Abstract

TAYROSE, MICHELLE PARKER. The Relationship between WISC-IV Scores and North Carolina State Achievement Test Scores. (Under the direction of Jeffery P. Braden.)

The historical correlation between intellectual functioning and academic achievement is largely based on individually administered achievement tests. However, the standards-based reform movement and recent special education legislation emphasize group-based achievement tests and allow for the use of state-based tests in eligibility determinations. Importantly, there does not exist research examining the IQ-achievement relationship using standards-based tests; thus, the current study evaluated the correlation between the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) and those tests comprising the North Carolina assessment program (i.e., End-of-Grade (EOG) tests). Five NC school psychologists provided archival information on students (n = 76) and results yielded support for the two hypotheses, which postulated that confidence intervals placed around correlation coefficients between Full Scale IQ (FSIQ) scores and EOG-Reading and –Mathematics scores would capture r =.6, which is roughly halfway between the range of historical correlations. Implications regarding the validity of the WISC-IV and its potential use in identifying learning disabilities in which achievement is measured by tests linked to state standards, as well as suggestions for future research are discussed.

The Relationship between WISC-IV Scores and North Carolina State Achievement Test Scores

by
Michelle Parker Tayrose

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Psychology

Raleigh, North Carolina

2008

APPROVED BY:

Patricia W. Collins, Ph.D.

Gail Jones, Ph.D.

Jeffery P. Braden, Ph.D.
Chair of Advisory Committee

Biography

Born and raised in Chapel Hill, North Carolina, Michelle Parker Tayrose is the daughter of loving parents, Edythe and Barnett Parker, and older sister to wonderful siblings, Allison and Jeremy. Michelle graduated from East Chapel Hill High School in 1998 and then attended the University of North Carolina at Chapel Hill where she majored in Psychology and minored in Advertising. Upon receiving her Bachelor of Arts degree in 2002, Michelle worked for two years at the Duke University ADHD Program on the Multimodal Treatment Study of Children With and Without Attention-Deficit/Hyperactivity Disorder (MTA) before entering the North Carolina State University School Psychology program in 2004. She is now pursuing doctoral studies in School Psychology at NCSU. Although she has definitely traveled "Tobacco Road," Michelle remains none other than a Tarheel. She currently resides in Chapel Hill with her amazing husband, and best friend, Greg, and their dog, Bumble.

Acknowledgements

First and foremost, I would like to thank my advisor, Jeff Braden, who pushed me to think critically, and helped me to “find the horse.” He has invested an extraordinary amount of time and effort, and has provided much guidance in my professional development. I would also like to thank Ann Schulte for sharing writing materials, statistical consults, and formatting assistance, all of which were integral to this final product. I would like to extend my gratitude to Patsy Collins and Gail Jones for taking the time to read my thesis, provide insightful feedback, and participate in my defense.

This thesis would not have been possible without the help of those school psychologists who compiled and entered the archival student information. I know they are extremely busy and thus deeply appreciate the time and effort necessary for their participation in this project.

It is virtually impossible to show due appreciation to my parents. They have been there for me and have encouraged me from “Day One,” showing me constant love and support and allowing me to accomplish that which I perhaps did not believe possible. Just knowing I will make them proud is a deeply motivating and satisfying force. I also wish to thank my brother and sister, whose unselfish happiness in my achievements means so much, and whose love and support helped me through this endeavor.

Thanks go, as well, to my extended family; namely my aunts and cousins. They are all such admirable people who have provided me with much encouragement and confidence throughout my life. And thanks, too, to my Aunt Deb for the statistical assistance and late-night chats about the horrors of graduate school. I take this opportunity to offer deepest

thanks to my grandparents who had such a lasting, positive impact on my life. I know they are here in spirit and would be so proud of me, and for these reasons I will be eternally grateful.

And last but certainly not least, I would like to thank my husband, Greg, whose unwavering love and support made this work possible. Whether it was a much-needed hug, a reminder to “hang loose” and that I am “the woman for the job,” or a just fun, relaxing time together, he has been there for me in more ways than I can express. I can’t imagine a better teammate.

Table of Contents

List of Tables

Chapter 1

Review of the Literature and Statement of the Problem

The Standards and Validity

The *Standards* is a joint publication intended to "provide criteria for the evaluation of tests, testing practices, and the effects of test use" (AERA, APA, & NCME, 1999, p. 2). Here, validity is deemed to be a unitary concept and is defined as "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p. 9). The various sources of evidence that can contribute to determining the validity of a test include evidence based on: (a) test content, or the degree to which test items are linked to the construct the test is designed to measure, (b) response processes, or the fit between the construct and the nature of examinees' responses, (c) internal structure, or the extent to which relationships among items are consistent with the construct being measured, (d) relations with other variables, or the examination of the test's relationship to other instruments intended to measure similar constructs, and (e) consequences of testing, including both intended and unintended consequences of test use (AERA, APA, & NCME, 1999).

Intelligence Testing and the Schools

The advent of modern intelligence testing began in the early 1900s when Binet and Simon developed the first practical intelligence scale. The lack of consensus that existed at that time with regard to the definition of intelligence has continued to the present day; however, early theorists considered the notion of adaptability to one's environment central to the construct and this element has been retained in more recent definitions (Sternberg, 1997;

Wasserman & Tulsky, 2005). The Binet-Simon Scale was designed to identify those children who were unlikely to benefit from (i.e., adapt to) a strongly academic education. Thus, intelligence tests were originally intended to predict educational achievement, and this remains a primary method of validating intelligence tests. IQs are the best estimates of future school performance, next to current achievement (Wasserman & Tulsky, 2005).

Braden (2003) identifies six distinct purposes that drive psychological assessment in the schools (i.e., screening, diagnosis, intervention, evaluation, selection, and certification) but notes that the primary function of these assessments, and more specifically, intelligence tests, is the diagnosis and classification of educational disabilities. Given that the majority of school-based psychological assessments are initiated due to learning problems, these assessments are often conducted to assist in understanding and improving these types of problems and to facilitate the determination of special education eligibility (Braden, 2003; Reschly & Grimes, 2002). Generally, school-based cognitive ability assessments involve individually administered tests of intellectual functioning, in which an examiner gives a standardized test to an individual student (Braden, 2003). The Wechsler intelligence scales are the most frequently used and widely researched of these contemporary intelligence tests (Zhu & Weiss, 2005).

WISC-IV revisions. The most recent edition of the Wechsler tests for school-age children, the WISC-IV, was updated in 2003. The revision goals as outlined in the WISC-IV manual are five-fold: (1) to update the instrument's theoretical foundations, (2) to enhance clinical utility, (3) to increase developmental appropriateness, (4) to improve psychometric properties, and (5) to increase user-friendliness (Wechsler, 2003). Arguably the most

significant of these revisions (Burns & O'Leary, 2004) is the replacement of the Verbal and Performance IQ scores with four composite scores: (a) Verbal Comprehension Index (VCI), (b) Perceptual Reasoning Index (PRI), (c) Processing Speed Index (PSI), and (d) Working Memory Index (WMI). All four of the Indexes contribute to the Full Scale IQ (FSIQ) score, which is considered to be a clinically meaningful estimate of a child's general cognitive functioning (Saklofske, Prifitera, Weiss, Rolfhus, & Zhu, 2005; Weiss, 2003). Based on theory, and supported by clinical and factor analytic research, this new four-factor structure is intended to better represent the core domains of cognitive functioning and to better align with the current research (Burns & O'Leary, 2004; Wechsler, 2003; Williams, Weiss, & Rolfhus, 2003). More specifically, the FSIQ now includes greater contributions from fluid reasoning, working memory, and information processing speed, as contemporary research in the area of intelligence testing emphasizes these as essential components of cognitive ability (Wechsler, 2003; Weiss, 2003).

The Historical IQ-Achievement Correlation

There exists an historical linkage between intelligence and achievement. Cognitive test scores and academic achievement test scores tend to be positively correlated, with correlations usually ranging between $r = .4$ and $r = .7$ (Reschly & Grimes, 2002; Sternberg, Grigorenko, & Bundy, 2001). This correlation can provide evidence for the validity of an instrument by demonstrating its relation to other variables, thus enhancing our understanding of it as an assessment tool (Smith, 2005). Therefore, the existence of this correlation also serves as a primary method by which to validate measures of cognitive ability.

*WISC-III and group-based achievement tests***.** The last version of the Wechsler Intelligence Scales, the WISC-III, is comprised of 13 subtests yielding three composite scores including the Verbal IQ (VIQ), the Performance IQ (PIQ), which emphasizes perceptual-motor abilities, and the Full Scale IQ (FSIQ), thought to be the best indicator of overall intellectual functioning. From these three composite scores come four factor-based Index scores: (1) Verbal Comprehension (VCI), (2) Perceptual Organization (POI), (3) Freedom from Distractibility (FDI), and (4) Processing Speed (PSI). The manual for the WISC-III, demonstrates evidence for its validity by providing relationships with other measures, including the following group-administered achievement tests: the Comprehensive Tests of Basic Skills (CTBS), the Iowa Tests of Basic Skills (ITBS), the California Achievement Test (CAT), the Metropolitan Achievement Test, and the Stanford Achievement Test. When correlated with these group-based achievement tests, a high correlation of $r = .74$ was found between the WISC-III FSIQ and Total Achievement, as well as the VIQ and Total Achievement. Of the Index scores, the VCI correlated most highly with Total Achievement ($r = .70$; Wechsler, 1991).

*WISC-IV and achievement tests***.** Unfortunately, there are no data to link the WISC-IV and group tests of achievement. The WISC-IV manual provides evidence of its relation to achievement by presenting its correlations with the subtest and composite scores of the Wechsler Individual Achievement Test, Second Edition (WIAT-II). Comprised of 10 core and 2 supplemental subtests, the WISC-IV yields a Full Scale IQ score (FSIQ), as well as four Index scores: (1) Verbal Comprehension (VCI), (2), Perceptual Reasoning (PRI), (3) Working Memory (WMI), and (4) Processing Speed (PSI). With the exception of the PSI-

Total Achievement correlation (r = .58), each of the Index (VCI, PRI, WMI) scores as well as the FSIQ correlated highly with Total Achievement (r = .80, r = .71, r = .71, r = .87 respectively). It is important to note that, although no evidence of a correlation between the WISC-IV and group-based achievement tests was provided, the high correlation between the WISC-III and WISC-IV FSIQ scores (r = .89) suggests a great deal of overlap in the construct of these two instruments. Thus, it can be concluded that the WISC-IV is also likely to correlate with group-based measures of achievement (Wechsler, 2003), although the substantial changes to the WISC-IV described earlier may affect correlations with group achievement tests, particularly for substantially altered indexes (e.g., PRI).

*Diagnosing Specific Learning Disabilities (SLD)***.** In addition to providing evidence for the validity of an instrument, the correlation between intellectual functioning and academic achievement is relevant to the school context. This correlation is most notable with regard to the diagnosis of learning disabilities (LD), as these evaluations account for over half of those conducted by school psychologists (Reschly & Grimes, 2002). More specifically, 52% of those students identified as having a disability and thus receiving special education services under the Individuals with Disabilities Education Act (IDEA) qualify under the LD category and, since the enactment of federal legislation, comparisons between intellectual ability and achievement have served as the primary condition by which to determine special education eligibility (Gresham, 2005; Smith, 2005; Wechsler, 1991).

Currently, SLD is defined as

> "A disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest

> itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. This term includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. This term does not include children who have learning problems that are primarily the result of visual, hearing, or motor disabilities; mental retardation; or environmental, cultural, or economic disadvantage" (§602.26, p. 13).

Historically, the diagnosis of SLD rests on comparing students' achievement levels to the achievement that is expected given their level of intellectual functioning. This discrepancy approach to LD classification is based on intellectual ability, as measured by a standardized test of cognitive ability (i.e., IQ), and academic achievement, as measured by a standardized test of school achievement. LD is said to exist when the comparison between ability and achievement shows a "severe discrepancy" in at least one academic area (e.g., reading; Braden, 2003; Dombrowski, Kamphaus, & Reynolds, 2004, p. 366).

The widespread usage of the discrepancy formula has caused some to refer to it as the "cornerstone of LD eligibility decisions" (Dombrowski, Kamphaus, & Reynolds, 2004, p. 365). However, changes to the new Individuals with Disabilities Education Improvement Act (2004) are intended to modify SLD diagnosis. For one, although states may permit the IQ-achievement discrepancy to identify SLD, they may not require it. Current regulations now emphasize students' performance with regard to state standards, and specifically allow states to develop models for identifying SLD using tests of state standards. Thus, it is

possible that states will develop alternative models by which to diagnose SLD that incorporate these state-based achievement tests.

The High-Stakes Culture and Rise of Group-Based Tests

The recent standards-based reform movement spawned numerous changes in the school context, including an increase in the use and the relevance of group-based standardized achievement tests to assess student knowledge. The movement, which began in the latter part of the 1980s, aims to improve the quality of education in the United States by delineating the information and skills educators should teach and students should learn and by holding schools and students accountable for meeting predetermined performance standards. Most states have responded to governmental legislation (e.g., Title I of the Elementary and Secondary Education Act) by administering large-scale standardized achievement tests in grades three through eight, and at least once in high school, to assess student knowledge.

In terms of the actual tests, states may choose to develop their own or to purchase them from the test publisher; however, they most often contract with publishers to create tests customized to their specific educational standards. Currently, more than 40 states contract with one of three publishers: (1) CTB/McGraw Hill (TerraNova), (2) Harcourt Brace (Stanford Achievement Test), or (3) Riverside Publishing (Iowa Tests of Basic Skills). Although tests are produced to match state standards, they are generally modeled after these three national group-based achievement tests, and thus tend to be similar with regard to format, number of items, reliability, and validity (Elliott, Braden, & White, 2001).

Often referred to as high-stakes tests, these statewide group achievement test batteries have serious consequences for states, schools, and (in some states and schools) individual students and teachers. Individual consequences such as grade promotion, the attainment of a high school diploma, and tracking are not specified by federal law, policy, or other elements of the reform movement, but many states elect to attach such consequences to the tests (required by federal law) they use to monitor schools. Thus, this reform movement has resulted in fundamental changes in curriculum, instruction, and testing leading to what Allington and McGill-Franzen (1992) refer to as an "accountability culture" (p. 397; see also McDonnell, McLaughlin, & Morison, 1997; Mehrens & Kaminski, 1989; National Academy of Sciences, n.d.).

These high-stakes achievement tests differ from other measures of achievement in two fundamental ways. For one, because a school-based psychological assessment is intended to identify the presence of educational disabilities, it is only administered to students who have been referred by either parents or teachers. High-stakes achievement tests, on the other hand, are given to all students in a particular grade or course. The second major way in which high-stakes tests can be distinguished from other forms of assessment is that high-stakes tests are used in the making of general education decisions, such as promotion and retention, whereas psychological assessment is typically used in the determination of special education eligibility (Braden, 2003).

However, the Individuals with Disabilities Education Improvement Act (2004; IDEIA-04) is likely to affect the notion that only psychological assessment is involved in special education decisions. Not only does this recent legislation allow for students to be

classified as having an SLD via the use of the discrepancy formula or response to intervention, but, as noted earlier, it also places increasing emphasis on the state-approved grade-level standards by permitting states to incorporate them into their eligibility process. Previous federal definitions have not made explicit reference to standards, and the implicit measurement of how students meet those standards (e.g., high-stakes tests). If students are unable to adequately achieve, make sufficient progress, or if they demonstrate a pattern of strengths and weaknesses such that they are unable to meet state-based grade-level standards, they are eligible to be identified as having an SLD. Students' ability to meet these state-based grade-level standards is presumed to be best indicated by their performance on state tests. Because individualized tests of achievement are not specifically developed to measure a state's standards, they are unlikely to be acceptable as measures of "state-approved grade-level standards." In contrast, high-stakes tests are specifically designed to measure and be aligned with state standards. Thus, these high-stakes tests may now be an integral part of special education eligibility decisions.

North Carolina's ABC assessment program. North Carolina's assessment program was developed before federal legislation mandated these large-scale achievement tests; however, the state tests were designed and modeled on other large-scale achievement tests. North Carolina's School-Based Management and Accountability Program, referred to as the "ABCs," is the state's primary school improvement program for public education. Considered to be a high-stakes accountability program (Jones et al., 1999), the ABCs largely affects North Carolina's curriculum, instruction, assessment, and school personnel. The goals are as follows: (a) to provide strong local school accountability, (b) to emphasize a

mastery of basic subjects, namely reading and mathematics, and (c) to allow for as much local decision making as possible (Public Schools of North Carolina, 2005a, 2005b). The assessments that comprise North Carolina's program include the End-of-Grade (EOG) Tests of Reading Comprehension and Mathematics, which are administered each school year to those students in grades 3-8, and the End-of-Course (EOC) tests, which cover a variety of academic areas (e.g., Biology, United States History) and are administered at the high school level to each student in that respective course (Public Schools of North Carolina, 2005c). However, the focus of the present study will be on the EOGs in Reading and Mathematics, for the following reasons: (1) overall ease of study, (2) students are generally not identified as having an SLD at the secondary level, and (3) technical information surrounding the EOCs is unavailable.

The purpose of the North Carolina state-mandated achievement tests is "(a) to assure that all high school graduates possess those minimum skills and that knowledge thought necessary to function as a member of society, (b) to provide a means of identifying strengths and weaknesses in the education process in order to improve instructional delivery, and (c) to establish additional means for making the education system at the state, local, and school levels accountable to the public for results" (Public Schools of North Carolina, 2004b, p. 1). These tests are aligned with North Carolina's curriculum, the *Standard Course of Study*, and are thus intended to provide an accurate measurement of mastery of material in particular content areas (Public Schools of North Carolina, 2004a, 2005d). Because the test development process of these tests is similar to that of other large-scale achievement tests

(e.g., employing item characteristic curves, item-to-total ratios), it is likely that they will behave in much the same way.

Factors Affecting Correlations between Tests of Intelligence and Tests of State Standards

Three relationships previously presented lead to the claim that underlies the proposed study, which is that the WISC-IV is likely to correlate positively with the North Carolina achievement tests. One, there exists great overlap in the construct of the WISC-III and WISC-IV, as indicated by the high FSIQ-FSIQ correlation. Two, there is a high correlation between the previous version of the Wechsler Scales (i.e., the WISC-III) and various group-based tests of achievement. Three, because the North Carolina achievement tests were developed using similar procedures to those of other large-scale achievement tests, they are likely to behave in similar ways.

However, high-stakes achievement tests differ from other achievement tests with respect to how results are reported, and in the type of content assessed. Previously, achievement tests emphasized reports of norm-referenced scores and scales, which compared a student's score to that of other students of the same age or grade. The numerical approach adopted in norm-referenced evaluation produces a normal distribution of scores. However, the standards-based reform movement focuses on the attainment of state-approved educational standards, rather than normative position. Consequently, states emphasize criterion-referenced scores and scales in their assessment programs, in part because No Child Left Behind (NCLB) requires states to identify whether a student is proficient or advanced with respect to grade-level state standards. Unlike norm-referenced evaluation, criterion-referenced evaluation does not tend to produce a normal distribution of scores (e.g., all

students could potentially score 4s on the EOGs). Thus, this fundamental shift away from valuing norm-referenced statements of results to criterion-referenced statements of results might weaken the correlation with intelligence tests.

Another way in which these high-stakes achievement tests can be distinguished from other achievement tests is via content. Because NCLB requires state tests to align with state standards, these tests assess a student's ability to meet educational standards and are, thus, more restrictive in content then are the nationally standardized individual and group achievement tests, which cover a broad or random sample of academic achievement domains. The restrictive nature of the high-stakes tests may result in their being less associated with intellectual ability than a test consisting of more broadly sampled content. An additional issue surrounding test content relates to the cognitive complexity of state standards (i.e., some states may have fairly superficial or less cognitively demanding standards than others), which is likely to influence the degree to which performance is linked to intelligence (Braden, 2003; Langenfeld, Thurlow, & Scott, 1996; Overton, 2006).

In addition to distinctions in reporting (i.e., scaling) and content between nationally standardized tests and tests of state standards, there are some additional factors that could influence the relationship between tests of intelligence and achievement. Some of these are considered in the following paragraphs.

Changes in referral patterns. Because students with disabilities used to be excluded from large-scale accountability systems, they would often be reclassified so they would not be required to participate in standard state assessments (Langenfeld, Thurlow, & Scott, 1996; McDonnell, McLaughlin, & Morison, 1997). To stop these exclusionary practices, certain

provisions were developed to allow for accommodations and modifications in administration procedures as well as alternate assessments for students whose disabilities preclude their involvement in the general assessments (Schulte, Villwock, Whichard, & Stallings, 2001; US Department of Education, 2005). Schulte and her colleagues (2001) evaluated the impact of these provisions by analyzing the reading scores of 461 students with learning disabilities across a 5-year period (i.e., 1994-1998) in one school district. Results indicated that testing program participation did increase by 11% during the specified time period, but there was a simultaneous increase (20%) in the number of students identified as learning disabled. Thus, it is likely that the high stakes attached to these types of assessments is causing shifts in referral patterns, such that teachers are now referring students whom they would not have previously deemed eligible. It is possible that teachers are now referring all students they consider to be lower functioning as opposed to only those who they believe will meet the ability-achievement discrepancy criterion. If true, this change would result in a restriction of possible achievement values, thereby affecting the IQ-achievement correlation.

Narrowing of the curriculum. An additional concern surrounding these high-stakes tests is that teachers will engage in curricular narrowing by limiting their instruction to merely those domains covered on the tests. Jones and her colleagues (1999) evaluated the effect of high-stakes testing on the instruction practices of teachers in North Carolina. Teacher surveys indicated that the majority of teachers (80%) believed students to be spending more than 20% of their time practicing for end-of-grade tests. Furthermore, results demonstrated that teachers spend most of their class time preparing students in reading (401 minutes), mathematics (292 minutes), and writing (198 minutes), as these are the domains

tested on the state's assessments. Compare the mean number of minutes spent on each of these areas with 102 minutes in social studies, 99 in science, and 61 in physical education. According to Mehrens (1998), this narrowing of the curriculum spoils the inferences one can make between the test and the domain being sampled. More specifically, higher test scores do not necessarily indicate increased learning because repeated exposure to the same material is likely to result in inflated test scores. Thus, curricular narrowing is likely to weaken the correlation between IQ and achievement.

Corruptibility and teaching to the test. Another potential factor affecting the IQ-achievement correlation is corruptibility, or unethical behavior resulting in misleading information. Gay (1990) examined the presence of irregularities in standardized test administration by surveying 168 teachers with an instrument developed to target those testing irregularities delineated by Buss and Novick (as cited in Gay, 1990). These include: (1) inaccurate timing, (2) altering answer sheets, (3) coaching, (4) errors in scoring/reporting, (5) student cheating, and (6) teaching to the test. Results indicated 35% of teachers were cognizant of or had participated in these irregularities. Teaching to the test, by copying test sections and using them to teach students, was found to be the most common infraction. The survey suggested teachers tend to engage in these irregularities because of pressure from superintendents, principals, peers, and parents to improve test results.

A study conducted by Firestone and Mayrowetz (2000) to examine teacher responses to assessment programs in Maryland (high stakes) and Maine (no stakes) also found a linkage between teaching to the test and external pressures associated with testing. In Maryland, where external pressures were high, teachers incorporated items similar to those on the tests

into their instruction, discussed with the students how the tests were scored, and explained how to succeed on them. In Maine, teachers did not exhibit these responses to testing.

Gay (1990) claims that "infractions of the publishers' guidelines destroy the validity of test results" (p. 93). Thus, like curricular narrowing, issues of corruptibility (e.g., teaching to the test) are likely to attenuate the correlation between IQ and achievement. Central to the construct of intelligence is the rapidity with which individuals acquire information and the degree to which they can apply that information to novel situations. Cheating and/or prolonged exposure to the same material disassociates the link between intellectual ability and test scores by reducing the degree to which intellectual ability affects information acquisition.

Rationale for this Study

In summary, the proposed study will extend the historical study of IQ-achievement correlations, using the WISC-IV to assess intellectual functioning and the high-stakes tests comprising North Carolina's assessment program to measure achievement. It is important to understand the nature of the relationship between these instruments on two levels. On a contextual level, evidence of a correlation between the WISC-IV and group-administered state-based achievement tests would provide support for the validity of the WISC-IV by demonstrating its relation to other variables (here, achievement). In addition, evidence of this correlation will enhance the test user's understanding of the WISC-IV as an assessment tool. On a conceptual level, evidence of a correlation between the WISC-IV and these high-stakes tests might inform future policy procedures that use state-based tests in the identification of LD. Such procedures seem more likely, in that recent legislation

incorporates the attainment of state-approved grade-level standards into its procedures for SLD identification. However, there are no current studies of the relationship between state tests and tests of intelligence to inform the issue; this study would be a first step to develop information that might inform policies and practices relating SLD to state tests.

Hypotheses

As noted earlier, there is a long history establishing the relationship between measures of intelligence and achievement. Specifically, the Wechsler Scales have shown strong positive correlations with group tests of achievement (i.e., the WISC-III), and with individual measures of achievement (e.g., the WISC-IV and its predecessors). Therefore, the predictions for the proposed study reflect these relationships:

1. A confidence interval placed around the correlation coefficient for WISC-IV Full Scale IQ (FSIQ) scores and EOG-Reading scores (norm-referenced) will capture the historic range of correlations (r = .4 and r = .7), and will not include r = 0.
2. A confidence interval placed around obtained correlation coefficient for WISC-IV Full Scale IQ (FSIQ) scores and EOG-Mathematics scores (norm-referenced) will capture the range of r = .4 and r = .7, and will not include r = 0.

Hypotheses will be supported if confidence intervals placed around obtained correlation coefficients capture r = .6, as this is roughly halfway between the range of historical correlations, and do not include r = 0.

Because there is no research in the extant literature that examines the relationship between IQ and high-stakes tests (based on a search of the PsycInfo database), the proposed study is exploratory in nature. Thus, the relationship between the other WISC-IV Indexes

and EOG-Reading and EOG-Mathematics scores will be evaluated and reported, but no predictions surrounding these relationships will be made, with the exception that these relationships are hypothesized to exceed $\rho = 0$.

Chapter 2

Methods

Participants and Students

This study differentiates between participants, those individuals who were involved in supplying data, and students, whose scores provided the data for analyses. Participants were practicing school psychologists who were recruited through sign-up sheets at area conferences (e.g., North Carolina School Psychology Association [NCSPA] annual fall conference) and via a posting on an NC psychology-related listserve (see Appendices A and B). Students were those who attended North Carolina schools and had been tested on the WISC-IV and the NC EOGs (and possibly other tests of achievement). They must have been tested on the WISC-IV as part of a psychological evaluation (and therefore between the ages of 6 and 17) and have developmental scale scores from both North Carolina EOG educational tests (i.e., Reading and Mathematics).

Five school psychologists (participants) consented to participate and provided information on 80 students. Three of these psychologists (60%) described their student information as coming predominantly from the piedmont region of North Carolina, as compared to one psychologist (20%) who identified the mountain region and one (20%) who identified the coast. With regard to school district, two (40%) described the one in which they work as rural and three (60%) described theirs as both rural and urban. No psychologist considered his/her school district to be solely urban.

Although none of the 80 students for whom the participants provided information were excluded on the basis of disability, gender, ethnicity, or any other characteristic, four were removed from analyses due to their missing an EOG-Mathematics score (n = 1) or their

having taken a version of the EOG-Mathematics test that differed from that of the other students (n = 3). That is, a new version of the EOG-Mathematics with a different scale of scores was utilized beginning in 2006. Information for these three students was yielded from the 2005 version of the EOG-Mathematics. Removal of these data allowed for a cleaner data set (i.e., one comprised of EOG scores from the same version) without compromising the power of the obtained sample. The omission of these students resulted in a sample totaling 76 students. Of these, 49 (64.5%) were males and 27 (35.5%) were females. The students ranged in age from 9.0 years to 16.75 years, with the mean age being 13.0 years (SD = 2.0 years). The sample was comprised of those from White (44.7%), Black (39.5%), Hispanic (9.2%), Native American (5.3%), and Mixed (White/Black; 1.3%) backgrounds. The majority of the students were administered the WISC-IV as part of a triennial reevaluation (57.9%); however, some completed the test because of a special education referral (39.5%), a referral for the gifted program (1.3%), or some other reason (1.3%). At the time of the WISC-IV administration, most students had a diagnosed disability (68.4%) and were receiving special education services (64.5%). Only five students (6.6%) were classified as Limited English Proficient (LEP) when they were given the WISC-IV.

These and additional descriptive data delineated by student grade at the time of EOG administration are outlined in Table 1. These data were subjected to chi-square tests to determine if student information was normally distributed across grades. It is important to note that a chi square test was only calculated for two of the four possible responses to the item pertaining to the reason for WISC-IV administration (i.e., referral for special education

and triennial reevaluation) so as to avoid expected values near zero given the extremely small *n* values for each of the other responses (i.e., referral for gifted program, other).

Results from chi-square tests indicate that student information surrounding: (1) the reason for WISC-IV administration (χ^2 (5, N = 74) = 21.14, p = .00) and (2) whether the student was receiving special education services at the time of the WISC-IV administration (x^2 (5, N = 76) = 12.68, p = .03) do not follow random frequency distributions. That is, students were not equally likely to be classified under each of the possible conditions across grades. When evaluating each of these patterns, the data show: (1) students were more often described as having taken the WISC-IV for purposes of a special education referral during the younger grades and as part of a triennial reevaluation during the older grades, and (2) students in the older grades were more likely to have been receiving special education services than were those in the younger grades.

Procedure

Interested psychologists provided their names and email addresses in response to recruitment. Once psychologists expressed their interest in participation, they were sent a consent form and an outline of the steps required for participation (see Appendices C and D), as well as the IRB Letter of Approval, via email. The consent form included a space designated for the psychologist to supply a password. Due to the insecure nature of email, psychologists were asked to print, sign, and return the consent form by ground mail or facsimile. The password was then used to encrypt files in an effort to ensure data in the file remained confidential. It is important to note that minor changes were submitted to the IRB

since the time of the original exemption (i.e., the title of the study, the last name of the principal investigator, and the number of participants), but these were all approved.

Table 1

Descriptive Statistics by Student Grade at the Time of EOG Administration and for the Total Group

Grade	3	4	5	6	7	8	Total N
Gender							
Male	10	4	6	14	11	4	49
Female	6	4	3	2	11	1	27
Ethnicity							
White	7	6	3	9	6	3	34
Black	9	1	5	3	12	0	30
Hispanic	0	1	1	3	2	0	7
Native American	0	0	0	1	1	2	4
Mixed – White/Black	0	0	0	0	1	0	1
WISC-IV Testing Reason*							
Referral for Special Education	11	7	3	3	4	2	30
Referral for Gifted Program	0	0	0	1	0	0	1
Triennial Reevaluation	5	1	6	12	18	2	44

Table 1 (continued)

Diagnosed Disability at WISC-IV Testing							
No	8	4	2	4	3	2	23
Yes	8	3	7	12	19	3	52
Missing Data	0	1	0	0	0	0	1
Receiving Special Education Services at WISC-IV Testing*							
No	9	5	3	4	3	3	27
Yes	7	3	6	12	19	2	49
Limited English Proficient (LEP)							
No	16	7	9	14	20	5	71
Yes	0	1	0	2	2	0	5
Current Age in Years *M*	10.31	10.96	12.10	13.51	14.68	15.77	12.95
(SD)	(0.86)	(0.75)	(0.58)	(0.89)	(0.96)	(0.72)	(2.02)
Age in Years at WISC-IV Administration	9.97	10.57	12.10	12.38	14.00	15.27	12.31
	(0.93)	(0.87)	(0.68)	(0.71)	(0.87)	(0.58)	(1.88)
Age in Years at EOG Administration	9.57	10.33	11.80	12.70	13.72	14.83	12.12
	(0.67)	(0.83)	(0.37)	(0.66)	(0.82)	(0.42)	(1.88)

Table 1 (continued)

* $p < .05$.

Upon receipt of the consent form, psychologists were sent two pre-formatted Microsoft Excel spreadsheets to complete via email, one for demographic and one for test data. These spreadsheets were encrypted; thus, to be opened, psychologists were required to enter their password. The spreadsheets were also unique to each participant, meaning that they contained empty cells with pre-determined student numbers unique to the psychologist. Also, no other psychologist's data were shared with another psychologist.

The psychologist entered archival data describing the student. Once psychologists entered their information into the spreadsheets, they returned the encrypted files via email. Each psychologist was encouraged to include information about as many students as were available to the psychologist, provided the data were collected in the course of routine educational procedures (i.e., were archival, and not collected solely for research purposes).

Measures

WISC-IV. The WISC-IV is comprised of 10 core subtests and 5 supplemental subtests emphasizing verbal and nonverbal reasoning skills, as well as working memory and processing speed abilities. The 10 core subtests include Block Design, Similarities, Digit Span, Picture Concepts, Coding, Vocabulary, Letter-Number Sequencing, Matrix Reasoning, Comprehension, and Symbol Search. Supplemental subtests are Picture Completion, Cancellation, Information, Arithmetic, and Word Reasoning. WISC-IV standard scores range from 40 to 160 and have a mean of 100 and a standard deviation of15. Approximately

half of all children will have scores that fall between 90 and 109. A technical manual is available that provides psychometric, as well as normative data (Wechsler, 2003).

Evidence for reliability is presented in terms of internal consistency, test-retest stability, and interscorer agreement. Internal consistency was examined by calculating split-half correlations (or test-retest, depending upon the nature of the subtest). These correlations yielded average reliability coefficients ranging from .79 (Symbol Search and Cancellation) to .90 (Letter-Number Sequencing), with overall subtest reliability coefficients improving as compared to those obtained from the WISC-III. The reliability coefficients for the composite scales tend to be higher than those of the individual subtests, ranging from .88 (Processing Speed) to .97 (Full Scale). All subtest values were recalculated utilizing 16 special groups (e.g., ADHD, LD) and were found to be comparable. To compute test-retest reliability, the WISC-IV was administered to each of 243 children across five age groups (i.e., 6-7, 8-9, 10-11, 12-13, and 14-16) twice in an average of 32 days. Data indicate that WISC-IV scores "possess adequate stability across time for all five age groups," with average corrected stability coefficients ranging from the .70s (Picture Concepts, Cancellation, and Arithmetic) to .92 (Vocabulary) for subtest scores, and ranging from the high .80s to .90s for composite scores (p. 39). Interscorer agreement was computed following double-scoring by two independent observers, and was found to be quite high (.98 to .99). However, additional studies were conducted to evaluate interscorer agreement of those subtests deemed to require more judgment in scoring (i.e., Similarities, Vocabulary, Comprehension, Information, and Word Reasoning). Results indicate high interscorer reliabilities (.95 to .98) for these subtests as well.

As previously mentioned, various sources of evidence are provided to support the validity of the WISC-IV. With regard to test content, reviewers and experts examined the relationship between the test's content and the construct it is designed to measure. The authors note a "strong theoretical and empirical evidence of validity" (Wechsler, 2003, p. 48) as a result of literature reviews, expert consultation, and empirical and qualitative examinations conducted during the instrument's development, thereby providing evidence based on response processes. Evidence based on internal structure was evaluated through a series of covariance studies. Although all subtests should correlate positively, those subtests part of the same index should correlate most highly with each other. All intersubtest correlations were found to be statistically significant, and those subtests intended to measure similar functions correlated more highly with each other than did those measuring different types of functioning. The manual also presents data yielded from factor analysis. Results indicate each of the subtests having factor loadings above .60 with their intended index, and significantly lower factor loadings on each of the other indexes, with Picture Concepts as the only exception.

As noted earlier, another line of evidence is based on the instrument's relationship with other variables. Here, the relationships between the WISC-IV and other measures of cognitive ability, achievement, memory, adaptive behavior, giftedness, and emotional intelligence were examined. Correlations of WISC-IV scores with those obtained from achievement measures ranged from .20 to .80. Special group studies were conducted to evaluate scores from groups of children expected to score very differently from one another, including those with learning disabilities. Results were consistent with expectations based on

previous research, thus providing support for validity. However, no information concerning the IQ-achievement correlations for these special groups was provided. Although a simple warning is offered to professionals urging them to be cognizant of both intended and unintended consequences of test use, no actual evidence based on the consequences of testing is provided (Wechsler, 2003).

End-of-Grade Tests. The North Carolina End-of-Grade Tests of Reading Comprehension and Mathematics are multiple-choice tests administered during the last three weeks of each school year to those students in grades 3-8. For the EOG-Reading, students are asked to read eight or nine (depending on the grade) literary and informational texts, and then answer approximately 50 questions related to these selections. The test items are organized into four categories (i.e. Cognition, Interpretation, Critical Stance, and Connections) and are designed to assess reading comprehension and vocabulary. A technical manual providing reliability and validity information is available. Internal consistency reliability estimates are above .92 for all grade levels, with the standard error of measurement falling between 2 and 6 points for each individual test score, depending on the distance of the score from the mean. Content validity is described in terms of an "explicit statement of the constructs or concepts being measured" (Public Schools of North Carolina, 2004c, p. 64), here considered to be Cognition, Interpretation, Critical Stance, and Connections. Tables outlining the percentage of items by each of these four strands are provided for each grade level. Sources of criterion-related (concurrent) validity are correlations between EOG-Reading scores and teachers' judgments of student achievement, expected grade, and assigned achievement levels. Correlation coefficients range from .49 to .65, indicating

moderate correlations between scale scores and these variables (Public Schools of North Carolina, 2004c).

The EOG-Mathematics is designed to test student achievement in the areas of: (a) Number Sense, Numeration, and Numerical Operations, (b) Spatial Sense, Measurement, and Geometry, (c) Patterns, Relationships, and Functions, (d) Data, Probability, and Statistics, and (e) Algebra. The assessment is comprised of 80 questions and is divided into calculator-active (24 questions) and calculator-inactive (56 questions) sections (except in grade 8, for which the entire test is calculator-active), and requires students to interpret information in contextually relevant ways to generate an accurate response. Students unable to pass the grade 8 EOGs are provided with remedial instruction and additional opportunities to take the North Carolina Competency Tests in Reading and Mathematics, as meeting this competency standard is a requirement to receive a high school diploma (Public Schools of North Carolina, 2003, 2005c, 2005d). Internal consistency reliability estimates are .93 or .94 for all grade levels, with the standard error of measurement for individual test scores falling between 2 and 6 points depending on the distance of the score from the mean (Public Schools of North Carolina, 2006).

Evidence of content validity is provided in the explicit statement that Number Sense, Numeration, and Numerical Operations; Spatial Sense, Measurement, and Geometry; Patterns, Relationships, and Functions; and Data, Probability, and Statistics are the constructs or concepts measured. Instructional validity is measured by administering questionnaires to teachers asking them to evaluate how aligned the test content is with goals and objectives of the curriculum, how clear and concise the items are, whether content is balanced with regard

to such factors as ethnicity, race, and sex, and whether each of the items has only one best answer. All objectives were deemed to have been met at either a "superior" or "high" degree by at least 48% of the teachers sampled. Sources of criterion-related (concurrent) validity are the same as those described in the discussion of the EOG-Reading. Correlation coefficients range from .49 to .89, indicating moderate to strong correlations between scale scores and these previously mentioned variables. For both the EOG-Reading and the EOG-Mathematics population distributions of proficiency within grades approximate a normal distribution (Public Schools of North Carolina, 2004c; 2006).

EOG score reporting. Raw scores, or the number of questions answered correctly, are converted into *developmental scale scores* (sometimes referred to as simply *scale scores*). These scores allow the child's performance to be compared from year to year. However, scales are subject-specific, so scores cannot be compared across academic content areas. The reading scale scores range from 216 to 290 (as of 2002-2003), whereas the mathematics scale scores range from 311 to 384 (as of 2005-2006). Scores are also reported in terms of *percentiles*. This score allows the child's performance to be compared to all of those North Carolina students who took the test during the norming year, or the first year the test was administered. Specifically, percentiles range from 1 to 99, and tell the percentage of students to whom a student scored equal or better than on that test (Public Schools of North Carolina, 2003; 2007).

Developmental scale scores are also reported relative to performance standards that define grade-level expectations (i.e., criterion-referenced statements). These criteria were set with a combination of teacher judgment and data-driven iterations (Public Schools of North

Carolina, 2004c). Performance categories have four levels: (1) *I*, insufficient mastery of knowledge and skills in the subject area to be successful at the next grade level, (2) *II*, inconsistent mastery of knowledge and skills indicating minimal preparation to be successful at the next grade level, (3) *III*, consistent demonstration of mastery of knowledge and skills indicating good preparation for a more advanced level in the content area, and (4) *IV*, consistent superior performance clearly beyond that required to be proficient at grade-level work indicating very good preparation for a more advanced level in the content area (Public Schools of North Carolina, 2003, 2005d).

Chapter 3

Results

Preliminary Analyses

It was originally proposed that a one-tailed power analysis would be conducted to determine the power of the obtained sample to detect a significant correlation between scores of intellectual functioning and academic achievement, assuming a population correlation of $\rho = .4$ (here considered the lower bound of an acceptable correlation) and an alpha level of .01. However, no power analysis was performed because significant correlations between these two constructs were found (which are described below), thereby indicating that the sample provided sufficient power for rejecting the null hypothesis.

Before any analyses were conducted, all EOG achievement data (i.e., developmental scale scores) were rescaled to normalized scores for grades using state data available from the North Carolina Department of Public Instruction. More specifically, scale scores were transformed into normalized (i.e., *z*) scores derived from students who took the same grade test during the same year as the student. For example, if a student took the third grade EOG-Reading test in 2001, that score was converted to a z score using the mean and standard deviation of all NC students who also took the third grade EOG-Reading test in 2001 (i.e., [score – *M*grade]/*SD*grade). Because data yielded from those students who took the 2005 EOG-Mathematics test were dropped (as this version differed from those taken by the majority of the students) and information from the 2007 EOG was not available at the time of analyses, all scale scores were compared to those from the 2006 version. Table 2 presents the means and standard deviations used for score conversion, where z = (actual score –

M)/*SD*. Table 3 presents the comparative means of the obtained sample on both the WISC-IV and the EOGs.

Table 2

Mean Scale Scores and Standard Deviations for 2005-2006 EOG Tests

Grade	3	4	5	6	7	8
EOG-Reading Achievement						
M	248.6	253.1	257.1	259.2	261.9	263.9
(SD)	(8.8)	(8.6)	(7.8)	(8.1)	(8.6)	(8.6)
EOG-Math Achievement						
M	343.2	348.9	353.7	354.9	357.8	359.2
(SD)	(9.7)	(9.5)	(9.2)	(9.7)	(9.6)	(9.2)

The procedures performed as part of this study likely yielded more of an "at-risk" sample than would have been obtained had random selection methods been employed. Thus, restriction of range with regard to both intellectual functioning and academic achievement levels could have occurred. To ensure that obtained correlation coefficients were not attenuated by significantly homogeneous samples, an *F* test for equal variances was conducted in which the obtained variance for a given test was compared to the population variance for that test to determine whether they differed when the variance of the sample was

less than that of the population. Here, *F* tests were performed for the WISC-IV, as well as for both EOG tests. An *F* test of the variance of the WISC-IV FSIQs revealed no restriction of range [F (75, ∞) = 1.10, NS]. Before conducting an *F* test of the variance of both EOG

Table 3

Comparative Means of Descriptive Statistics by Student Grade and for the Total Group

Grade	3	4	5	6	7	8	Total
Full Scale IQ	85.25	96.62	86.89	93.12	87.68	79.20	88.61
	(13.36)	(12.05)	(20.79)	(13.68)	(11.21)	(17.23)	(14.28)
EOG-Reading	241.75	246.75	251.56	256.06	251.41	255.60	250.16
Achievement	(6.52)	(12.27)	(7.16)	(8.82)	(9.48)	(4.10)	(9.80)
EOG-Math	335.37	343.75	352.44	353.06	349.73	351.00	347.18
Achievement	(7.37)	(8.31)	(9.86)	(9.33)	(7.53)	(6.96)	(10.40)
Normalized EOG-	-0.78	-0.74	-0.71	-0.39	-1.22	-0.97	-0.82
Reading	(0.74)	(1.43)	(0.92)	(1.09)	(1.10)	(0.48)	(1.03)
Normalized EOG-Math	-0.81	-0.54	-0.14	-0.19	-0.84	-0.89	-0.58
	(0.76)	(0.87)	(1.07)	(0.96)	(0.78)	(0.69)	(0.89)

Note: normalized scores $M = 0$, $SD = 1$

tests, each score was converted into a z-score to allow for cross-grade comparisons (as each grade version of the test had its own population variance). The variance of the EOG-Reading scores was found to be 1.03, which is greater than the population variance of 1, so no *F* test was conducted, as the range could not be restricted. An *F* test of the variance of the EOG-Mathematics scores indicated there did not exist restriction of range [F (75, ∞) = 1.27, NS]. Thus, no correction formula was applied to these data.

As children age, their intellectual ability and their academic achievement become more correlated. Thus, differences in the IQ-achievement correlation across grades may occur. It is also possible that students' LEP status affects the degree to which their intellectual functioning is correlated with their academic achievement. Thus, a linear regression was performed to test whether grade and/or LEP status were uniquely significant predictors of academic achievement. FSIQ, grade, and LEP status were entered to predict EOG-Reading and then EOG-Mathematics scores. Grade was not a significant predictor of either reading ($\beta = -0.09$, $t(72) = -0.88$, NS) or math ($\beta = 0.00$, $t(72) = 0.04$, NS); likewise, LEP status did not predict either reading ($\beta = -0.08$, t (72) = -0.79, NS) or math ($\beta = -0.05$, t (72) = 0.53, NS) achievement. Therefore, all data were combined as opposed to being analyzed separately by grade or LEP status (see Appendix E for the source tables of these analyses).

Hypotheses

The two hypotheses in this study concerned the relationship between intellectual functioning and academic achievement. Hypothesis One predicted that the relationship between WISC-IV Full Scale IQ (FSIQ) scores and norm-referenced EOG-Reading scores would be both positive and significant, and that a confidence interval placed around the

correlation coefficient would capture the value $r = .6$, as this is roughly halfway between the range of historical IQ-achievement correlations of $r = .4$ and $r = .7$, and not include $r = 0$. Hypothesis Two predicted that this same relationship would hold for FSIQ scores and EOG-Mathematics scores. Two bivariate correlations were performed to test these hypotheses using FSIQ scores and normalized EOG scores. Because there was more than one contrast conducted on the same sample, the alpha level for these correlations was set at .01 to ensure an experiment-wise error rate less than .05. After obtaining the correlation coefficients, confidence intervals were placed around these values. Importantly, note that the asymmetrical nature of the confidence intervals is a function of the distribution of correlation coefficients, for Fischer's *z* was used. Table 4 below outlines the results of these correlations.

Table 4

Results of Bivariate Correlations of WISC-IV FSIQ and Adjusted EOG Scores

FSIQ	EOG-Reading	EOG-Mathematics
Lower (.5%)	.23	.44
r	.49*	.65*
Upper (99.5%)	.68	.79

* $p < .01$.

It was proposed that Hypotheses One and Two would be supported if the confidence interval surrounding each of the correlation coefficients included $r = .6$ but did not include zero. Both hypotheses were supported, thereby offering evidence for the validity of the WISC-IV. To test whether the correlation between FSIQ and EOG-Mathematics was significantly higher than the one between FSIQ and EOG-Reading, a *t*-test designed to examine the difference of two dependent correlations from the same sample was conducted. That is, because the correlations were yielded from the same sample and were therefore not independent, a special test was used in which *t* was calculated using the following formula:

$$t = (r_{xy} - r_{zy}) * \text{SQRT}\ [\{(n - 3)\ (1 + r_{xz})\} / \{2\ (1 - r_{xy}^2 - r_{xz}^2 - r_{zy}^2) + (2r_{xy} * r_{xz} * r_{xy})\}].$$

Here, "*n*" is the sample size, "x" is the EOG-Reading score, "y" is the FSIQ score and "z" is the EOG-Mathematics score. Because the absolute value of the calculated t-value was greater than the cut-off value on the t-table using *n* -3 degrees of freedom, the difference in the correlations is significant ($t(73) = -2.46$, $p < .05$).

Supplemental Analyses

In addition to testing each of the proposed hypotheses, supplemental analyses were conducted to further examine the relationship between intellectual functioning and academic achievement. More specifically, bivariate correlations between each of the WISC-IV Indexes (i.e., Verbal, Perceptual Reasoning, Working Memory, and Processing Speed) and adjusted EOG-Reading and -Mathematics scores were calculated, and confidence intervals

were placed around the obtained coefficients. Given the exploratory nature of this study, each correlation was tested to determine if $r > 0$ at an alpha level of .05. Results from these correlations are provided in Table 5.

Results from these correlations demonstrate that the relationship between each WISC-IV Index and EOG test is reliably greater than zero. With the exception of the correlation between PRI and EOG-Reading scores, which is significant at an alpha level of .05, all correlations are significant at the .01 level. To examine whether the correlations between the Indexes and each of the EOG tests were significantly different from one another, the aforementioned *t*-test comparing two dependent correlations from the same sample was conducted within each Index/EOG pair of correlations. Results indicate that there does not

Table 5

Results of Bivariate Correlations of WISC-IV Indexes and Adjusted EOG Scores

Index	EOG-Reading	EOG-Mathematics
Confidence Interval		
VCI		
2.5%	.36	.43
r	.54**	.60**
97.5%	.68	.73

Table 5 (continued)

PRI		
2.5%	.03	.35
r	.25*	.53**
97.5%	.45	.68
WMI		
2.5%	.11	.29
r	.33**	.49**
97.5%	.52	.64
PSI		
2.5%	.17	.16
r	.38**	.37**
97.5%	.56	.55

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

exist a significant difference in correlations for the VCI ($t(73) = -2.46$, NS), the WMI ($t(73) = -1.95$, NS), or the PSI ($t(73) = 0.66$, NS); however, the PRI was found to correlate significantly more highly with the EOG-Mathematics than with the EOG-Reading ($t(73) = -3.74$, $p < .01$).

Scores were garnered from the following individually administered achievement measures: (1) the third edition of the Woodcock-Johnson Tests of Achievement (WJ-III), (2)

the normative update of the third edition of the Woodcock-Johnson Tests of Achievement, and (3) the second edition of the Kaufman Test of Educational Achievement (KTEA-II). However, given the relatively few contributions from the latter test (n = 2), KTEA-II data were removed from analyses. Data points yielded from both versions of the WJ-III (i.e., standard and normative update) were then collapsed across subject area into the following categories: (1) Broad Reading, (2) Basic Reading, (3) Reading Comprehension, (4) Broad Mathematics, (5) Mathematics Calculation, (6) Mathematics Reasoning, and (7) Broad Written Language. It was originally calculated that 16 data points from each subject area would allow for a power equal to .80, assuming a population correlation of ρ = .6 at an alpha level of .05. This minimum was not met for two subject areas: (1) Basic Writing Skills and (2) Written Expression; thus, these variables were removed from analyses.

Bivariate correlations were computed examining the relationship between WISC-IV FSIQ scores and WJ-III scores, as well as between WJ-III scores and EOG-Reading and -Mathematics scores. To evaluate the relationship between WJ-III and EOG scores, information from each WJ-III subject area was correlated with the relevant EOG test (i.e., Broad Reading, Basic Reading, Reading Comprehension and EOG-Reading; Broad Math, Math Calculation, Math Reasoning and EOG-Mathematics). Information pertaining to written language was not correlated with EOG scores as this is not a skill explicitly tested on either of the EOG tests. Confidence intervals were then placed around the correlations obtained. Because these correlations included individually administered tests of achievement (i.e., those measures used in establishing the historical IQ-achievement correlation), a

population correlation of $\rho = .6$ and a more liberal alpha level of .05 were used. Tables 6, 7, 8, 9, and 10 present the findings yielded from these analyses.

Examination of the yielded correlations indicates a positive relationship between WISC-IV FSIQ scores and WJ-III Reading Comprehension, Broad Mathematics, and Mathematics Calculation subtest scores, significant at the level of .01. A significant correlation was also found between WISC-IV FSIQ scores and WJ-III Broad Written Language scores at an alpha level of .05. No significant correlations were found for WISC-IV FSIQ and Broad Reading, Basic Reading, or Mathematics Reasoning scores. An evaluation of the correlations obtained when comparing EOG-Reading scores with WJ-III reading scores reveals significant relationships with the Broad Reading and Reading Comprehension subtests at the .01 level, but no relationship between EOG-Reading and

Table 6

Results of Bivariate Correlations of WISC-IV FSIQ and WJ-III Reading Scores

FSIQ	Broad Reading	Basic Reading	Reading Comprehension
n	24	25	24
2.5%	-.11	-.05	.20
r	.31	.36	.56**
97.5%	.63	.66	.79

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

Table 7

Results of Bivariate Correlations of WISC-IV FSIQ and WJ-III Mathematics Scores

FSIQ	Broad Math	Math Calculation	Math Reasoning
n	23	25	16
2.5%	.54	.42	-.19
r	.78**	.70**	.33
97.5%	.90	.86	.71

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

Table 8

Results of Bivariate Correlations of WISC-IV FSIQ and WJ-III Broad Written Language Scores

FSIQ	Broad Written Language
n	32
2.5%	.02
r	.36*
97.5%	.63

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

Table 9

Results of Bivariate Correlations of Adjusted EOG-Reading and WJ-III Reading Scores

EOG-Reading	Broad Reading	Basic Reading	Reading Comprehension
n	24	25	24
2.5%	.43	-.10	.29
r	.71**	.31	.62**
97.5%	.87	.63	.82

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

Table 10

Results of Bivariate Correlations of Adjusted EOG-Mathematics and WJ-III Mathematics Scores

EOG-Math	Broad Math	Math Calculation	Math Reasoning
n	23	25	16
2.5%	.12	.28	-.31
r	.51*	.61**	.21
97.5%	.76	.81	.64

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

Basic Reading scores. As for the tests relating to mathematics, a significant relationship was obtained between EOG-Mathematics scores and both Mathematics Calculation and Broad Mathematics scores, at alpha levels of .01 and .05, respectively. However, a significant relationship was not found to exist between EOG- Mathematics and Mathematics Reasoning scores.

Chapter 4

Discussion

The major purpose of the present study was to examine the relationship between intellectual functioning and academic achievement by comparing the correlations between WISC-IV scores and North Carolina state achievement test scores (i.e., the EOGs) with the historical IQ-achievement correlations yielded from various measures of intellectual functioning and individually administered tests of achievement. Two hypotheses were generated based on relevant research and tested using correlational methods. Several supplementary analyses were performed to further explore the IQ-achievement relationship.

The two hypotheses postulated that the relationship between WISC-IV FSIQ scores and norm-referenced EOG scores (reading and mathematics) would capture the range of historic correlations. Both hypotheses were supported, as confidence intervals placed around correlations between FSIQ and EOG-Reading and FSIQ and EOG-Mathematics included r = .6 and excluded r = 0. Although much previous research exists surrounding the relationship between intelligence and achievement as measured by individually administered tests of achievement, and nationally developed and normed group achievement tests, no studies have examined this relationship using state-specific standards-based tests to measure achievement. Thus, these findings extend previous work in this area in that they provide additional support for the validity of the WISC-IV by demonstrating its relationship to other variables (i.e., reading and mathematics achievement) using an alternate measure of achievement. However, these results are also novel in that they are the first to suggest the plausibility of using standards-based tests in SLD identification procedures. Recent legislation allows states

to use state-based grade-level standards in their SLD identification procedures. Not only are these tests likely to reflect student performance with regard to state standards (i.e., they are aligned with state curricula), but this study demonstrates that student performance on these tests also correlates with intelligence.

Despite both study hypotheses having been supported, a test of dependent correlations from the same sample revealed the correlation between FSIQ and EOG-Mathematics to be significantly higher than the one between FSIQ and EOG-Reading. A variety of reasons may serve to explain this finding. It is plausible that teachers engaged in a type of curricular narrowing whereby they spent a substantially larger portion of class time preparing students in reading than in mathematics (e.g., Jones et al., 1999, found 401 minutes of reading instruction compared to 292 of math instruction). This phenomenon is likely to result in practice effects. Practice effects reduce the relationship between intelligence and achievement because repeated practice reduces differences in performance related to intelligence (e.g., learning speed, number of trials to criterion). Whatever factors are causing differences in skill acquisition following practice are then less related to intelligence and more related to some alternative variable (e.g., an underlying psychological processing problem). Therefore, increasing instructional time on a subject (here, reading) could account for the lower correlation between intelligence and reading scores.

Although it is possible that teachers focused more of their class time on reading-related activities than on those in other academic areas, a review of the comparative test score means indicates that, with the exception of third grade, students in all grades were lower than their peers on the EOG-Reading than on the EOG-Mathematics. Presumably, higher scores

in reading than in mathematics might be expected if teachers had spent more time teaching reading. Thus, an alternate explanation is that this sample is comprised of students who were referred largely as a result of their reading difficulties, which is plausible given that more children have been served under the disability category of SLD than any other category since IDEA was first enacted and 80% of children with learning problems have difficulties with reading (Joseph, 2002; US Department of Education, 2007). A sample of students with reading disabilities is likely to obscure or attenuate the relationship between intelligence and achievement because of the nature of the disability and its relationship to reading. More specifically, SLD is defined as a disorder in one or more of the basic psychological processes underlying language (e.g., phonological processing, phonemic awareness) that is not associated with general intellectual ability, which may therefore result in difficulties in a variety of academic areas. Thus, this processing deficit disrupts the typical IQ-achievement correlation by acting as a moderator, and is often invoked to explain why a discrepancy between ability and achievement indicates a disability.

An evaluation of the correlations between EOGs and WJ-III achievement scores reveals that the EOG-Reading is more highly correlated with measures of applied reading skills (i.e., Reading Comprehension) than with basic reading skills, whereas the EOG-Mathematics is more highly correlated with measures of basic mathematics skills than with applied mathematics skills. This pattern suggests that the correlation between FSIQ and EOG-Reading may be lower than the correlation between FSIQ and EOG-Mathematics because the EOG-Reading is not sampling from material students are learning at school. That is, to be successful at reading comprehension, which is a higher order skill, one must have

knowledge of basic skills. It is possible that at these younger ages reading instruction targets more rudimentary skills rather than reading comprehension and thus this test would not be measuring that which the students have been taught. However, this hypothesis is unlikely given that restriction of range was not found to exist on any measures including the EOG-Reading.

Another potential reason for the difference in correlations between FSIQ and each of the EOG tests relates to the construct of intelligence as it is measured by the WISC-IV. That is, an examination of the relationship between individual WISC-IV Indexes and EOG scores shows the Verbal Comprehension Index (VCI) to correlate more highly with EOG scores in both academic areas than any of the other three Indexes (i.e., PRI, WMI, PSI). This finding is consistent with the extant literature, which shows the VCI to correlate more highly than other Index scores with individual achievement composites (e.g., Reading, Mathematics, Written Language, Oral Language) as well as with total achievement. Those subtests that comprise the VCI are intended to measure verbal concept formation, reasoning, and comprehension, as well as acquired knowledge and attention to verbal stimuli (Zhu & Weiss, 2005). Thus, the Index aligns with what Cattell and Horn termed crystallized intelligence, or accessible knowledge stores and the ability to acquire additional knowledge via learning strategies, which is thought to represent those abilities influenced by acculturation, schooling, and language development (Wasserman & Tulsky, 2005).

The Perceptual Reasoning Index (PRI), however, was found to correlate more highly with the EOG-Mathematics than with the EOG-Reading. The PRI consists of subtests measuring fluid reasoning, spatial processing, attentiveness to detail, and visual-motor

integration and thus reflects what Cattell and Horn described as fluid ability, which is utilized when adaptation to new situations is required. Cattell argues that fluid ability is the most essential factor in contributing to *g*, or one's general intellectual capacity (Wasserman & Tulsky, 2005; Zhu & Weiss, 2005). Thus, it is possible that the present study found the correlation between FSIQ and EOG-Mathematics to be significantly higher than the one between FSIQ and EOG-Reading because that which is tested on the EOG-Mathematics represents more novel intelligence and therefore requires more fluid ability than does that which is tested on the EOG-Reading.

Because the correlations are influenced by the way in which the WISC-IV measures intelligence, the finding that neither the WMI nor the PSI were found to correlate significantly more highly with one EOG test than the other suggests that the EOG-Reading and –Mathematics share similar amounts of variance with regard to working memory and processing speed. However, the correlations between each of these Indexes and EOG test scores are lower than are those correlations between the VCI and PRI and the EOGs (with the exception of the PRI/EOG-Reading correlation) likely because both the WMI and PSI are less *g*-loaded than are the other Indexes and therefore share less variance with the EOGs.

An examination of the correlations between FSIQ scores and scores yielded from a measure of achievement other than the EOGs, here the third edition of the Woodcock-Johnson Tests of Achievement (WJ-III), indicates a pattern similar to the one found for FSIQ and EOG scores. That is, FSIQs correlate more highly with WJ-III broad mathematics scores (r =.78) than with WJ-III broad reading scores (r = .31). Thus, it is possible that the differences in the correlations reflect a real difference in the relationship between the

constructs (i.e., intelligence correlates more highly with mathematical abilities than with reading abilities). However, in comparing these findings with those presented in the WISC-IV technical manual, an inconsistency is noted. More specifically, the manual reports a correlation of $r = .78$ with both the math and reading composites of the second edition of the Wechsler Individual Achievement Test (WIAT-II). It is important to note that a search of multiple databases (i.e., ERIC, PsycInfo, SSCI) as well as various test company websites (i.e., The Psychological Corporation, Riverside Publishing) did not yield any studies providing information on the relationship between WISC-IV and WJ-III scores or on the relationship between WIAT-II and WJ-III scores. The only other study in the extant literature base to have analyzed the WISC-IV correlates of academic achievement (as measured by the WIAT-II) reveals correlations similar to but still somewhat higher than those obtained in the current study, which is interesting given that the study analyzed correlations for children diagnosed with Attention-Deficit/Hyperactivity Disorder (ADHD; Mayes & Calhoun, 2007). The fact that the results yielded in the present study are more similar to those of the aforementioned study than of those presented in the WISC-IV manual suggests that the pattern of correlations obtained here may better represent that of referred than non-referred samples. However, it is not possible to evaluate whether the differences in the obtained correlations between FSIQ and EOG-Reading and - Mathematics may actually be the result of differences in tests or in samples, given the lack of available studies examining the relationship between WISC-IV and WJ-III scores.

A final reason why the current study may have yielded this pattern of correlations is related to the types of items included on the North Carolina state tests. As noted earlier,

more than 40 states currently contract with CTB/McGraw Hill, Harcourt Brace, or Riverside Publishing, each of whom publish a major national group-based achievement test. Consequently, these states' tests tend to be modeled after the national tests and are therefore likely to be similar to them with regard to format, number of items, psychometric properties, and question types used in validating other intelligence tests. Because North Carolina does not contract with one of these publishers and therefore does not draw any of its test items from a national item pool, it is possible that its tests incorporate different kinds of test items than do other states' assessments, and are therefore less correlated with tests of intelligence than are other national tests of achievement.

Limitations and Future Research

Although the results of the current study offer insight into the nature of the relationship between intelligence and achievement as measured by state achievement tests, there exist a number of issues important to consider for future research. For one, participants of the present study were asked to provide any available achievement data to which they had access. The majority supplied scores from the WJ-III, which is not co-normed with the WISC-IV. Thus, future research should incorporate WIAT-II scores into its design, as the WISC-IV and WIAT-II have a linking sample that facilitates standard scores comparisons.

Another limitation of the present study is the nature of the sample. Although varied enough to detect significant correlations, replication of the present study with more representative subjects would help determine whether the correlations in this study are representative of the actual IQ/EOG relationship, or whether they are influenced by the non-representative nature of the referred sample.

Future research in this area may wish to extend this study to examine the relationship between intelligence and achievement using standards-based tests from other states. Most state tests are modeled after one of three national group-based achievement tests; however, North Carolina's tests are not. Thus, correlations between intelligence and achievement utilizing another state's assessments may be higher than those yielded as part of this study.

Conclusions

In summary, the results yielded from this study provide additional evidence for the validity of the WISC-IV with regard to its relation with other variables (i.e., achievement). Significant correlations between FSIQ and EOG-Reading and EOG-Mathematics scores also suggest the potential for such standards-based tests to be incorporated into SLD identification policies. This study serves as the first to evaluate the relationship between intelligence and achievement using tests of state standards and can thus be used to guide future investigations in this area.

References

Allington, R. L., & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York. *Educational Policy, 6*, 397-414.

American Educational Research Association. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Braden, J. P. (2003). Psychological assessment in school settings. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Vol. 10. Assessment psychology.* (pp. 261-290). New York: John Wiley & Sons, Inc.

Burns, T. G., & O'Leary, S. D. (2004). Wechsler intelligence scale for children-IV: Test review. *Applied Neuropsychology, 11*, 235-238.

Dombrowski, S. C., Kamphaus, R. W., & Reynolds, C. R. (2004). After the demise of the discrepancy: Proposed learning disabilities diagnostic criteria. *Professional Psychology: Research and Practice, 35*, 364-372.

Duncan, D. B. (1955). Multiple range and multiple F tests. *Biometrics, 11*, 1-42.

Elliott, S. N., Braden, J. P., & White, J. L. (2001). *Assessing one and all: Educational accountability for students with disabilities.* Arlington, VA: Council for Exceptional Children.

Firestone, W. A., & Mayrowetz, D. (2000). Rethinking "high stakes": Lessons from the United States and England and Wales. *Teachers College Record, 102*, 724-749.

Gay, G. H. (1990). Standardized tests: Irregularities in administering of tests affect test results. *Journal of Instructional Psychology, 17*, 93-103.

Gresham, F. M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 467-519). Mahwah, NJ: Erlbaum.

Individuals with Disabilities Education Act (IDEA) Amendments of 1997, P. L. 105-117, 20 U.S.C. §1400 *et seq.*

Individuals with Disabilities Education Improvement Act of 2004, P. L. 108-146, 20 U.S.C. §1400 *et seq.*

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*, 199-203.

Joseph, L. M. (2002). Best practices in planning interventions for students with reading problems. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV: Vol. 1.* (pp. 803-816). Maryland: National Association of School Psychologists.

Langenfeld, K. L., Thurlow, M. L., & Scott, D. L. (1996). *High stakes testing for students: Unanswered questions and implications for students with disabilities* (Synthesis Report No. 26), Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved March 7, 2006, from http://education.umn.edu/NCEO/OnlinePubs/Synthesis26.htm

Mayes, S. D. & Calhoun, S. L. (2007). Wechsler Intelligence Scale for Children – Third and – Fourth Edition predictors of academic achievement in children with Attention-Deficit/Hyperactivity Disorder. *School Psychology Quarterly,* 22, 234-249.

McDonnell, L. M., McLaughlin, M. J., & Morison, P. (Eds.). (1997). *Educating one and all: Students with disabilities and standards-based reform.* Washington, DC: National Academy Press.

Mehrens, W. A. (1998, July 14). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives, 6*, Article 13. Retrieved March 7, 2006, from http://epaa.asu.edu/epaa/v6n13.html

Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless or fraudulent? *Educational Measurement: Issues and Practices, 8*, 14-22.

National Academy of Sciences (n.d.). Executive summary. [Executive summary of the book *High stakes: Testing for tracking, promotion, and graduation]*. Retrieved January 4, 2006, from books.nap.edu/execsumm_pdf/6336.pdf

Overton, T. (2006). *Assessing learners with special needs: An applied approach* (5th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Public Schools of North Carolina (2003). *Understanding your child's End-of-Grade test scores: Grades 3, 4, and 5.* Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2004a). *Assessment brief: Understanding the North Carolina End-of-Course Tests.* Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2004b). *North Carolina End-of-Grade Test of Reading Comprehension-Grade 3.* Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2004c). *North Carolina Reading Comprehension Tests* (Technical Report). Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2005a). *ABCs 2005 accountability report and background packet*. Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2005b). *Evolution of the ABCs*. Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2005c). *The North Carolina testing program.* Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2005d). *Understanding the individual score report for the North Carolina End-of-Grade Tests: Grades 6, 7, and 8.* Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2006). *The North Carolina Mathematics Tests* (Technical Report). Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services.

Public Schools of North Carolina (2007). *The North Carolina State Testing Results*. Raleigh, NC: Public Schools of North Carolina. Department of Public Instruction, Division of Accountability Services/Test Development Section.

Reschly, D. J., & Grimes, J. P. (2002). Best practices in intellectual assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV: Vol. 2.* (pp. 1337-1350). Maryland: National Association of School Psychologists.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112-118.

Saklofske, D. H., Prifitera, A., Weiss, L. G., Rolfhus, E., & Zhu, J. (2005). Clinical interpretation of the WISC-IV FSIQ and GAI. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-Practitioner Perspectives* (pp. 33-69). New York: Elsevier Academic Press.

Schulte, A. C., Villwock, D. N., Whichard, S. M., & Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. *School Psychology Review, 30*, 487-506.

Smith, T. C. (2005). IDEA 2004: Another round in the reauthorization process. *Remedial and Special Education, 26*, 314-319.

Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist, 52*, 1030-1037.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly, 47*, 1-41.

United States Department of Education. (2005). *Accountability for students with disabilities: Accountability plan amendments for 2004-2005.* Washington, DC: Author. Retrieved February 2, 2005, from http://www.ed.gov/policy/elsec/guid/raising/disab-acctplan.html

United States Department of Education. (2007). *The condition of education 2007.* Washington, DC: National Center for Education Statistics. Retrieved February 10, 2008, from http://nces.ed.gov/programs/coe/2007/pdf/07_2007.pdf

Wasserman, J. D., & Tulsky, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 3-22). New York: Guilford Press.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children: Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children: Fourth edition*. San Antonio, TX: The Psychological Corporation.

Weiss, L. G. (2003). *Theoretical model and test blueprint* (WISC-IV Technical Report #1). San Antonio, TX: The Psychological Corporation.

Williams. P. E., Weiss, L. G., & Rolfhus, E. (2003). *Theoretical model and test blueprint* (WISC-IV Technical Report #1). San Antonio, TX: The Psychological Corporation.

Zhu, J., & Weiss, L. (2005). The Wechsler scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 297-324). New York: Guilford Press.

Appendices

Appendix A

Conference Recruitment Posting

Are WISC-IV Scores Valid Predictors of NC Test Scores?
Wanna Find Out??

Jeffery P. Braden, PhD
NC State University

The WISC-IV was published in 2003 to upgrade and improve its predecessor, the WISC-III. There are a number of differences in the WISC-IV, not the least of which is the elimination of Verbal and Performance IQs in favor of Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed Indexes. These changes reflect an attempt to align the WISC-IV with contemporary research in cognitive abilities and neuropsychology. The test publisher (The Psychological Corporation) provides substantial data to justify this shift, and to show that WISC-IV composites are related to academic achievement. However, the WISC-IV documentation does not include information linking these new composites to achievement data unique to North Carolina such as the tests used in our ABCs assessment program.

In collaboration with NCSPA and NC State University's Psychoeducational Clinic, I would like to recruit North Carolina psychologists to contribute WISC-IV scores and NC ABCs test results to a common database. We will use this "data cooperative" to explore the degree to which the WISC-IV relates to achievement data specific to North Carolina. We will send participating psychologists a blank Excel spreadsheet already formatted to receive data. Psychologists will insert data, without identifying information, into the spreadsheet and return it to me via email for analysis. Although the scope of such a study is typically beyond any one psychologist or entity, a collaborative effort could produce a large and varied sample of North Carolina students. Note that, because the study will collect anonymous, archival data, it will NOT be necessary to obtain permission from individual students or their parents/guardians for participation.

We held a meeting at NCSPA's annual conference in Wilmington last month to discuss the nature of this study, solicit feedback, and recruit interested participants. We are now looking to invite even more psychologists and professionals to become involved. Let's do our part to study how the WISC-IV relates to North Carolina achievement tests, and in so doing, inform and improve professional practice in our state. I hope you will contact me (jeff_braden@ncsu.edu) for information about this project.

Appendix B

Listserve Recruitment Posting

Dear NCSPA Members,

I am looking for North Carolina school psychologists who could help provide data on the relationship between WISC-IV scores and North Carolina End-of-Grade (EOG) Test scores! (A more formal description of the study can be read below.) Essentially, your participation would require you to provide anonymous, archival data (WISC-IV and EOG scores) on as many students as you could. I estimate that it will take about 10-15 minutes per student to enter all the data. I would only need information on students who have already been tested and who have EOG scores, so no additional testing would be needed!

If you think you might be interested and/or would like more information, please contact me via email at michelle_parker@ncsu.edu or phone at 919-969-8095. I am a student in the NC State School Psychology Program, and my advisor for the project is Dr. Jeff Braden (jeff_braden@ncsu.edu).

Thank you so much in advance for your consideration! I really appreciate it!
Michelle Tayrose

The proposed study, "The Relationship between WISC-IV Scores and North Carolina State Achievement Tests Scores," will seek to estimate the correlation between the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV), and those tests comprising the North Carolina assessment program (i.e., scores from the ABCs [*A*ccountability and High Standards, *B*asics, and Maximum Local *C*ontrol] of Public Education). There are three reasons why the research may be of value: (1) previous research has predominately used individually administered tests of achievement, not group tests, (2) the standards-based reform movement and recent legislation now emphasizes group achievement (or "high-stakes") tests for general and special education, and (3) there is no prior research examining the relationship between intelligence and state-based achievement tests. The current study will address these gaps in the literature, and in so doing, will provide evidence regarding the validity of the WISC-IV, including information that may inform its use in identifying learning disabilities in which achievement is measured by tests of state standards.

The study is particularly important for NC school psychologists because our professional standards require us to understand the validity of the tests we use. Given that there are no data linking WISC-IV scores to scores on any state assessments, it is very important for WISC-IV users to contribute data so the validity of the WISC-IV for NC school children can be better understood.

Appendix C

North Carolina State University
Informed Consent Form for Research

Title of Study The Relationship between WISC-IV Scores and North Carolina State Achievement Test Scores
Principal Investigator Michelle Tayrose Faculty Sponsor Jeff Braden, PhD.

The Relationship between WISC-IV Scores and North Carolina State Achievement Test Scores – Consent Form

"The Relationship between WISC-IV Scores and North Carolina State Achievement Test Scores" is a research study aimed to evaluate the newly updated version of the WISC as it relates to the performance of North Carolina students on state tests, which are a part of the ABCs assessment program. This study will be conducted by Dr. Jeff Braden and an NCSU graduate student, Michelle Tayrose. Knowledge of this topic will provide essential information that will help to improve the research knowledge base on which all North Carolina school psychologists practice. This study invites you to provide anonymous archival data describing the performance of North Carolina students on the WISC-IV and other academic and achievement tests. You are asked to print this consent form, sign below, and indicate a password that will be used to encrypt your data for the duration of this study. You will then mail this consent form to Dr. Jeff Braden Box 7650 North Carolina State University Raleigh, NC 27695-7650. After receiving this form, you will be emailed two blank Excel spreadsheets in which to enter relevant student information. Access to these spreadsheets will require the use of your password. Once complete, you will email your spreadsheets to Michelle Tayrose at michelle_parker@ncsu.edu.

The goal of this study is to recruit enough participants to provide at least 60 pairs of student scores (i.e., WISC-IV/EOG-Reading and WISC-IV/EOG-Mathematics). Thus, it is likely that you will be one of approximately 70-80 psychologists or other professionals participating in this study. It is estimated that you will need about 15 minutes per client to enter all the test and demographic data into the spreadsheets.

This study will not involve any contact with those individuals whose data will be used in this study. Rather all data will be archival, having already been collected for other purposes. Two steps will be taken to ensure data are not used to identify students. First, you will be required to provide demographic data in a separate file from the WISC-IV test data to minimize the ability of an unauthorized individual to associate demographic data with test results. Second, you will encrypt the files when sending them via email to further protect the confidentiality of the data.

Because the study will use extant archival data that will be kept anonymous, the study does not pose harm or risk to participants. There are no negative consequences for declining participation, and you can withdraw at any time without penalty. An electronic copy of the results of the study will be available free of charge to anybody who asks, whether or not they decide to participate.

If you have any questions or concerns that arise in connection with your participation in this study, you may contact the Principal Investigator, Dr. Jeff Braden, at (919) 513-7393, jeff_braden@ncsu.edu or Michelle Tayrose at michelle_parker@ncsu.edu. This study has been reviewed and approved by the NCSU Institutional Review Board (IRB). If you have questions or concerns about your rights or the rights of the North Carolina students indirectly involved in this study, please contact the NCSU IRB office at IRB administrator, Ms. Debra Paxton, at (919) 515-4514 or debra_paxton@ncsu.edu.

I have read the above and have been given the opportunity to discuss it and to ask questions. I agree to participate in this research with the understanding that I may withdraw at any time.

_____	_____	_____
Date	Signature	Password

Appendix D

Steps for Participation

What Have I Gotten Myself Into?
(Completing the Study in 5 *Easy* Steps)

1. Express your interest either by signing up after today's session or by contacting Michelle Tayrose via email at michelle_parker@ncsu.edu or Jeff Braden at jeff_braden@ncsu.edu.

2. After receiving your name and contact information, you will be emailed a consent form. We ask that you print the form, sign where indicated, **and designate a password**, which will be used to secure your data for the duration of this study. You will then "snail" mail this letter back to Dr. Jeff Braden.

3. Once we receive your letter, we will email you two Excel spreadsheets (one for demographic data and one for test data) in which you will enter the specified data. You will be required to supply your password in order to access the spreadsheet. Please note that this spreadsheet will be unique to you, so if you know of someone who would also like to participate in this study, please ask them to contact Michelle Tayrose for a separate spreadsheet.

4. Once you have finished entering your data, you will send the completed Excel spreadsheets to Michelle Tayrose via email.

5. Pat yourself on the back!!! You're all done and we thank you for your participation. You can feel pleased that you've helped to improve the research knowledge base on which all NC School Psychologists practice.

Appendix E

Linear Regression Analyses

Results of a Linear Regression Analysis for Variables Predicting EOG-Reading Scores (N = 76)

Variable	*B*	*SE B*	B
FSIQ	0.04	0.01	0.49**
Grade	-0.06	0.06	-0.09
LEP Status	-0.34	0.42	-0.08

Note. R^2 = .26

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

Results of a Linear Regression Analysis for Variables Predicting EOG-Mathematics Scores (N = 76)

Variable	*B*	*SE B*	B
FSIQ	0.04	0.01	0.65**

Appendix E (continued)

Grade	0.00	0.05	0.00
LEP Status	-0.17	0.32	-0.05

Note. R^2 = .42

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