.4	101.6	148	54	21.7
Non-IEP	110.6	150	61	19.5	108.4	150	61	20.9
ESL	89.7	136	61	26.1	98.0	130	71	18.7
Non-ESL	108.6	150	57	20.5	107.1	150	54	21.3

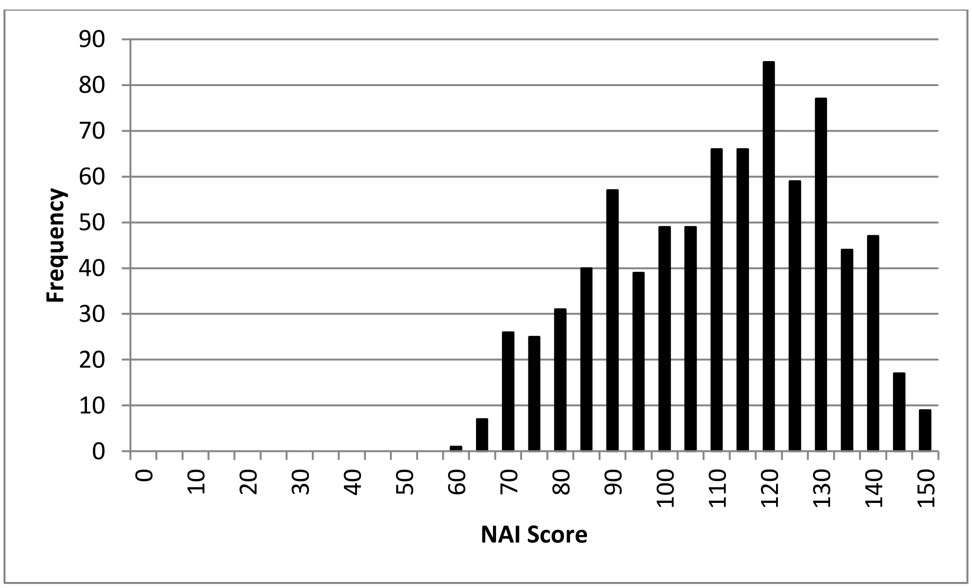


Figure 4.1. Nonverbal Ability Index histogram, Cohort 1.

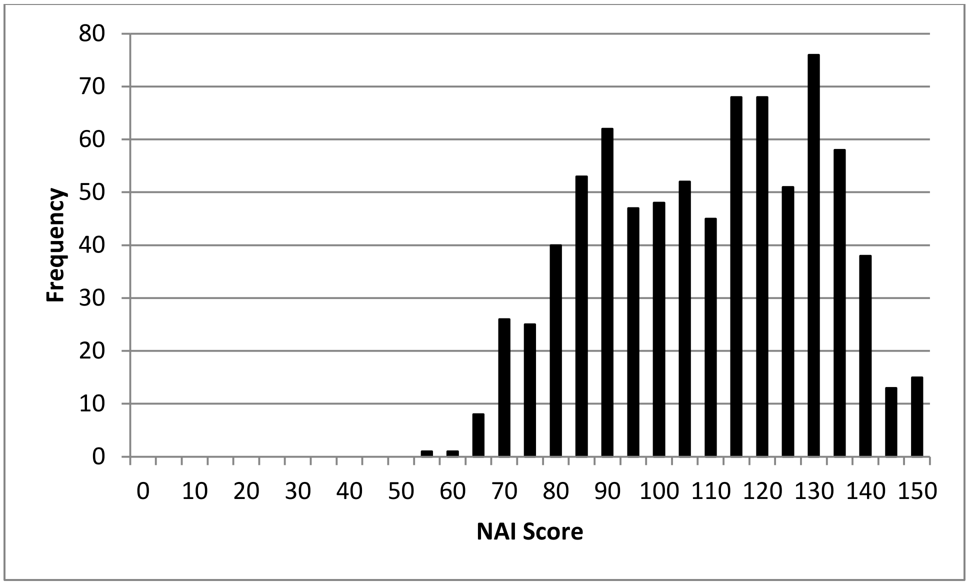


Figure 4.2. Nonverbal Ability Index histogram, Cohort 2.

Missouri Assessment Program. Cohort 1's MAP scores had a mean of 651.6 with a standard deviation of 34.4. The scores had a skewness of -0.40 and kurtosis of -1.91. Cohort 2's MAP scores had a mean score of 655.4 which was just slightly higher (3.8) than that of Cohort 1. The MAP standard deviation for Cohort 2 was 34.0 with a skewness of -0.28 and kurtosis of 1.22. Descriptive statistics for both cohorts' MAP scores (including subgroups) are presented in Table 4.3.

A visual inspection of the MAP histograms (Figures 4.3 and 4.4 below) revealed more normal distributions than the NAI data presented above. The MAP distributions exhibited normal characteristics from end to end.

Table 4.3

Missouri Assessment Program Statistics

	Cohort 1				Cohort 2			
	M	Max	Min	*SD*	*M*	Max	Min	*SD*
Total	651.6	762	492	34.3	655.3	790	512	34.0
Males	645.5	762	492	34.7	651.4	769	512	33.1
Females	658.3	761	548	32.6	659.4	790	559	34.4
FRL	624.0	701	493	36.2	627.5	707	512	37.8
Non-FRL	655.8	762	492	31.9	659.4	790	532	31.3
White	655.5	762	492	30.9	659.6	790	532	30.0
Black	612.2	657	493	31.6	620.2	707	545	38.9
Hispanic	619.0	693	496	52.6	642	733	512	44.7
Asian	662.2	761	580	34.5	659.3	730	559	35.5
Native Am.	640.0	640	0	0.0	0.0	0	0	0.0
IEP	618.1	712	492	35.3	631	725	512	36.1
Non-IEP	658.4	762	545	29.8	661.7	790	545	30.3
ESL	610.3	679	533	42.6	605.9	642	578	23.1
Non-ESL	652.0	762	492	34.0	655.9	790	512	33.6

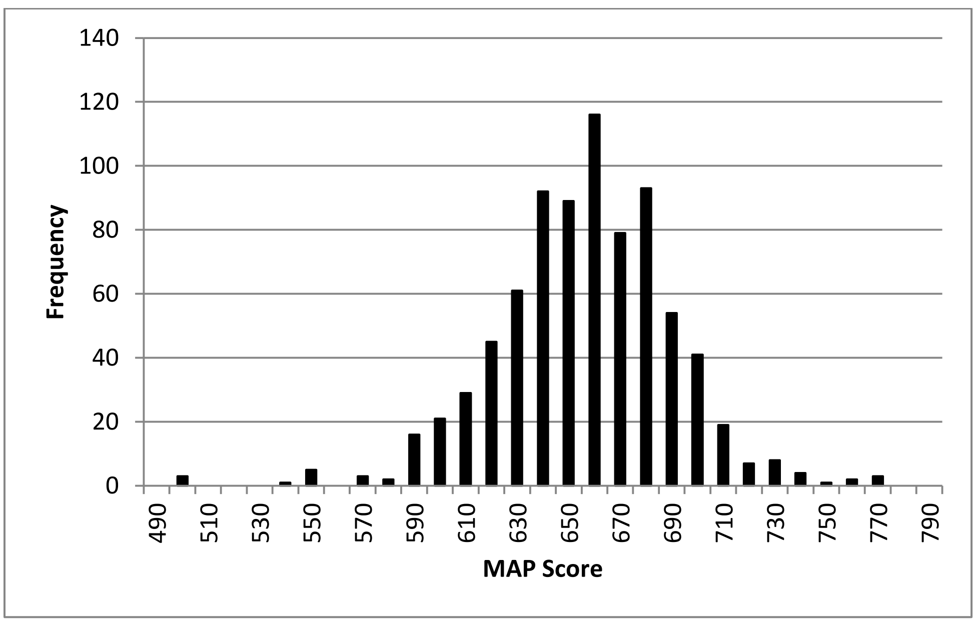


Figure 4.3. Missouri Assessment Program histogram, Cohort 1.

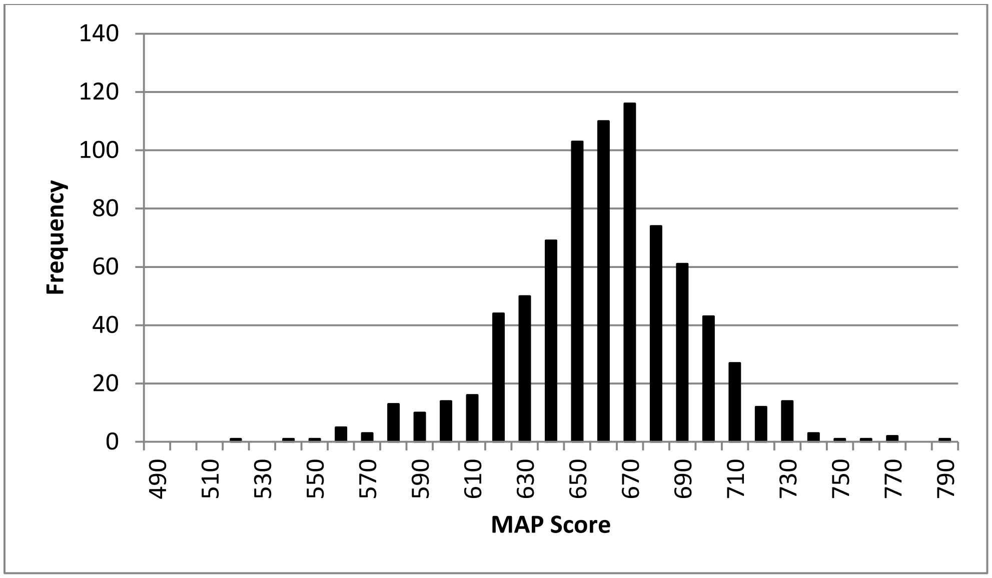


Figure 4.4. Missouri Assessment Program histogram, Cohort 2.

The initial inspection of the scatterplots (Figures 4.5 and 4.6) for NAI (x-axis) and MAP (y-axis) showed a clear linear relationship between the two scores. Students who scored lower on the NNAT test also tended to score lower on the MAP test nearly four years later. A clear correlation existed. The upward slope of the trend line suggested that the eventual regression calculations will be able to draw significant conclusions from the data. The width of the spread suggested that other factors are at play in determining eventual success on the MAP test, but the trend line indicated that there is some correlation and predictability of MAP which can be inferred from NAI scores. The scatterplot distribution is very similar between the two cohorts suggesting that the two assessments likely have a consistent relationship from year to year even though there was nearly a four-year time span between administering the two tests.

With the exception of a few outliers, the spread of data along the trend lines for both cohorts is quite consistent from top to bottom suggesting that variance from the mean will remain consistent for both high scorers and low scorers. This even, parallel variance band suggests homoscedasticity of data.

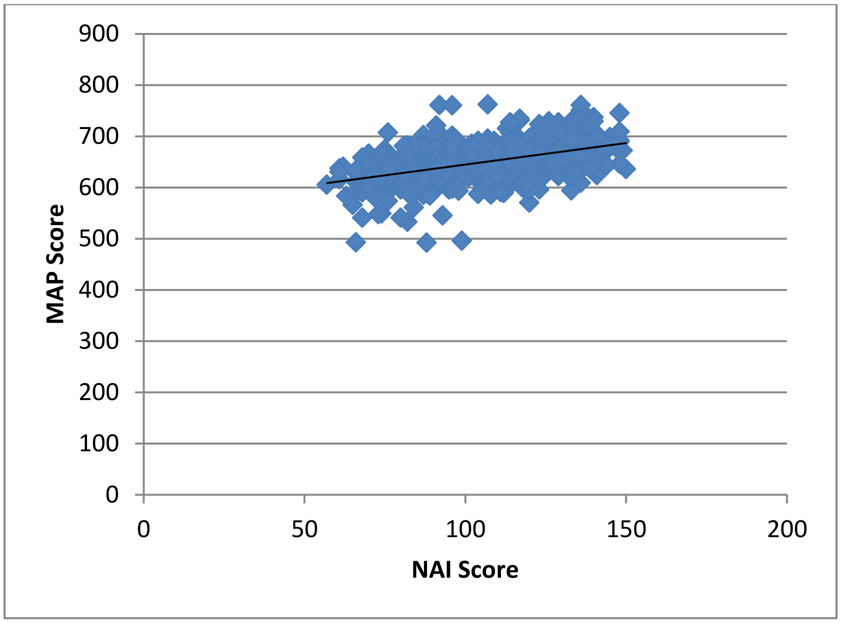


Figure 4.5. Nonverbal Ability Index and Missouri Assessment Program scatterplot, Cohort 1.

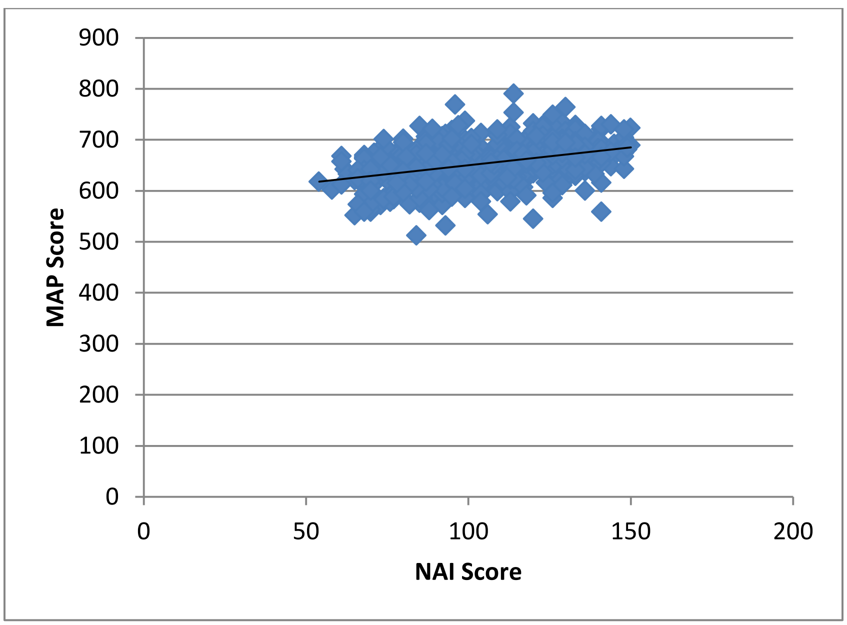


Figure 4.6. Nonverbal Ability Index and Missouri Assessment Program scatterplot, Cohort 2.

Regression Analysis

A Pearson correlation analysis determined to what extent third-grade MAP scores could be inferred from kindergarten NAI scores. In Cohort 1, the total population *r* value was .50. Based on Naglieri's use of Cohen's (1977) interpretive guidelines, this *r* value is a sufficiently strong correlation to draw inferences about the MAP test score from the NAI score. Similar results were observed in Cohort 2. In Cohort 2, the total population *r* value was .44.

The regression model as presented in Chapter 3 was MAP = $b_0 + b_1$*NAI + *e*. The regression statistics for both cohorts are presented in Table 4.4.

Table 4.4

Nonverbal Ability Index Regression Statistics

Variable	Cohort 1	Cohort 2
Intercept	561.01**	580.00**
NAI	0.84**	.70**
r	.50	.44
r^2	.25	.19

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

In addition to the study of the overall relationship between the NAI test and MAP scores, additional tests were conducted to assess the impact of various demographic variables on the MAP score and the relationship between the two test scores. Those demographic variables included the study of gender, race (White, Black, Asian, and Hispanic), income (as represented by participation in the free and reduced-price lunch program), and presence of an individual educational plan (IEP). Two reported subgroups revealed very low *n* counts in both cohorts. From this point on, the Native American (*n* = 1, 0) and ESL (*n* = 8, 10) populations, were not analyzed as individual subgroups in this study due to the limited data. However, those students remained in the overall samples and were included as members of other subgroups (such as FRL) when appropriate. For the remainder of the student subgroups, similar analysis was done of the scatterplots of the NAI scores. As a general observation, within each demographic group and for each cohort, the scatterplot showed an apparent linear relationship between MAP score and NAI score. Moreover, the variance bands were generally consistent across the range of MAP scores, again mitigating any potential concerns over heteroskedasticity of the data. The scatterplots of each demographic group are presented in Appendixes B through Q.

For each demographic group a regression was completed to determine the effect of the subgroup variable and the interaction of NAI and the subgroup variable. For example, male students were studied using the regression equation: MAP = $b_0 + b_1$*NAI + b_2*Male + b_3*MaleNAI + e. In this equation, the Male variable is a binary variable that takes the value 1 if the student is a male and 0 if the student is female. The MaleNAI variable is the product of the NAI score of the student and the Male variable which represents the interaction between the two variables.

In situations when the interaction variable (e.g. MaleNAI) was not significant, it was removed from the equation and a second regression, Model 2, was executed. The second equation was MAP = $b_0 + b_1$*NAI + b_2Male + e.

Regression statistics: Male versus female. The MaleNAI interaction variable was not statistically significant in Cohort 1, so regression Model 2 was executed. The Male variable then tested to be significant. In Cohort 2, the interaction variable was significant when tested in Model 1. The results of the regression analysis for the gender variable are presented in Table 4.5.

Table 4.5

Regression Analysis for Gender (Male)

	Cohort 1		Cohort 2
Variable	Model 1	Model 2	Model 1
Intercept	564.23**	573.63**	573.51**
NAI	.86**	.77**	.81**
Male	5.85	-10.24**	13.69
MaleNAI	-0.15		-0.21*
r^2	.27	.26	.21

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

The calculations presented above produced the regression lines below. For cohort 1, regression Model 2 is graphed showing the statistically significant male variable line. For Cohort 2, Model 1 is graphed showing the interaction of the Male variable and the NAI variable (Figures 4.7 and 4.8).

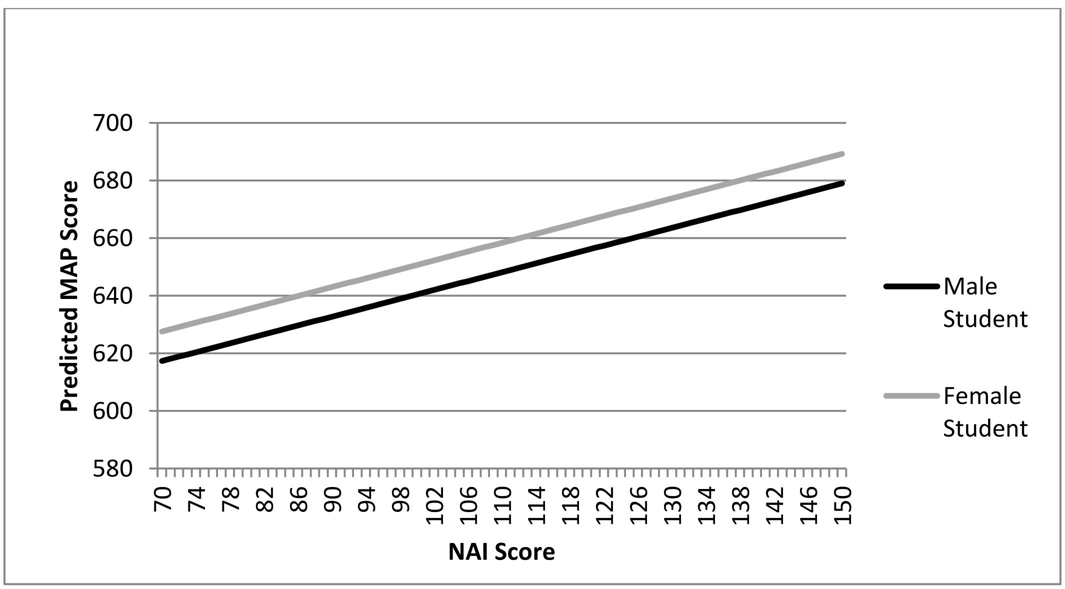


Figure 4.7. Cohort 1: Missouri Assessment Program predictions for male and female students, using Nonverbal Ability Index scores.

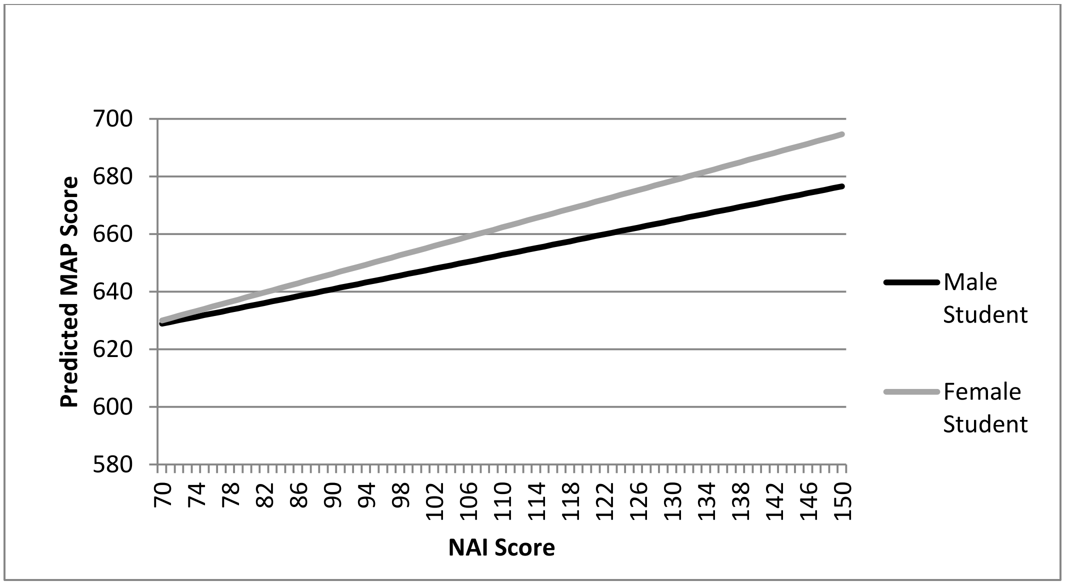


Figure 4.8. Cohort 2: Missouri Assessment Program predictions for male and female students, using Nonverbal Ability Index scores. Intersection point is at Nonverbal Ability Index = 71.

Regression statistics: White versus non-White. Regression Model 1 produced statistically significant results in both cohorts when the White variable was tested (*n* = 627, 603). The statistics are presented in Table 4.6 and the regression lines with intersection points are presented in Figures 4.9 and 4.10 below.

Table 4.6

Regression Analysis for White Variable

	Cohort 1	Cohort 2
Variable	Model 1	Model 1
Intercept	522.31**	549.36**
NAI	1.13**	.91**
White	63.87**	49.40**
WhiteNAI	-0.50**	-0.35**
r^2	.28	.23

*p < .05. **p < .01.

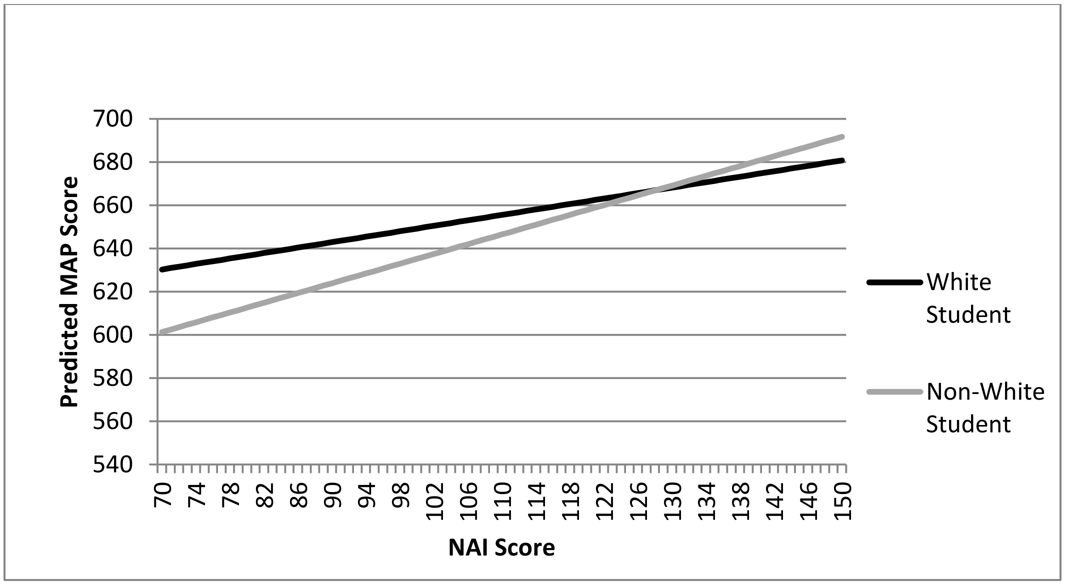


Figure 4.9. Cohort 1: Missouri Assessment Program predictions for White and non-White students, using Nonverbal Ability Index scores. Intersection point is at Nonverbal Ability Index = 128.

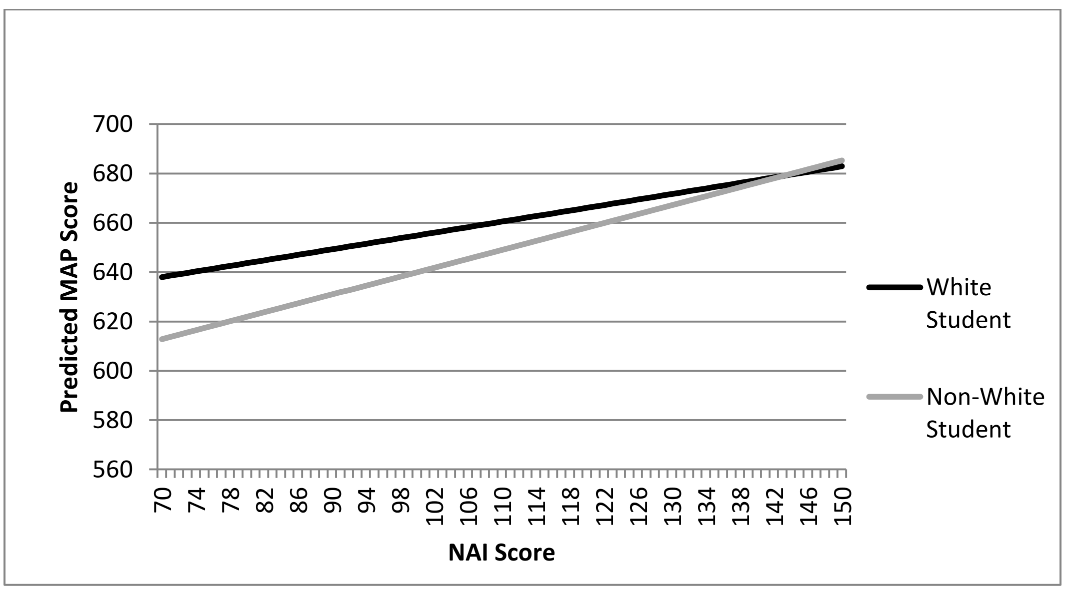


Figure 4.10. Cohort 2: Missouri Assessment Program predictions for White and non-White students, using Nonverbal Ability Index scores. Intersection point is at Nonverbal Ability Index = 143.

Regression statistics: Black versus non-Black. Similar to the White group, the Black groups (n = 66, 70) produced statistically significant coefficients for the BlackNAI variables in both cohorts. The regression statistics (Table 4.7) and line graphs (Figures 4.11 and 4.12) are presented below.

Table 4.7

Regression Analysis for Black Variable

	Cohort 1	Cohort 2
Variable	Model 1	Model 1
Intercept	575.72**	596.24**
NAI	.72**	.58**
Black	5.80	-63.23**
BlackNAI	-0.35*	0.40*
r^2	.28	.25

*p < .05. **p < .01.

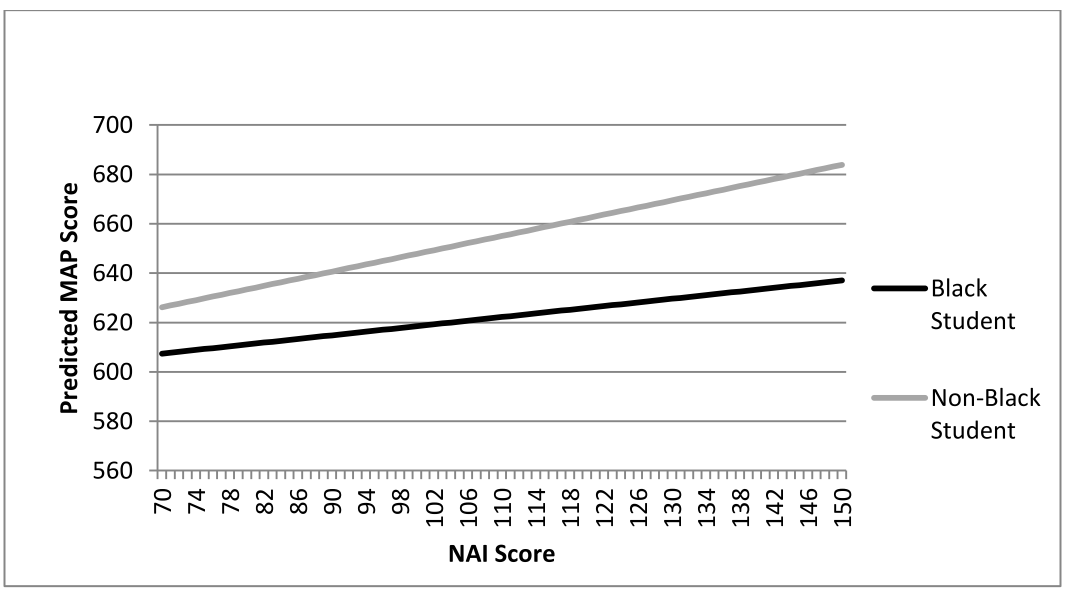


Figure 4.11. Cohort 1: Missouri Assessment Program predictions for Black and non-Black students, using Nonverbal Ability Index scores.

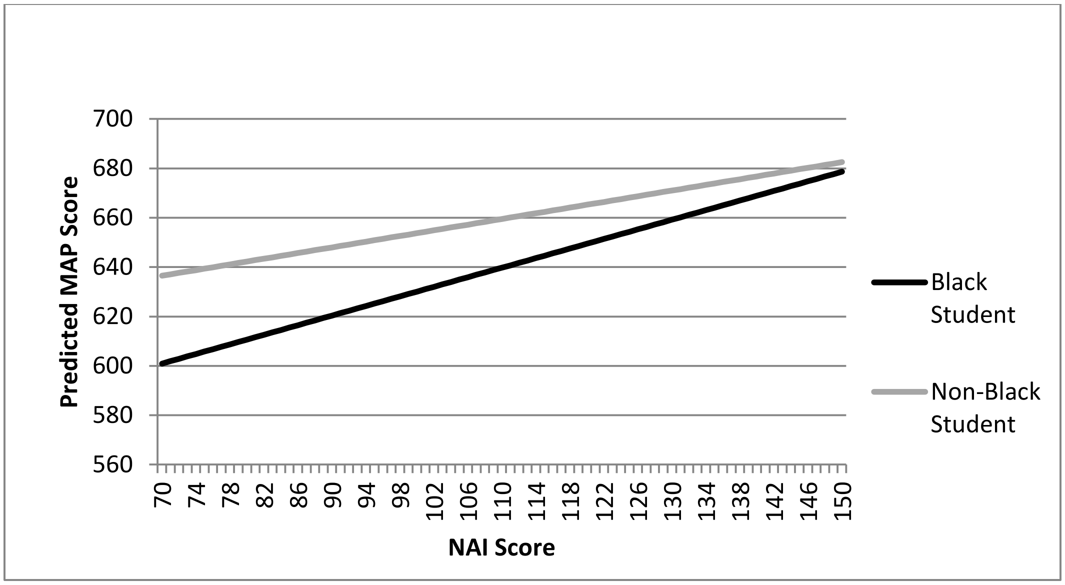


Figure 4.12. Cohort 2: Missouri Assessment Program predictions for Black and non-Black students, using Nonverbal Ability Index scores.

Regression statistics: Asian versus non-Asian. The Asian group in Cohort 1 (*n* = 60) produced significantly significant coefficients for all variables tested. The AsianNAI interaction variable in Cohort 2 (*n* = 78) was not significant, therefore regression Model 2 was executed. Unlike the gender group, when the interaction variable was removed from the equation, the model still did not produce a statistically significant Asian variable. The regression statistics are presented in Table 4.8 and the graph for Cohort 1, Model 1 is presented in Figure 4.13.

Table 4.8

Regression Analysis for Asian Variable

	Cohort 1	Cohort 2	
Variable	Model 1	Model 1	Model 2
Intercept	559.04**	578.33**	580.47**
NAI	.85**	.72**	.70**
Asian	58.11**	21.70	.03
AsianNAI	-0.44**	-0.19	
r^2	.26	.19	

*p < .05. **p < .01.

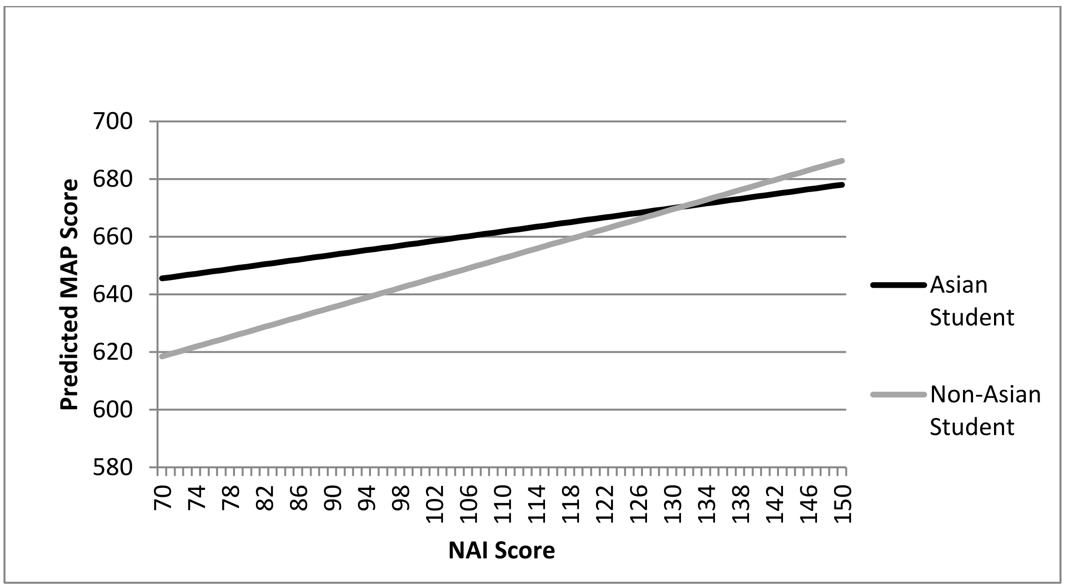


Figure 4.13. Cohort 1: Missouri Assessment Program predictions for Asian and non-Asian students, using Nonverbal Ability Index scores. Intersection is at Nonverbal Ability Index score = 131.

Regression statistics: Hispanic versus non-Hispanic. The Hispanic group in Cohort 1 produced four statistically significant coefficients even though the group had a relatively low *n* (15). The regression test for the Hispanic group in Cohort 2 ($n = 23$) did not produce a significant interaction variable. When the interaction variable was removed, the Hispanic variable in Model 2 was also not statistically significant. The results are presented in Table 4.9 and the regression graph for Cohort 1 follows (Figure 4.14).

Table 4.9

Regression Analysis for Hispanic Variable

	Cohort 1	Cohort 2	
Variable	Model 1	Model 1	Model 2
Intercept	562.80**	581.42**	581.14**
NAI	.82**	.69**	.70**
Hispanic	56.00**	-17.24	-8.33
HispanicNAI	-0.78**	0.09	
r^2	.25	.19	.19

*p < .05. **p < .01.

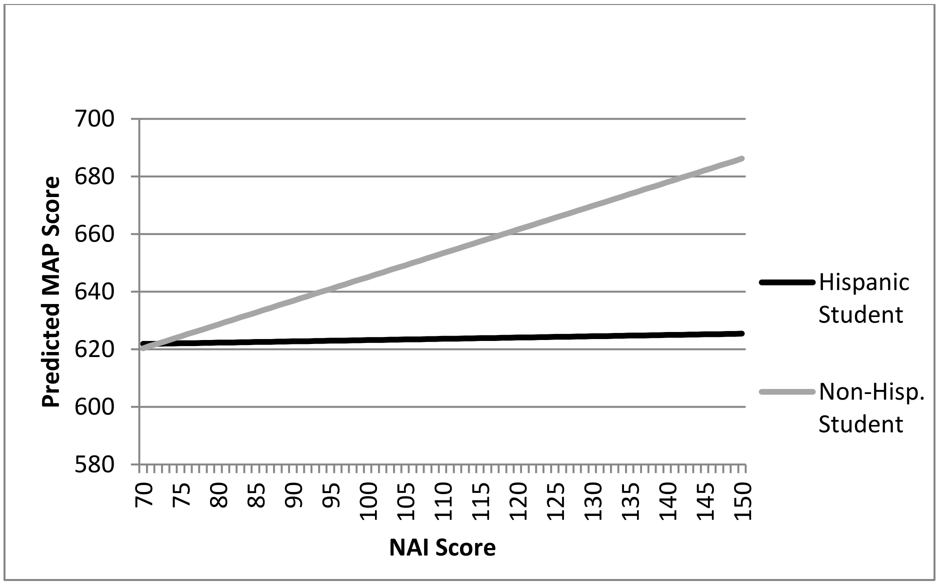


Figure 4.14. Cohort 1: Missouri Assessment Program predictions for Hispanic and non-Hispanic students, using Nonverbal Ability Index scores. Intersect at Nonverbal Ability Index score = 72.

Regression statistics: Individual education plan versus no individual education plan. The IEP groups in both cohorts (*n* = 134, 166) failed to produce statistically significant interaction variables (IEPNAI). As a result, Model 2 was executed for both cohorts and both resulted in significant IEP variables. The statistics are presented in Table 4.10 and the two cohorts' graphs for regression Model 2 are shown in Figures 4.15 and 4.16.

Table 4.10

Regression Analysis for Individualized Education Plan Variable

	Cohort 1		Cohort 2	
Variable	Model 1	Model 2	Model 1	Model 2
Intercept	584.14**	584.97**	597.57**	592.99**
NAI	.67**	.66**	.59**	.63**
IEP	-26.71*	-30.05**	-46.26**	-26.50**
IEPNAI	-0.03		.19	
r^2	.34	.34	.29	.29

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

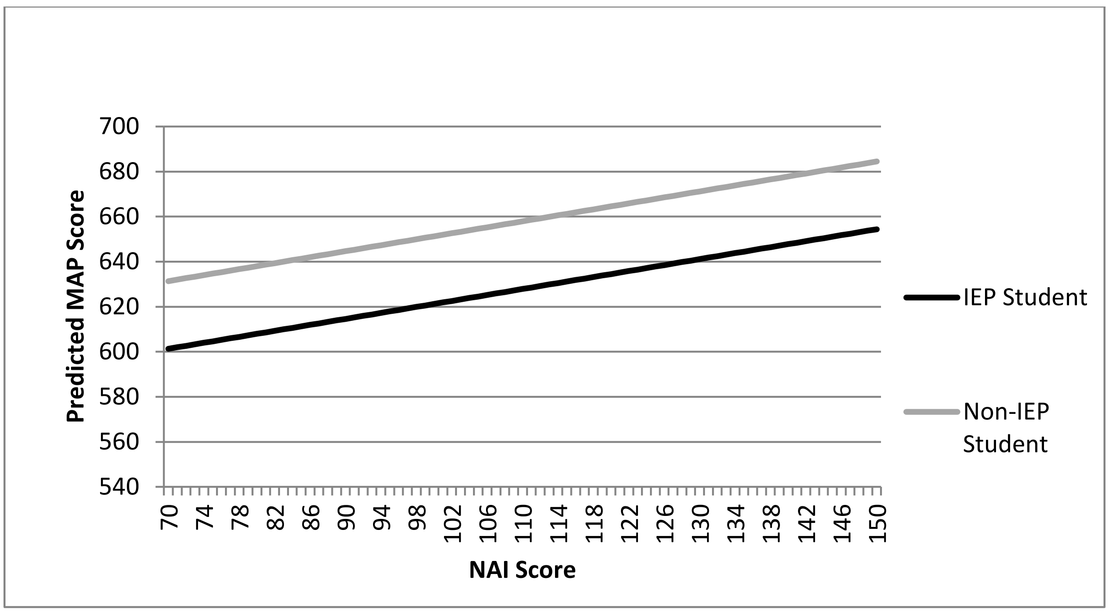


Figure 4.15. Cohort 1: Missouri Assessment Program predictions for students with and without an individual education plan, using Nonverbal Ability Index scores.

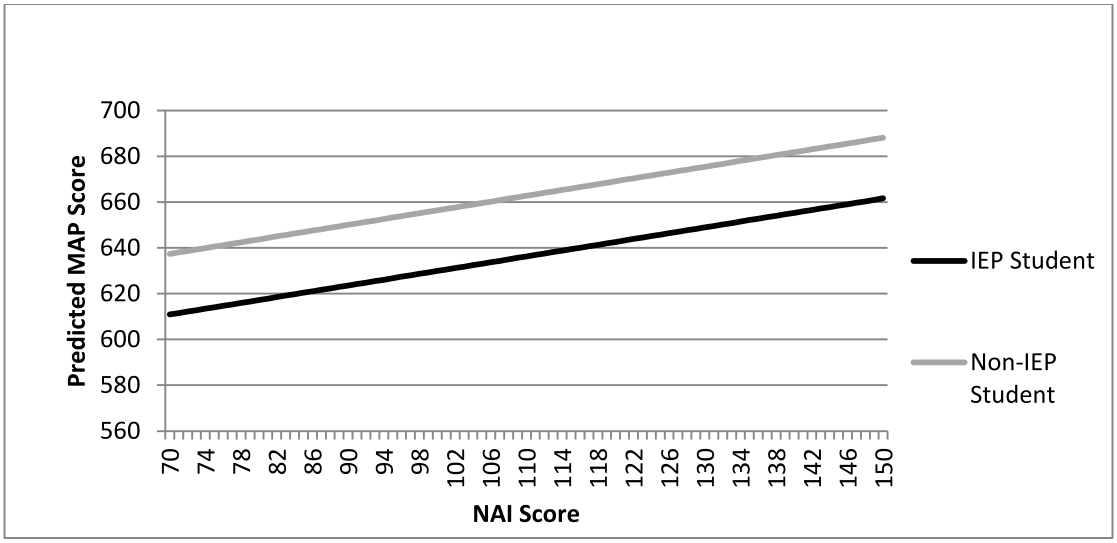


Figure 4.16. Cohort 2: Missouri Assessment Program predictions for students with and without an individual education plan, using Nonverbal Ability Index scores.

Regression statistics: Free or reduced-price lunch versus no free or reduced-price lunch. The FRL group in Cohort 1 did not produce a statistically significant interaction variable (FRLNAI). As a result, the variable was removed and Model 2 was executed. Model 2 produced a significant FRL variable as presented in Table 4.11 and Figure 4.17. Cohort 2 however produced a statistically significant interaction variable which is presented visually in Figure 4.18.

Table 4.11

Regression Analysis for Free or Reduced-Lunch Variable

	Cohort 1		Cohort 2
Variable	Model 1	Model 2	Model 1
Intercept	582.33**	578.20**	597.57**
NAI	.67**	.70**	.57**
FRL	-47.90**	-20.73**	-59.86**
FRLNAI	0.28		.40*
r^2	.28	.28	.24

*p < .05. **p < .01.

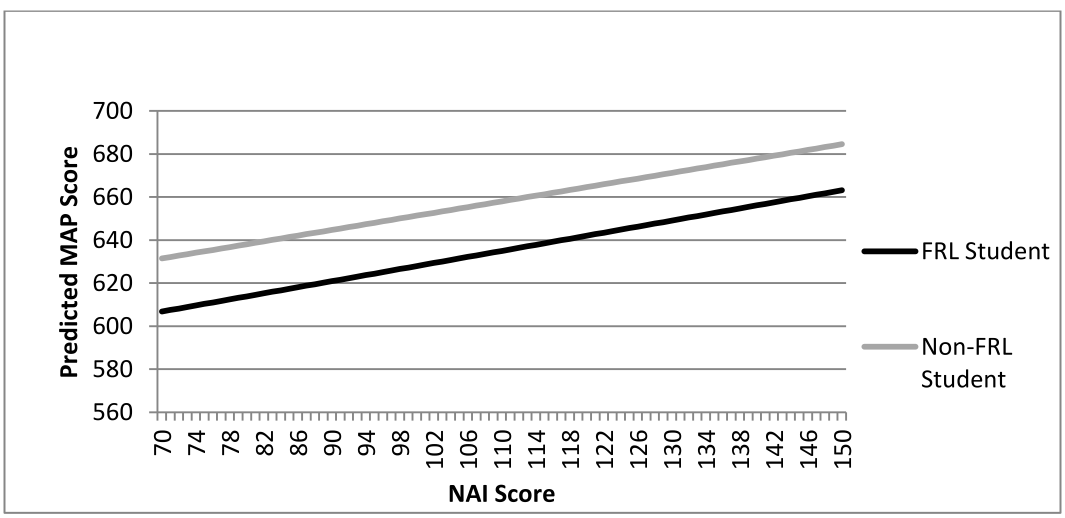


Figure 4.17. Cohort 1: Missouri Assessment Program predictions for students with and without free or reduced-price lunch, using Nonverbal Ability Index scores.

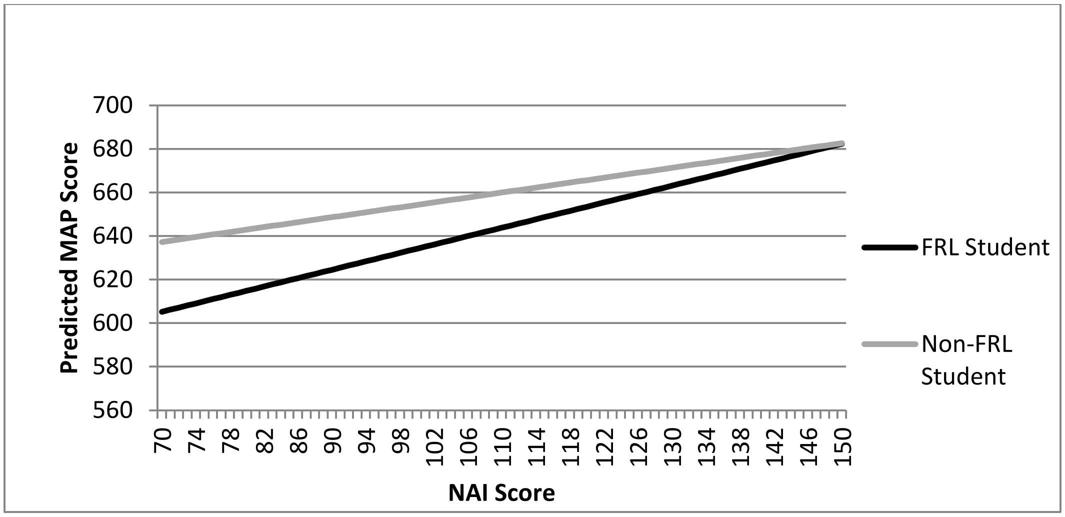


Figure 4.18. Cohort 2: Missouri Assessment Program predictions for students with and without free or reduced-price lunch, using Nonverbal Ability Index scores.

Chapter 5

Conclusions and Implications

The presentation of conclusions and implications for this study begins with a comparison of the revealed correlational values with those of other similar studies presented in Chapter 2. Following this comparison, the subgroup regression results are analyzed beginning with a discussion of implications of the variables testing statistically significant or insignificant in both regression Models 1 and 2. Then, the effect of the coefficients in each of the seven subgroups is analyzed including some suppositions of how certain variables may be influencing the coefficients of other variables in separate regression models. Next, analysis of the subgroups' r and r^2 values is presented which leads into the final discussion of the implications these relationships have for educators and policymakers. Through this discussion, the limitations of the study are revisited and suggestions for further study are offered.

Correlation Conclusions

The primary purpose of this research was to determine whether or not the levels of correlation found by Naglieri and Ronning (2000a) between the NNAT and the SAT-9 would be consistent with the levels of correlation between the NNAT and MAP (communication arts) when the assessments were administered nearly four years apart. Naglieri and Ronning found r values between .49 and .56, thus it was hypothesized in this study that the r values between NNAT and MAP would be within that range. The hypothesis was proven true given the analysis of the data in Cohort 1 (r = .50) but did not hold true for the second cohort (r = .44). Though the second cohort did not meet the .49 threshold, the r value was only .05 less than the bottom end of the target range presented in the hypothesis. Again, the *r* values obtained by Naglieri and Ronning (2000a) were calculated using data from two assessments (one testing nonverbal ability and the other testing reading achievement) that were given in the same year of schooling. The current study examined data in which the nonverbal ability scores were obtained in kindergarten, and then statistically compared with reading and writing scores obtained near the end of third grade. Even with the large gap between the administrations of the assessments, the *r* values were generally within .1 of those that were found when the tests were given nearly concurrently.

Table 5.1

Correlation Values for Missouri Assessment Program and Nonverbal Ability Index

Population	Cohort 1	Cohort 2
r	.50	.44
r^2	.25	.19

Using the interpretive guidelines employed by Naglieri and Ronning (2000b) which were previously suggested by Cohen (1977), *r* values may be characterized as small if greater than .10, medium if greater than .30, and large if greater than .50. Using these standards, this study found medium to large correlations between nonverbal ability in kindergarten and reading and writing achievement in third grade. But again, these standards are subject to interpretation. Remembering the caution extended by Hinkle et al. (2002), any conclusions drawn with respect to an *r* value must carefully consider the meaning of the variables rather than simply following a set categorical scale. Having said that, those researchers' rule of thumb guidelines suggest that correlations between .30 and .50 are low and correlations between .50 and .70 are moderate in the behavioral sciences. Using those standards, the results of this study would be classified as low to moderate correlations.

Regardless of the categorical labels used to describe the relationship, the findings of this study were quite consistent with those that other researchers recognized. Some of the questions posed early on in the study targeted the general predictive ability of cognitive ability tests to infer eventual success in reading and writing. The *r* values (.50 and .44) are in line with those found by Naglieri and Ronning (2000b). Likewise, they are similar to those of other correlational studies which were presented in detail in chapter 2. Those *r* values are presented again in Table 5.2 below.

Table 5.2

Correlation Values of Previous Studies

Authors	Correlation variables	*r*
Hammill and McNutt (1981)	WISC and reading achievement	.44
Hammill and McNutt (1981)	Stanford-Binet and reading achievement	.46
Stanovich, Cunningham, and Freeman (1984)	Various ability tests and various reading achievement tests	.19 to .76
Stanovich, Cunningham, and Freeman (1984)	General ability and reading achievement at the end of first grade, third grade, and fifth grade	.33, .42, .56
Carver (1990)	Nonverbal ability and reading achievement	.5 (average)

The findings in this study when compared with those depicted above are quite congruent with respect to the *r* values. Even when compared with individually administered assessments such as the WISC or Stanford-Binet, the group-administered

NNAT test is revealing similar correlation statistics when compared with communication arts achievement at the end of third grade.

So the results of this study are consistent with the findings of similar research and the strength of the *r* value can certainly be debated. But rather than further discussing the subjective descriptors for the revealed *r* values (such as low, moderate, or large), it may be more beneficial to consider the meaning of the *r squared* values.

Considering the *r* squared values presented in Table 5.2 above, it can be generally concluded that between 19% and 25% of a child's third-grade achievement in reading and writing can be inferred from their kindergarten nonverbal ability scores. With that in mind, individual researchers and educators will have to determine on their own how confidently they can use the findings presented in this study or that of Naglieri and Ronning (2000a). Likely, it will depend on the nature of their research and what other variables they are considering. However, given the fact that both cohorts in this study revealed that the two variables (NAI and MAP) are statistically related at the 95% confidence level, there are clearly a few implications (both positive and negative) for children and educators that will be discussed throughout the remainder of this chapter.

Regression Conclusions

First though, it is necessary to further investigate the meaning of the results from the subgroup regression analysis. In all cases (each subgroup regression considering the subgroup variable and the subgroup variable times NAI), the overall model was clearly statistically significant, with F-statistics ranging from 63.5 to 137.1. This was a convincing indicator that both the approach and data used were sound. Strong relationships clearly exist between the MAP and NAI scores. Additionally, there is clear statistical evidence of difference between the performances of the various subgroups on MAP. The next question to answer is whether the differences identified are simply differences in the inherent ability of these subgroups on their MAP performance or whether the NAI is a better predictor of MAP for these subgroups, or whether both of these conditions are true.

Cases in which Model 1 revealed a significant interaction variable. The simplest conclusions to draw were in cases where the subgroup interaction variables (e.g. MaleNAI) were statistically significant in regression Model 1. This was the case for the following subgroups: Gender (Cohort 2), White (both cohorts), Black (both cohorts), Asian (Cohort 1), Hispanic (Cohort 1), and FRL (both cohorts).

In these cases, it can be statistically concluded with 95% confidence that the NAI has inferential power with respect to MAP that differs between subgroups. For example, for the Gender variable (Cohort 2), the intercepts and regression slopes are different for males than females. The same can be said for Whites and non-Whites, etc. Discussion and implications related to specific subgroup examples such as these are presented later in this chapter.

Cases in which Model 1 did not reveal a significant interaction variable, but Model 2 revealed a significant subgroup variable. In other cases where the interaction variable (e.g. MaleNAI) did not test to be statistically significant in regression Model 1, the subgroup variable (e.g. Male) was significant in Model 2. This was the case for the following subgroups: Gender (Cohort 1), IEP (both cohorts), and FRL (Cohort 1). It is possible—and in some cases based on the statistics of the model, likely—that when tested in Model 1, the interaction variable (MaleNAI) was not significant because it was bringing similar information to the model as the subgroup variable (Male) creating conflict between the two. In these cases, testing the subgroup variable independently helped to clarify the situation. In the cases listed above, the subgroup variables alone tested to be significant, indicating that there is a natural performance differential between the two groups (males and females), but the NNAT test does not provide additional information regarding likely MAP performance for one group relative to the other. Again, further discussion and implications specific to particular subgroup variables will be presented later in this chapter.

Cases in which neither Model 1 nor Model 2 revealed a significant interaction variable or subgroup variable, respectively. The Asian (Cohort 2) and Hispanic (Cohort 2) groups did not produce statistically significant coefficients in either Model 1 or 2. It was not possible to draw conclusions regarding these two cohort groups.

Subgroup Coefficient Analysis

The coefficients from the individual subgroup regressions provided an interesting vantage point from which to view the subgroups and how their MAP achievement may be inferred by nonverbal ability. And though it was not the focus of this study, it certainly is of further interest to note how the variables impacted the MAP scores both with and without taking into account the impact of the nonverbal ability score variable. The remainder of this section is an analysis of the coefficients for each subgroup variable that tested statistically significant in the regression study. For each significant variable, references are made to the regression lines of the two groups, i.e. the variable group (e.g. Male), and the opposite group (Female). The regression lines presented in Figures 4.7 through 4.18 depict the predicted third-grade MAP scores given a specific kindergarten

NAI score. The slopes of the lines demonstrate how scores from different subgroups behave in the regression models. In some cases, the interaction variable was significant, resulting in non-parallel lines. In the cases where the interaction variable was not significant at the 95% confidence level, only the significant subgroup variable was graphed resulting in parallel regression lines.

Male versus female. Figures 4.7 and 4.8 depict similar relationships that seem to exist between males and females with respect to NNAT and MAP performance. Typically, it appears that on the third-grade communication arts assessment, females score higher than their male counterparts with similar nonverbal ability scores in kindergarten. For Cohort 1, regression Model 2 was presented showing that females tend to score approximately 10 points higher than males with the same NAI score. Because the interaction variable was not significant in Model 1, it is not possible to determine how the predictive capability of NAI changes between higher and lower scorers.

Model 1 however was significant for Cohort 2 and revealed a similar trend as Cohort 1, though as NAI scores were lower, the eventual MAP gap between males and females diminished. At NAI = 71, there was no longer an advantage for females on eventual MAP performance. A kindergartner with the average NAI score (107) for Cohort 2, would be predicted to score 9 points higher on the MAP test if the student were female. It is beyond the scope of this study to determine why females have higher scores, but the data presented here may be of interest to researchers studying the relationship between gender and academic achievement.

White versus non-White. Two interesting characteristics were revealed when comparing the regression graphs of the White and non-White subgroups as presented in

Figures 4.9 and 4.10. First, the intercept for Cohort 1 was 64 points higher (49.4 for Cohort 2) for White students than non-White students. This means in the absence of an NAI score, a White student would be predicted to score 50 to 60 points higher on MAP than a non-White peer. However, the second interesting feature of this regression model can be found in the slopes of the regression lines. Though the White students have a substantially higher intercept values, the slopes of the non-White students' trend lines are steeper. In Cohort 1, for example, a White student with an NAI score of 80 is predicted to score 24 points higher than a non-White student with the same kindergarten NAI score of 80. But if the two students had kindergarten NAI scores of 110, the White student's MAP score would only be 9 points higher. Continuing up the trend line, we would predict that two students with NAI scores of 126 would have even MAP scores. At this level of non-verbal ability in kindergarten, there is no longer a MAP achievement advantage for being in the White group. And, given the regression lines, White students with very high NAI scores above the 130 level will actually score lower on the MAP than their non-White peers of similar ability in kindergarten. Cohort 2 revealed similar regression lines though the intersection point was higher at NAI = 141.

White is the largest racial subgroup in the study (*n* = 627, 603)—its data will receive further analysis in the discussion of subsequent racial subgroups.

Black versus non-Black. The examination of the regression coefficients for the Black subgroup resulted in both cohorts having statistically significant interaction variables (BlackNAI) when tested with regression Model 1. As might be expected following the previous discussion of the White subgroup, the Black groups revealed regression lines that are well below that of the non-Black groups (Figures 4.11 and 4.12).

However, two very different slopes were observed between the two cohorts. In Cohort 1, the slope of the Black regression line was lower than that of the non-Black line. As students scored higher on the kindergarten NNAT test, the non-Black students' eventual MAP achievement increased much more significantly than those in the Black group. When students scored lower on the NNAT in kindergarten, their eventual MAP scores were more similar, though the non-Black group still scored higher.

Curiously, Cohort 2 produced regression lines with rather opposite relationships from Cohort 1. In Cohort 2, Black students still scored lower than their non-Black peers, but as NAI scores increased, the MAP gap significantly decreased. For example, if a student in the Black group and another student in the non-Black group scored 80 on the NNAT, their predicted MAP scores would be 611 and 642 respectively. This is a difference of 31 points. However, if those same two students had scored 130 on the NNAT, their respective predicted MAP scores would be 659 and 670—a difference of only 11 points. The relationship of the scores in Cohort 2 suggests that students in the Black group are at a significant relative disadvantage if they attain a low score on the kindergarten nonverbal ability test. The disadvantage also exists for a student who scored the Cohort's mean score (NAI = 107), though the eventual MAP disadvantage is not as great (21 points lower). But even though the gap reduces as NAI scores increase, indicators for the gap in eventual reading and writing achievement—commonly referred to as the "achievement gap"—seem to manifest in the kindergarten year or earlier. Even more discouraging is that the mean NAI score for Black students in Cohort 2 is 17 points lower than the total cohort mean. This lower mean score combined with similar standard deviations to the other racial subgroups indicate that there are likely to be few Black students on the higher end of the NAI scale taking advantage of the upper end of the steeply sloped trend line where the gap closes.

The steep slope of the Black subgroup line in Cohort 2 logically coincides with the observations made for the White subgroup in Cohort 2. Figure 4.10 depicted the non-White trend line eventually intersecting and surpassing the White regression line. Given that the Black subgroup is one of the larger remaining race categories (*n* = 70), it seems appropriate to infer that this group is having an influence on the slopes of the White and non-White lines.

However, this same inference does not transfer well to the relationship analysis for Cohort 1. The Black group in Cohort 1 did not make relative gains on the non-Black group as NAI increased. Yet the regression line for the White group in Cohort 1 was more quickly overtaken (intersection at NAI = 128) by the non-White group than it was in Cohort 2 (intersection at NAI = 143). This discrepancy may indicate that it is the Asian (*n* = 60) group combined with the Black (*n* = 66) group which is more strongly influencing the non-White regression lines discussed previously.

Asian versus non-Asian. The regression analysis of Model 1 for the Asian subgroup in Cohort 1 revealed statistically significant variables and the trend lines (Figure 4.13) are similar to those for the White subgroup. Both the Asian and White subgroups have slopes that intersect with their respective inverse groups near NAI = 130. Also similar to the White group is the fact that non-Asian students with higher kindergarten NAI scores (above 130) will have predicted MAP scores lower than their non-Asian peers with similar kindergarten nonverbal ability scores. Since the three largest populations in Cohort 1 were White (*n* = 627), Black (*n* = 66), and Asian (*n* = 60), it may be inferred that it is the combination of the Black group (mean NAI = 88) and the Asian group (mean NAI = 114) which is having the most significant impact on the non-White slope differential. Since the groups combine to establish the slope of this line, the steep slope may be a result of a teeter-totter effect with high Asian scores tipping the right side up and the lower Black scores tilting the line down on the left. Thus, even though White and Asian groups have similar trend lines, they appear to be surpassed by the non-White and non-Asian trend lines likely as a result of the influence of the combined Black score relationships.

One must be cautious when reviewing these regression trends as inaccurate conclusions could easily be drawn when simply viewing them independently at face value. In doing so, it would be logical for one to assume that given the non-White regression line, a Black or Asian student scoring at the upper end of the NAI scale would outperform a White student scoring at the same level. However, when considering the combined stories told by the interaction of all three graphs (i.e. all six regression lines), a more logical conclusion would be that a Black student with a high NAI score may not outperform the White student with the same nonverbal ability given the slope of the Black regression line. This of course is because of the discrepancy between the Asian line and the Black line on their respective graphs, combined with the variance in the two groups' means and their similar standard deviations. Again, the slope is likely the result of a weighted teeter-totter effect.

Unfortunately, this reasoning cannot be accurately tested with the data in Cohort 2 due to two factors: the data from the Black group in Cohort 2 performed much differently than in Cohort 1, and the Asian group in Cohort 2 did not produce significant variables in either Model 1 or Model 2. The Asian group in Cohort 2 therefore must be disregarded.

This is unfortunate because it leaves behind a bit of a mystery regarding the interplay between the regression lines of the subgroups in Cohort 2. Unlike Cohort 1, the Black group in Cohort 2 has a steeper regression line than the non-Black group (Figure 4.12). This is not surprising given that this cohort's Black NAI mean score is also higher (90) and the Asian and White NAI mean scores are lower (112, 109). This would reduce the teeter-totter effect observed in Cohort 1 and suggest that the Black subgroup is now having a positive influence on the non-White regression line rather than the negative influence inferred from the data in Cohort 1. Given the nation's profound achievement gap in the Black subgroup, one can only hope that the data in Cohort 2 is more typical in other populations than that produced by the Black group in Cohort 1. With data such as Cohort 2's, the achievement gap still exists, but the trend at least rationally allows for students with higher kindergarten nonverbal abilities to begin closing the achievement gap by third grade.

Hispanic versus non-Hispanic. One of the more unusual regression trends was generated by the Hispanic population in Cohort 1 (the variables were statistically significant, unlike Cohort 2). As seen in Figure 4.14, the regression line for Hispanic students is relatively flat and low, indicating that Hispanic students are projected to score below their non-Hispanic peers regardless of their NAI score in kindergarten. Furthermore, Hispanic students who score higher in nonverbal ability seem to have little to no advantage over other Hispanic students who score lower on the NNAT. One conclusion at which one could easily arrive would be that a lack of success on the MAP may be attributed to a lack of experience with the English language. But remember, the only students who were represented in this study were those who were in the same English-speaking school system from kindergarten through third grade. The students upon taking the MAP test near the end of the third-grade year had been exposed to English in school for at least four years prior to taking the communication arts assessment. Students who were new to the school system in second or third grade and may have had limited exposure to English were not included in this study's data samples. Further analysis of research regarding the acquisition of language and its effect on student achievement would need to be consulted before such conclusions could be drawn.

Another conclusion to caution against is concerning the possible lack of English experience at the time of the administration of the kindergarten NNAT assessment. Again, the NNAT is completely nonverbal. There are no speaking, listening, or reading components on the assessment—all items are progressive figural matrices. The test was designed not to bias against those who speak a non-English language.

Perhaps the strongest caution for interpreting these scores comes from the *n* value. The Hispanic subgroup in Cohort 1 only had 15 students. Though the data presented itself to be statistically significant at the 95% confidence level, the low *n* count leaves doubt as to the transferability of the regression. If conclusions are to be drawn for this subgroup, further research with a larger *n* is warranted.

Individualized education plan versus no individualized education plan. Neither cohort produced a regression model in which the IEP interaction variable (IEPNAI) was statistically significant at the 95% confidence level which limits the conclusions that can be drawn for this subgroup. However, using regression Model 2, both cohorts did produce statistically significant IEP variables. When graphed parallel to the non-IEP subgroups (Figures 4.15 and 4.16), the data from both cohorts produced very similar results. In both cohorts, the IEP group underperformed the non-IEP group on MAP achievement. The differences in predicted MAP scores between the groups in Cohort 1 and Cohort 2 respectively were 30 points and 27 points.

When considering IEP data it is important to remember that some students with IEPs receive accommodations which change their testing environments in third grade. While this may also be the case for NNAT testing in kindergarten, many students at the younger age may not have yet been diagnosed with an educational disability. Therefore, the testing accommodations may have been different for the students between the test administrations, and certainly may have been different from the testing environments of their nondisabled peers. For many reasons, the IEP subgroup must be analyzed with caution.

Free or reduced-price lunch versus no free or reduced-price lunch. Similar to the gender subgroup, the FRL subgroup had one Cohort that produced a statistically significant interaction variable (FRLNAI, Cohort 2) and one that did not. Model 2 however was executed in Cohort 1 and revealed a significant subgroup variable allowing at least a parallel regression graph to be presented. The graphs for these groups can be found in Figures 4.17 and 4.18.

In both cohorts, the FRL group underperformed the non-FRL group. The difference in cohort 1 was 21 MAP points. In Cohort 2, the FRL regression trends have a similar pattern to the Black subgroup in the same cohort. FRL students with lower NAI scores are predicted to score much lower than their non-FRL peers (28 MAP points lower at NAI = 80). However, at the higher NAI end of the graph, FRL students have nearly caught their non-FRL counterparts closing the gap as NAI increases. For example, at NAI = 130, the MAP gap is only eight points. But, for the Black and FRL subgroups, students who have a lower kindergarten NAI seem to have a difficult time keeping up with the achievement levels of other non-Black or non-FRL peers who also have a lower NAI score. However, in both subgroups, it appears that students who demonstrate a high NAI score in kindergarten can negate the predicted performance deficiencies by which the rest of the group is affected.

Subgroup *r* Square Values

Each subgroup produced a statistically significant regression model with F values ranging from 63.5 to 137.1. The *r* square values reveal important information with regards to interpreting the results of the study. Educators and policymakers alike must realize that the *r* square value suggests the percentage of the MAP scores that can be inferred by the kindergarten NAI scores. For example, the *r* square value for the gender variable in Cohort 1 was .26. The *r* square values for the other variables and cohorts were presented in Tables 4.4 to 4.11.

The .26 *r* square value for the gender variable in Cohort 1 suggests that 26% of a child's predicted MAP score (using the results of the statistical tests for the gender variable) can be attributed to the child's kindergarten NAI score. This of course means that 74% of the MAP score is influenced by other factors than what is measured as nonverbal ability. For educators and students, this can be both encouraging and discouraging. If 75% of the capacity to achieve can be influenced by instruction, programming, and other factors, teachers and school systems clearly have an opportunity to impact the success of their students, regardless of their ability when they first walk through the school doors. Schools that truly believe that all children can learn can find hope in the *r* square values of this and other similar studies.

The downside, though, is that nonverbal ability does clearly have a predictive relationship to eventual achievement in reading and writing. A child who enters school with a low nonverbal ability level will be facing a significant obstacle. And even with the best efforts made by parents, educators, and school systems, the low nonverbal ability is still likely to negatively contribute to the eventual achievement level. With proficiency targets being raised and higher expectations placed on student achievement, these students may have an even more difficult time overcoming the impact of a lower nonverbal ability.

Suggestions for Future Study

The results of this study along with the evidence from many other studies presented in Chapter 2 raise questions for researchers and policymakers to consider. Since kindergarten nonverbal ability impacts a child's future achievement levels in the areas of reading and writing, we must consider whether there are ways to increase nonverbal ability. Do students who have more experience playing with puzzles perform better on figural matrix assessments? What are the factors that cause higher or lower nonverbal abilities, and how can we enhance these abilities in children? Is the nonverbal ability of a child innate or can it be influenced in the early stages of child development? Future studies may wish to also track students who were exposed to certain interventions such as preschool experience, reading intervention in the early school years, and even sociological circumstances such as divorce, one- or two-income families, and other potentially relevant factors.

Perhaps the most encouraging finding in this study is that some subgroups that demonstrated lower third-grade achievement levels, had steeper regression lines which eventually approached or intersected with the lines generated from higher-performing subgroups. That is, students from lower-achieving subgroups who had higher nonverbal abilities in kindergarten were able to eventually demonstrate achievement levels just as high as or higher than similar students from higher-achieving subgroups. In many cases, a child with a high NAI is likely to achieve highly in reading and writing even when being a member of an otherwise underperforming subgroup. So again, we must continue to seek whether early interventions exist that can boost a child's nonverbal ability.

Future studies also may wish to replicate this procedure with assessments containing a verbal component. Though there are advantages to nonverbal assessments being used, particularly at an early age (discussed in chapter 2), obviously a child's verbal ability may also influence—possibly even more so—eventual success in reading and writing.

Also of interest may be to study the impact of nonverbal ability on the eventual achievement in mathematics. This study only considered achievement levels in communication arts, but this study could be easily replicated to discover the relationship between nonverbal ability and mathematics. It seems plausible that the visual perception skills needed for success on the nonverbal figural matrices may positively relate to the areas of mathematics which also require visual perception skills such as geometry. It would be of further interest to explore whether a nonverbal ability assessment gleans insight into eventual success with number sense. A study replicating the procedures of this research with a different dependent variable (e.g. Mathematics MAP subscores) may provide interesting insight to this question.

But even using the same variables, it may be of interest to replicate this study with a different population of children. The descriptive statistics presented at the beginning of Chapter 4 indicated that the sample population had an overall mean NAI of 108 and 107 for Cohorts 1 and 2 respectively. Though within the average range, the nonverbal ability of this population is skewed above average. Replicating the study with a more average or below-average population may be of interest for comparison.

One main purpose of this study was to determine whether or not the relationship reported to exist between nonverbal ability and achievement would remain consistent when the testing sessions were extended between kindergarten and third grade. Since the results were affirmative, future studies may wish to consider extending the time gap even further to determine whether this relationship maintains itself over even larger periods of time. A prudent hypothesis may in fact be that the relationship will continue to exist since this study was successful in demonstrating that similar relationships found between nonverbal ability and achievement were present when tested over a four-year span as compared to those results that were discovered when tested concurrently.

References

Berninger, V. W., Thalberg, S. P., DeBruyn, I., & Smith, R. (1987). Preventing reading disabilities by assessing and remediating phonemic skills. *School Psychology Review*, *16*, 554–565.

Blachman, B. A. (1983). Are we assessing the linguistic factors critical in early reading? *Annals of Dyslexia*, 33, 91–109.

Blachman, B. A. (1994). What we have learned from longitudinal studies of phonological processing and reading, and some unanswered questions: A response to Torgesen, Wagner, & Rashotte. *Journal of Learning Disabilities*, *27*, 287–291.

Bracken, B. A., & McCallum, R. S. (1998). *Universal nonverbal intelligence test.* Itasca, IL: Riverside.

Brown, A. L., Palincsar, A. S., & Purcell, L. (1986). Poor readers: Teach, don't label. In U. Neisser (Ed.), *The school achievement of minority children: New perspectives* (pp. 105–143). Hillsdale, NJ: Erlbaum.

Carver, R. P. (1990). Intelligence and reading ability in grades 2–12. *Intelligence*, *1*, 449–455.

Clay, M. (1993). *Reading Recovery: A guidebook for teachers in training*. Hong Kong, China: Heinemann.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York, NY: Academic.

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, *22*, 8–15.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*(1), 3–17.

Gredler, G. R. (1997). Issues in early childhood screening and assessment. *Psychology in the Schools*, *34*, 99–106.

Hammill, D., & McNutt, G. (1981). *The correlates of reading*. Austin, TX: Pro-Ed.

Hayes, S. C. (1999). Comparison of the Kaufman Brief Intelligence Test and the Matrix Analogies Test–Short Form in an adolescent forensic population. *Psychological Assessment*, *11*, 108–110.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2002). *Applied statistics for the behavioral sciences* (5th ed.). Independence, KY: Wadsworth.

Jansky, J. J. (1978). A critical review of some developmental and predictive precursors of reading disabilities. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 331–347) New York, NY: Oxford University Press.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437–447.

Kennedy, M. M., Birman, B. F., & Demaline, R. E. (1986). *The effectiveness of Chapter 1 services*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Mann, V. A. (1993). Phoneme awareness and future reading ability. *Journal of Learning Disabilities*, *26*, 259–69.

Missouri Assessment Program Technical Report. (2008). Monterey, CA: McGraw-Hill.

Missouri Census Data Center. (2009). Trends in Selected Demographic Items for Missouri, 1990–2007. Jefferson City, MO: Missouri Census Data Center.

NAEP 2000 Reading. (2001). A report card for the nation and the states. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Naglieri, J. A. (1985). *Matrix Analogies Test–Expanded Form*. San Antonio, TX: The Psychological Corporation.

Naglieri, J. A. (1997). *NNAT multilevel technical manual*. San Antonio, TX: Harcourt Brace.

Naglieri, J. A., & Ronning, M. E. (2000a). Comparison of White, African American, Hispanic, and Asian Children on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, *12*(3), 328–334.

Naglieri, J. A., & Ronning, M. E. (2000b). The relationship between general ability using the Naglieri Nonverbal Ability Test (NNAT) and the Stanford Achievement Test (SAT) reading achievement. *Journal of Psychoeducational Assessment*, *18*, 230–239.

Obrzut, J. E., Jones, M. J., Bolocofsky, D. N., & Heath, C. P. (1981). An investigation of the DIAL as a pre-kindergarten screening instrument. *Educational and Psychological Measurement*, *41*, 1231–1241.

Oka, E., & Paris, S. (1986). Patterns of motivation and reading skills in underachieving children. *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities*. Hillsdale, NJ: Erlbaum.

Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Author.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific Reading Disability: A View of the Spectrum* (pp. 77–121). Timonium, MD: York Press.

Scott, M. S., Delgado, C. F., Tu, S., & Fletcher, K. L. (2005). Selecting and validating tasks from a kindergarten screening battery that best predict third grade educational placement. *Education and Training in Developmental Disabilities, 40*(4), 377–89.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education*, *1,* 1–57.

Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly*, *19*, 278–303.

Taylor, B. M., Short, R. A., Frye, B. J., & Shearer, B. A. (1992). Classroom teachers prevent reading failure among low-achieving first-grade students. *The Reading Teacher*, *45,* 592–597.

Teisl, J. T., Mazzocco, M. M. M., & Myers G. F. (2001). The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Journal of Learning Disabilities*, *34*(3), 286–93.

Torgesen, J. K. (2002). The Prevention of reading difficulties. *Journal of School Psychology*, *40*(1), 7–26.

Torgesen, J. K., & S. R. Burgess (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning reading* (pp. 161–188). Hillsdale, NJ: Erlbaum.

VanDerHeyden, A. M., Witt, J. C., Naquin, G., & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *The School Psychology Review, 30*(3), 363–382.

Appendix A

Permission to Perform Research

[redacted] SCHOOL DISTRICT [redacted]

Program Evaluation

June 24, 2011

[redacted]

Dear Aaron:

This letter will serve as official notification that your proposed research project, "The relationship between kindergarten nonverbal ability and third-grade reading achievement" has been approved for implementation in the [redacted] School District. A copy of your signed Application to Perform Research is attached.

Thank you for your interest in our district, and the best of luck with your research endeavors.

Sincerely,

[redacted]
Coordinator, Program Evaluation

Attachment

[redacted]

[redacted]

JUN 24 2011

Director, Program Evaluation

[redacted]

Application to Perform Research

Division of Program Evaluation

[redacted]

I. **Name of Primary Investigator** Aaron Wills

Position [redacted] Affiliation [redacted]

Office Address [redacted]

Home Address [redacted]

Office Phone [redacted] Home Phone [redacted]

E-mail Address [redacted]

Names of additional members of research team:

Name ______ Phone ______

Name ______ Phone ______

Name ______ Phone ______

II. **Project Title** The relationship between kindergarten nonverbal ability and third-grade reading achievement

Description

Using correlation and regression analysis, this study will determine the statistical relationships between the kindergarten NNAT and third-grade reading MAP scores. Data from two cohorts of [redacted] students will be used in the study: students who took the third-grade MAP test in the year 2009 and also those who took the test in 2010.

Note: Please attach copies of any measures to be used (e.g. tests, questionnaires, surveys, etc.)

III. Participant Involvement

	Number of Subjects	Time Requirements
Pupils:	approximately 1500	None
Teachers:	0	
Administrators:	0	
Parents:	0	

Describe the involvement required of subjects (or access to records if subjects are not required).

Access to existing test score records is being requested.

Number of persons visiting sites in connection with project: 0

Frequency of visits during a school year: ______

Total contact hours of the project: ______

IV. Project Requirements

Number and type of school:

Elementary (K-5) [redacted] Middle school (6-8) ______ High school (9-12) ______

Early Childhood Education (birth to kindergarten) ______

Adult Basic Education ______ Other ______ Grades required ______

Total number of schools [redacted] Total number of classrooms ______

Other school characteristics:

Do you require any specific schools?________ If yes, please provide names:

Starting date of research: June 16, 2011 Ending date of research: December 1, 2011

Frequency of contact with subject: ____________________

V. **Results**

What is the anticipated value of the research?

In general:

To determine whether the relationships calculated by Nagliari and Ronning (2000) between nonverbal ability and achievement are similar when measured over a four-year span and with a different achievement test instrument (MAP).

To the [redacted] District:

This study will statistically describe the relationship between two assessments given annually through the [redacted] District's assessment program.

VI. **Dissemination**

How will the results of your study be used? Will they be available to the public in any form? If so, what groups will have access to the results? Will the [redacted] District, or any individuals within [redacted] be identified in your reports?

The completed study will be available to the public through the University of Missouri St. Louis Library or upon request from the researcher. Neither student identities nor the [redacted] District will be named in the study, but a broad description of the sample population will be provided.

VII. References (You may omit names, if you have promised confidentiality.)

Are other school systems involved in this research? No

Please name ____________________

Have you conducted research in other school systems? No

Please name ____________________

Date(s) ____________________

VIII. Human Subjects' Protection

Has this research been approved by a university or other institutional review for protection of human subjects?

Yes _____ No X

If yes, please indicate who reviewed the proposal and when:

____

If no, please explain why this proposal has not been reviewed for protection of human subjects:

This proposal is being reviewed concurrently with this request and the research will only proceed if approved by both IRB processes at [redacted] and the University of Missouri St. Louis.

Note: All researchers who plan to collect information from or about individual students should attach copies of their proposed consent forms and a brief description of planned procedures for obtaining informed consent. Research involving individual students may require the informed consent and signed agreement of parents or legal guardians.

[redacted]

IX. **Upon completion of the research you will be required to submit three copies of the report (or summary) to the Director of Program Evaluation.**

By signing this application, the applicant certifies that the research herein described involves an investigation which:

1. Promises to produce information of value to [redacted] or the field of education;
2. Provides adequate safeguards for participants' rights;
3. Does not detract from the primary mission of instruction; and
4. Is not-for-profit in nature.

The documents can be expected by (date) June 15, 2012

1. [redacted] 6/16/2011
signature of applicant — date

2. [redacted] Univ. Mo. St. Louis
PRINT – name of institutional advisor, professor or supervisor — institution

3. [redacted] [redacted]
signature of advisor, professor or supervisor — office telephone

---

(For district use only)

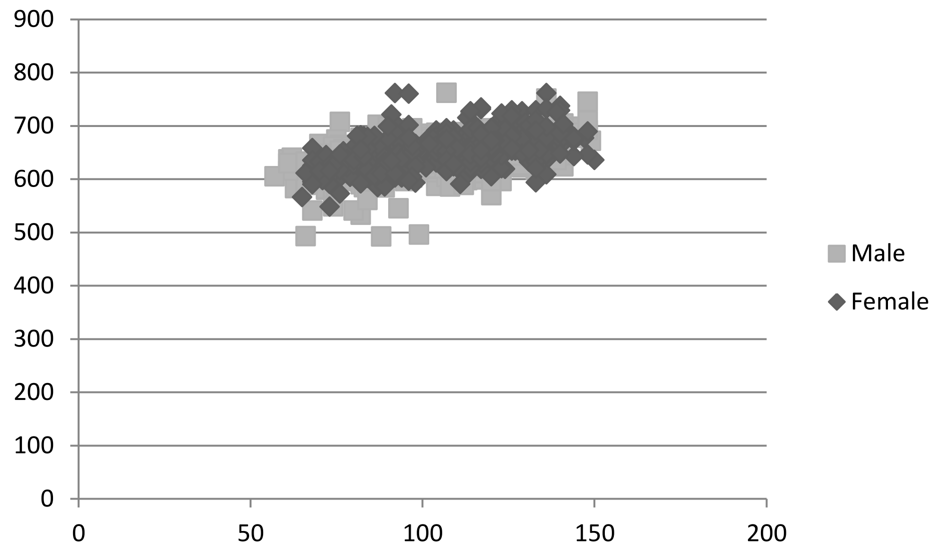
1. 6/24/2011
signature of Director of Program Evaluation — date

2.
signature(s) of administrator(s) affected — date

[redacted]

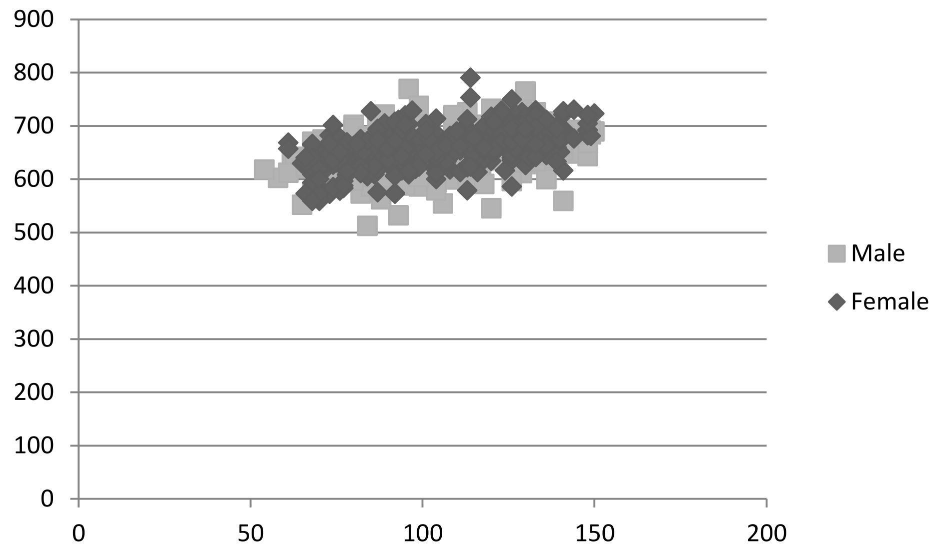
Appendix B

Scatter Plot for Gender Variable in Cohort 1



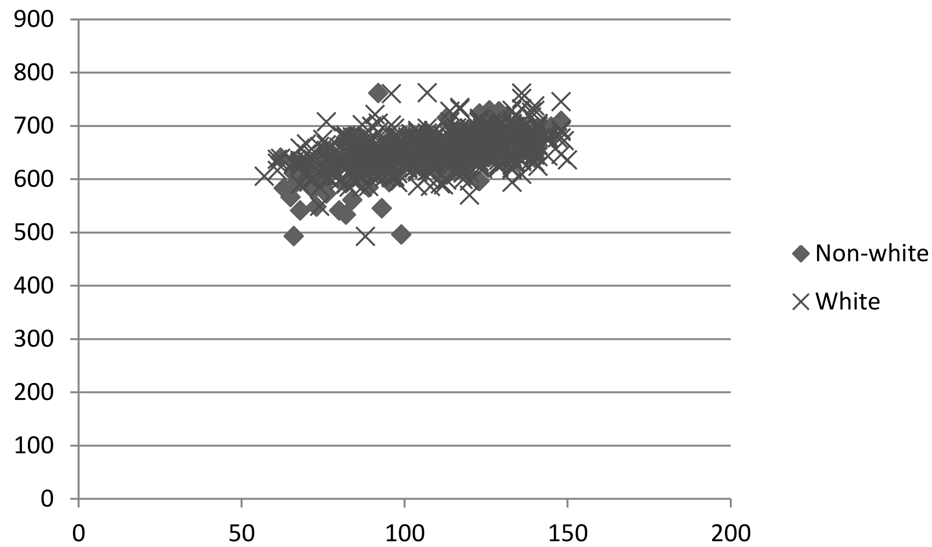
Appendix C

Scatter Plot for Gender Variable in Cohort 2



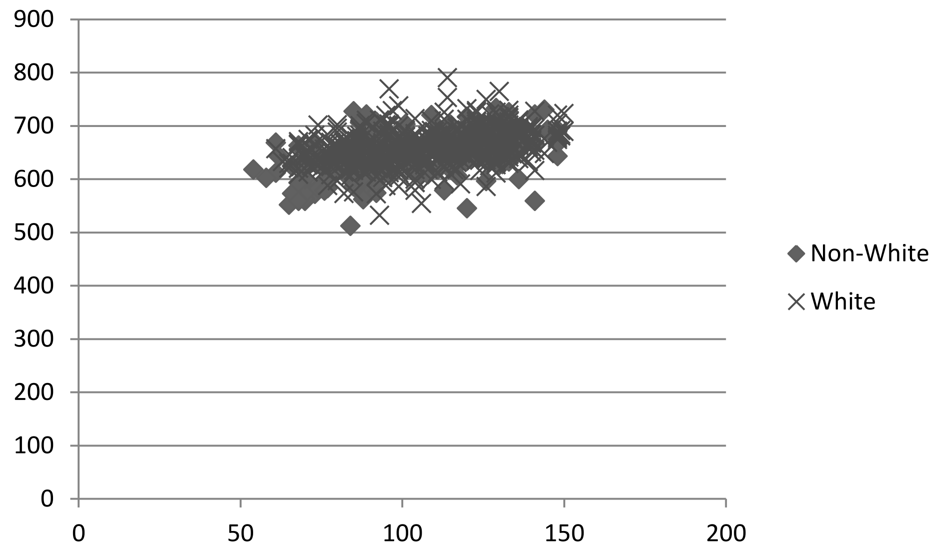
Appendix D

Scatter Plot for White Variable in Cohort 1



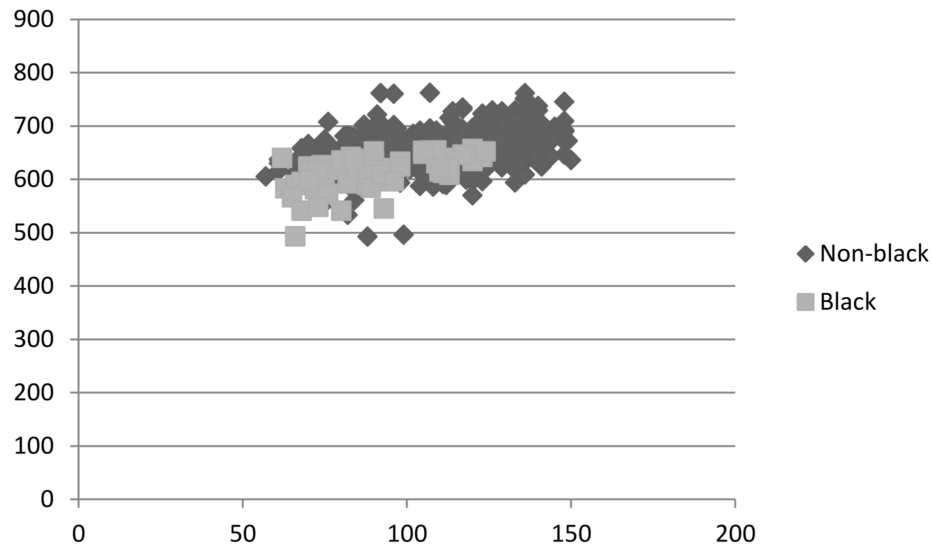
Appendix E

Scatter Plot for White Variable in Cohort 2



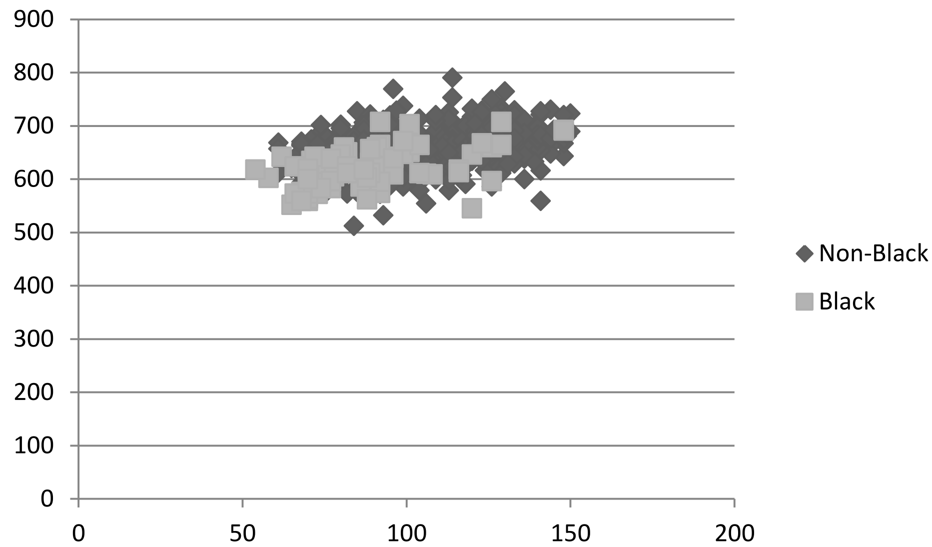
Appendix F

Scatter Plot for Black Variable in Cohort 1



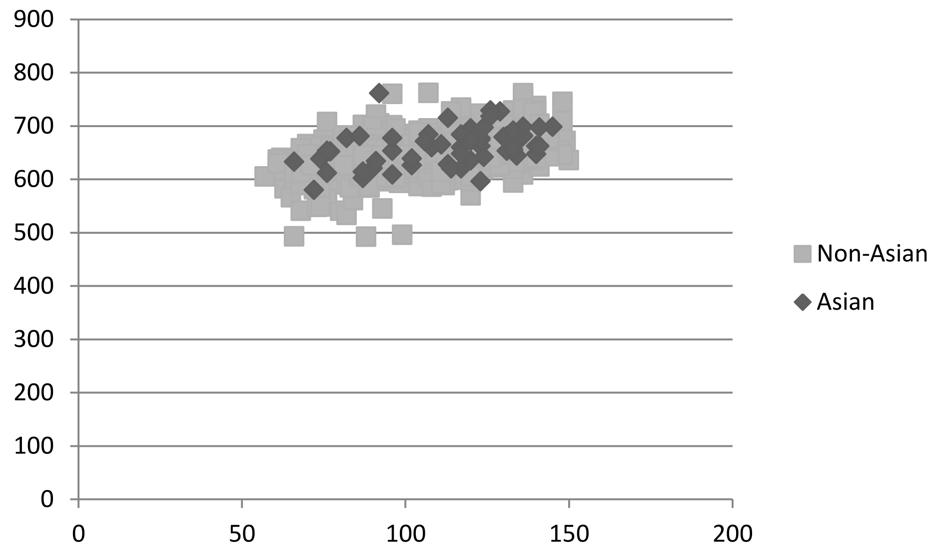
Appendix G

Scatter Plot for Black Variable in Cohort 2



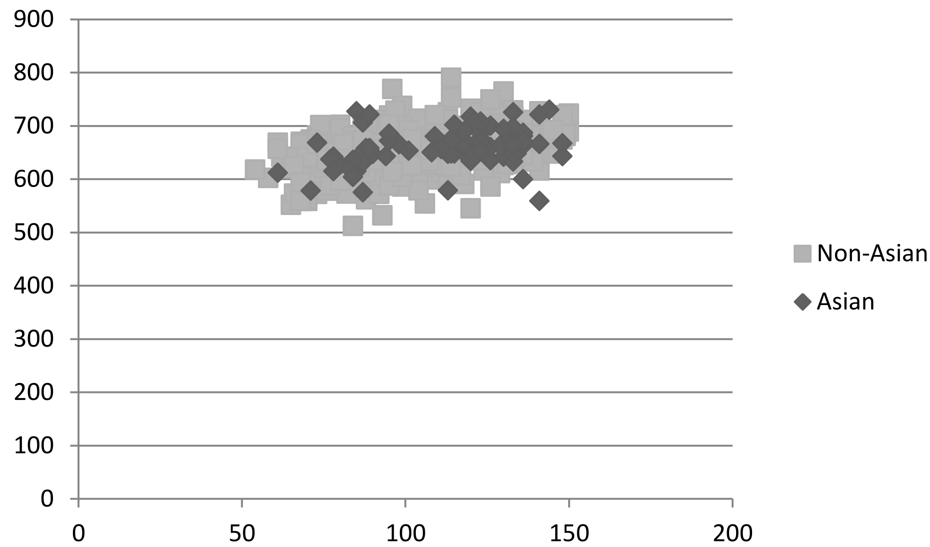
Appendix H

Scatter Plot for Asian Variable in Cohort 1



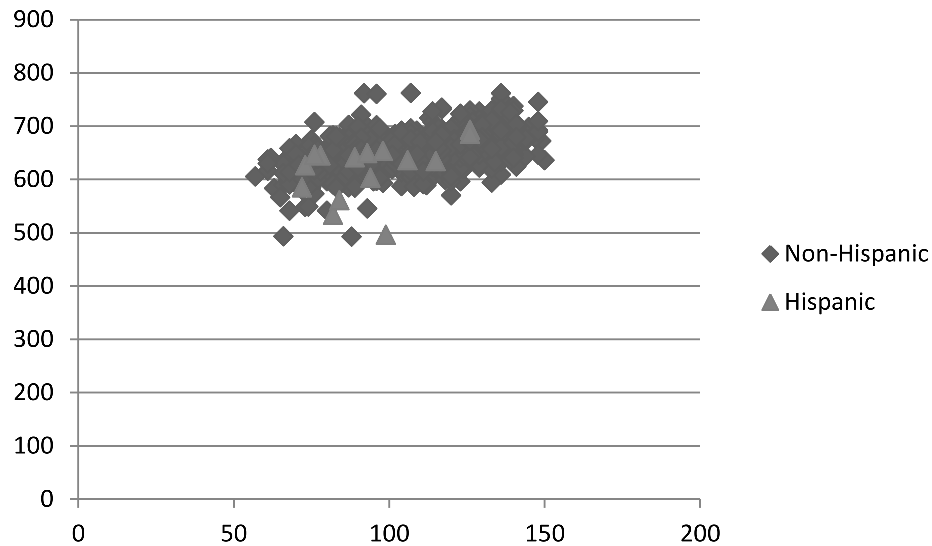
Appendix I

Scatter Plot for Asian Variable in Cohort 2



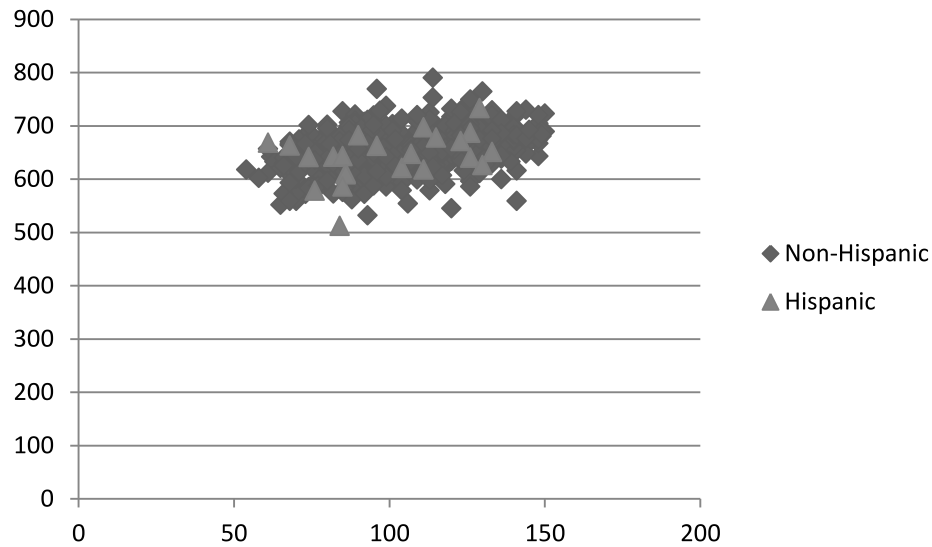
Appendix J

Scatter Plot for Hispanic Variable in Cohort 1



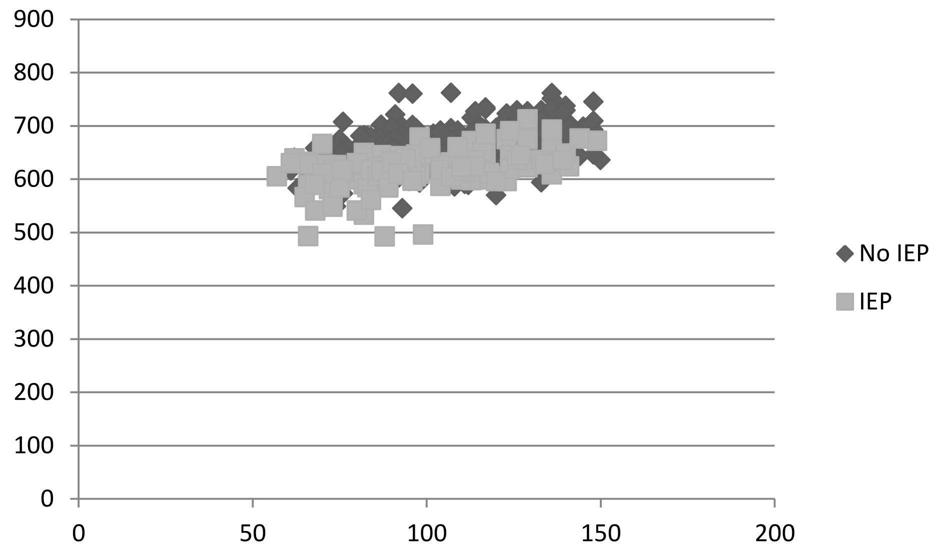
Appendix K

Scatter Plot for Hispanic Variable in Cohort 2



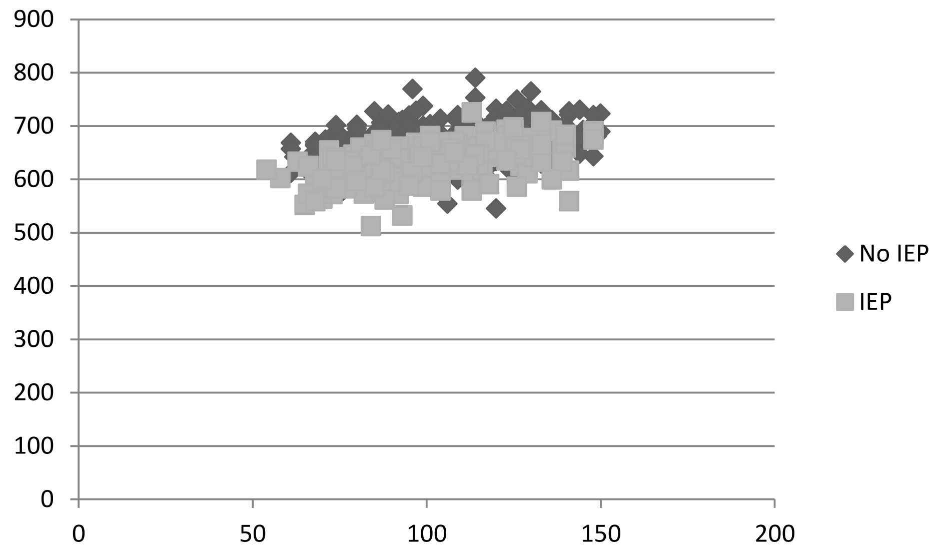
Appendix L

Scatter Plot for Individualized Education Plan Variable in Cohort 1



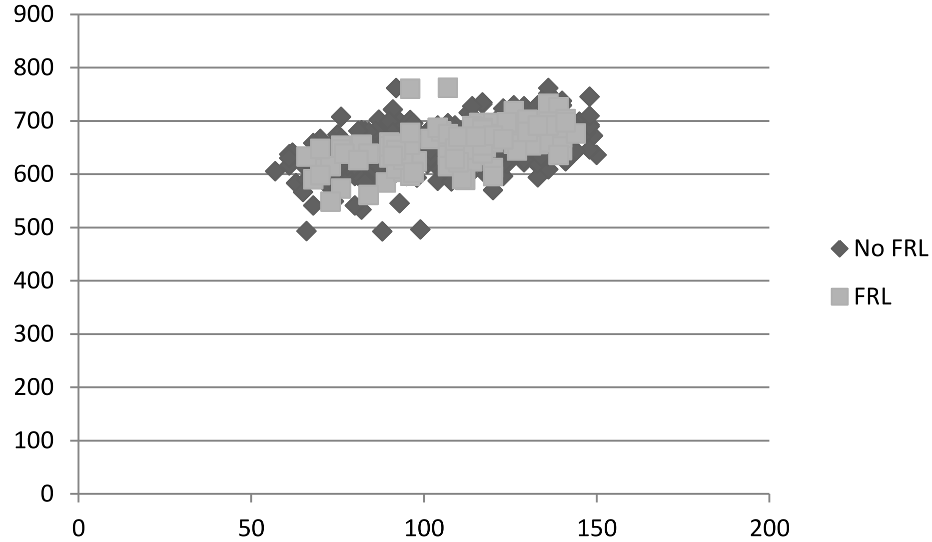
Appendix M

Scatter Plot for Individualized Education Plan Variable in Cohort 2



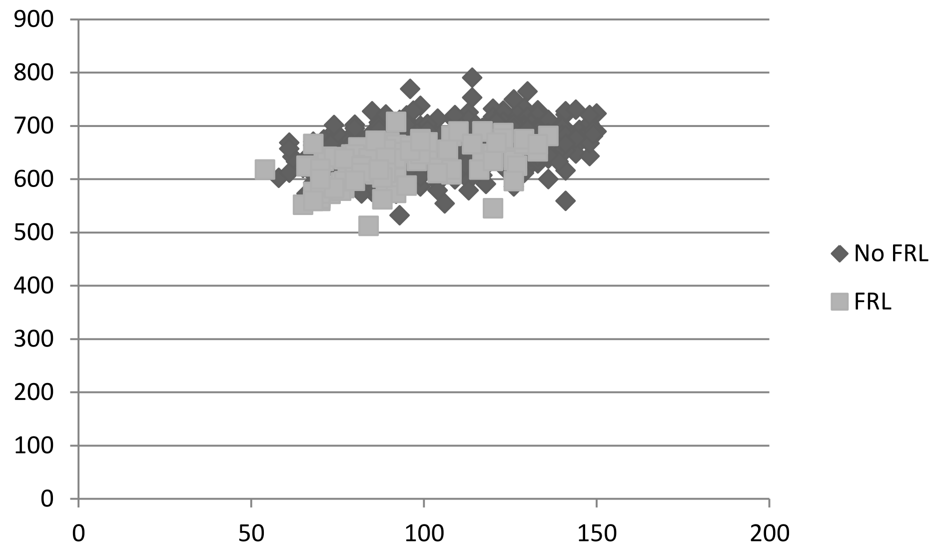
Appendix N

Scatter Plot for Free or Reduced-Price Lunch Variable in Cohort 1



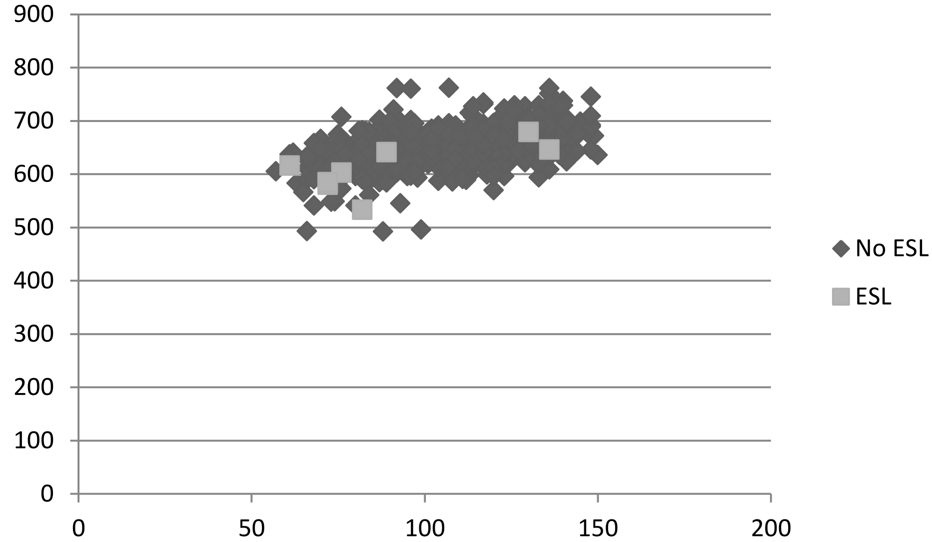
Appendix O

Scatter Plot for Free or Reduced-Price Lunch Variable in Cohort 2



Appendix P

Scatter Plot for English as a Second Language Variable in Cohort 1



Appendix Q

Scatter Plot for English as a Second Language Variable in Cohort 2