1995

The comparison of the test of cognitive skills and the wechsler intelligence scale for children-revised for students with learning disabilities.

Jennifer McGiverin
Indiana State University

VITA

Jennifer L. McGiverin
3796 East 73 1/2 Avenue
Terre Haute, IN 47805
(812) 466-6747

EDUCATION

1989-1995	Ph.D. Guidance and Psychological Services, Indiana State University (Expected Graduation Date: August 1995).
1985-1988	M.Ed. School Psychology, Indiana State University.
1976-1983	Graduate studies--27 hours towards M.S. in Counseling Psychology, Frostburg State University, Frostburg, MD.
1972-1976	B.S. Mathematics, Frostburg State University.

EXPERIENCE

1993-1994	Doctoral Intern, Clay Community Schools Special Services.
1992-1993	Intern, School Psychology, Clay Community Schools Special Services.
Spring 1992	School Psychologist, Covered Bridge Special Education District.
1981-1982	Instructor, Allegany Community College, Cumberland, MD.

AWARDS

May 1993	Becker Award. Meritorious award given by Department of Educational and School Psychology, Indiana State University
1992-1993	Paul A. Witty Scholarship. Meritorious award for scholars of gifted education.
1990-1991	Paul A. Witty Scholarship. Meritorious award for scholars of gifted education.

PUBLICATIONS

McGiverin, J., Gilman, D., & Tilletski, C. (1989). A meta-analysis of a two-year, state-wide program of reduced class size. Elementary School Journal, 90, 47-56.

PRESENTATIONS

McGiverin, J., & Walker, K. (1995, April). Group intelligence tests: Can they be used by the school psychologist? Presented at the annual national convention of the National Association of School Psychologists, Chicago, IL.

McGiverin, J. (1992). Mathematically gifted children: Parents and teachers as partners in the identification process. Presented at the annual national convention of the National Association of School Psychologists, Nashville, TN.

McGiverin, J., & Miller, M. (1991, November). Indicators of the mathematically precocious: The IMP Scale. Presented at the annual national convention of the National Association of Gifted Children, Kansas City, MO.

McGiverin, J. (1991, February). Helping teachers to identify the qualitative characteristics of their mathematically precocious students. Presentation to Phi Delta Kappa, Indiana State University Chapter, Terre Haute, IN.

Barratt, W., & McGiverin, J. (1990, November). Consultation: Facilitating the home-school relationship. Paper presented at the annual orientation and education sessions for school counselors at Indiana State University, Terre Haute, IN.

McGiverin, J. (1989, April). Identifying mathematically gifted children and youth. Presented at the annual meeting of the Indiana Association of the Gifted, Indianapolis, IN.

Rasico, J., & McGiverin, J. (1988). The relationship among kindergarten programs and measures of basic concept acquisition and reflectivity/impulsivity. Poster session for the annual national convention of the National Association of School Psychologists, Chicago, IL.

ORGANIZATIONS

American Association of University Women
American Psychological Association
National Association of School Psychologists
National Association of Gifted Children
Phi Kappa Phi
Sigma Zeta (National Honorary Math/Science Fraternity)

THE COMPARISON OF THE TEST OF COGNITIVE SKILLS AND THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN-REVISED FOR STUDENTS WITH LEARNING DISABILITIES

A Dissertation

Presented to

The School of Graduate Studies

Department of Educational and School Psychology

Indiana State University

Terre Haute, Indiana

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Jennifer McGiverin

December 1995

APPROVAL SHEET

The dissertation of Jennifer McGiverin, Contribution to the School of Graduate Studies, Indiana State University, Series III, Number 649, under the title The Comparison of the Test of Cognitive Skills and the Wechsler Intelligence Scale for Children-Revised for Students with Learning Disabilities is approved as partial fulfillment of the requirements for the Doctor of Philosophy Degree.

9/18/95
Date

Kenneth P. Walker
Chairperson

Committee Member

Committee Member

Committee Member

Committee Member

9/21/95
Date

For the School of Graduate Studies

ABSTRACT

This study investigated the criterion-related validity of the Test of Cognitive Skills (TCS) as an estimate of cognitive ability for students with learning disabilities. Research indicates that knowledge of scores from group ability tests significantly influences teachers' instruction and treatment of students. In contrast, school psychologists rarely consider group administered test scores and must devote substantial time to administering and interpreting individual intelligence measures. The present study compared the TCS to the Wechsler Intelligence Scale for Children--Revised (WISC-R) for students with learning disabilities in order to contribute to the valid use of the TCS by both teachers and school psychologists. More specifically, the study examined the validity of using the TCS (a) to screen for students with learning disabilities, and (b) as a substitute for the WISC-R as an estimate of cognitive ability for students with learning disabilities.

Subjects were 118 elementary and middle school students enrolled in learning disability services in a midwestern county school system. Results of *t*-tests indicated that the TCS Cognitive Skills Index (CSI) significantly underestimated the cognitive abilities operationally defined by the WISC-R. Pearson correlations revealed significant positive relationships between the CSI and each of the

WISC-R IQs and factor scores. The standard error in estimating the WISC-R Full Scale IQ (FSIQ) from the CSI was 8.19 standard score points; thus, a confidence range of approximately 33 points was required to estimate the FSIQ from the CSI at a 95% probability of accuracy.

Post-hoc analyses revealed that subjects tended to score lower on the CSI as FSIQ-CSI differences increased. Further, subjects with weaknesses in reading skills obtained significantly lower CSIs than subjects with average reading skills. The FSIQs of these two groups were not significantly different. Moreover, subjects with weaknesses in attention/concentration skills obtained significantly lower CSIs than subjects with average attention/ concentration skills. However, as measured by the WISC-R Verbal Comprehension factor, the cognitive abilities of these two groups were not significantly different.

Based on the results of this study, the TCS CSI cannot be recommended as an estimate of cognitive ability as operationally defined by the WISC-R for students with learning disabilities. Further, using the CSI in screening students for learning disabilities is contraindicated since the CSI is likely to screen out students with learning disabilities. Additional research is required to examine the validity of the CSI as an estimate of cognitive ability for both special education and regular education.

ACKNOWLEDGEMENTS

In completing this dissertation, I wish to express my deep appreciation to several special people. From the conceptualization to the completion of this document, Dr. Kenneth Walker, my committee chair, was a scholarly advisor and an enduring source of encouragement. Dr. Maurice Miller, a member of my committee, contributed vital insights, keen-eyed critiques, and continued support to me as I progressed through this work. I also appreciate the assistance of the other members of my committee: Dr. Gnanaolivu Aaron, Dr. David Gilman, and Dr. Edward Kirby.

To my husband, Rolland, and our children, Brian and Kellie, I owe a great debt for their loving encouragement, understanding, and humor in sharing this project with me. They have promoted my personal growth and continue to provide me an ever-present reminder of the truly meaningful facets of my life.

I dedicate this work to my mother, Patricia Brooks McCray. A petite and astute British woman, my mother survived both WWII and me, and by virtue of her role-modeling taught me to be polite, to accept life with sturdiness, and to embrace cultural diversity.

TABLE OF CONTENTS

Chapter Page

List of Tables

Table Page

Chapter 1

INTRODUCTION

Over 250 million standardized tests of ability, achievement, perceptual and motor skills, vocational interests, and socioemotional functioning are administered in education each year (Bersoff & Hofer, 1990). Group tests that purport to yield estimates of individual ability for achievement in the established academic environment constitute a substantial proportion of these tests. Referred to by a host of terms, such as "Standard Age Score," and "Cognitive Skills Index," the scores from group tests of ability are widely accepted as measures of intelligence, that is, "IQ" (Cunningham, 1986).

Scores gained from group intelligence tests are commonly reported in the cumulative folders for individual students in the public school system. Historically, these scores have been used to group children for instructional purposes; however, a volley of decisions derived from two decades of legal confrontations determined that special education decisions founded on group ability measures were discriminatory, and, therefore, illegal (Bersoff & Hofer, 1990; Cunningham, 1986). Federal mandates, such as the 1990

IDEA (Individuals with Disabilities Education Act) and its predecessors, demand that the evaluation of cognitive abilities of referred students be effected only through the use of standardized instruments that are individually administered exclusively by certified professionals.

Despite legal prohibitions against the use of group administered ability measures for making special education decisions, a majority of classroom teachers continue to give these scores considerable weight in making many decisions from individual instructional planning to special education referral (Dusek & Joseph, 1985; Fields & Kumar, 1982). Moreover, knowledge of ability scores has been found to (a) influence teacher acceptance of behavioral interventions (Martens & Meller, 1989), (b) modify teacher judgments of children based on social class (McCombs & Gay, 1988), and (c) affect teacher perception of students for both emotional and behavioral attributes more potently than student gender (Prawat & Jarvis, 1980). Thus, group administered intelligence tests produce individual student scores that not only provide data utilized by teachers for specific instructional purposes but also modulate teacher judgments regarding a multitude of individual student characteristics.

In contrast to teachers' common practice of using group intelligence scores, school psychologists rarely utilize information from group administered tests of ability in evaluating students (Elliott, Piersel, & Galvin, 1983). Concerns have been expressed regarding the use of group

administered tests even for screening purposes (Hartsough, Elias, & Wheeler, 1983). While disregarding group administered data, school psychologists place a heavy emphasis on information garnered from individually administered tests of ability. Experts have called for a reduction in the use of individually administered tests of intelligence and have charged that such tests yield little information that is relevant to intervention (e.g., Haywood, Brown, & Wingenfield, 1990; Naglieri, Das, & Jarman, 1990; Reschly, & Grimes, 1990; Reschly & Wilson, 1990); however, surveys indicate that school psychologists conduct individual intelligence assessment for virtually every referred student (Elliott, Piersel, & Galvin, 1983; Reschly, Genshaft, & Binder, 1987). This practice results in the monthly administration of an average of seventeen intellectual evaluations, or about one per workday, by each school psychologist (Reschly, Genshaft, & Binder, 1987).

Underlying the divergence between the attitudes of teachers and school psychologists regarding group ability measures is the issue of validity: "The appropriateness of the inferences that can be made on the basis of test results" (Salvia & Ysseldyke, 1991, p. 145). The prioritization of individual intelligence measures is premised, in part, on a body of research that has established their criterion-related validity, both concurrent and predictive (Sattler, 1988). The validity of group measures, however, is largely in doubt due to a dearth

of relevant research that has compared these instruments to other measures of cognitive functioning (Salvia & Ysseldyke, 1988; 1991). Critical to resolving conflicting views and practices involving group ability tests as well as contributing to their valid utility are investigations that examine their relationship to designated criteria.

Concerns of both teachers and school psychologists must be further addressed through research that examines the validity of group administered scores for those students who may be referred for special education evaluation and placement. In particular, children who have reading difficulties, communication problems, and/or attention deficits may produce group administered test scores that underestimate their true cognitive abilities and, therefore, preclude valid utilization of many tests (Anastasi, 1982; Cunningham, 1986; Salvia & Ysseldyke, 1988; 1991).

If group administered tests do underestimate ability, for example, teachers who are concerned with a student's low achievement levels may conclude erroneously that the student's achievement is commensurate with the ability estimate derived from a group test. Given such a conclusion, teacher intervention or referral for evaluation may therefore be obviated.

On the other hand, if valid, group administered tests may not only facilitate teacher insight, but also, by contributing to the evaluation process, enable the school psychologist to devote more time to consultation and other

related services. Thus, children who qualify for learning disability services warrant a targeted examination in the comparison of group and individual instruments of cognitive functioning.

Purpose of the Study

The purpose of the present study was to investigate the criterion-related validity of the group administered Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1981). The study compared the standard scores from the TCS and the individually administered Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) for students with learning disabilities. The following research questions were posed:

Question 1. For students with learning disabilities, are there significant differences between the standard score means of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Verbal Comprehension factor (VC), Perceptual Organization factor (PO), and Freedom from Distractibility factor (FD)?

Question 2. For students with learning disabilities, do significant relationships exist between the standard scores of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Verbal Comprehension factor (VC), Perceptual Organization factor (PO), and Freedom from Distractibility factor (FD)? Given a significant relationship, what is the

range of confidence within which an individual's Full Scale IQ may be estimated from the Cognitive Skills Index?

Hypotheses

The following null hypotheses were tested in this study

1. For students with learning disabilities, there are no significant differences between standard score means of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Verbal Comprehension factor (VC), Perceptual Organization factor (PO), and Freedom from Distractibility factor (FD).

2. For students with learning disabilities, there are no significant relationships between standard scores of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Verbal Comprehension factor (VC), Perceptual Organization factor (PO), and Freedom from Distractibility factor (FD).

Definition of Terms

Cognitive Skills Index

The Cognitive Skills Index (CSI) is the age-based, composite standard score obtained from the group administered Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1981) for students in grades 2 through 12. An estimate of cognitive ability, the CSI has a mean of 100 and standard deviation of 16. The CSI and TCS are described in more detail in Chapter 2.

Eligibility

Eligibility for learning disability special services was determined by the Initial Case Conference Committee. For this determination, the committee must consider the results of: (a) an individually administered, standardized test of cognitive ability; (b) an individually administered assessment of academic achievement; (c) an observation of the student's performance in the general education environment; (d) the relevant medical history and present health conditions; (e) a developmental history; and (f) the written report of the multidisciplinary evaluation team.

In the written report, the multidisciplinary evaluation team states its conclusion regarding whether the child has a specific learning disability and the basis for this decision. The team must provide documentation of a severe ability-achievement discrepancy operationalized by individually administered instruments, and the exclusion of the effects of environmental, cultural, or economic disadvantage as primary causal agents in the student's learning difficulties (Article 7, Indiana State Board of Education, 1992; Rule S-1, Indiana State Board of Education, 1988).

Learning Disability

Defining learning disability, the Indiana State Board of Education (1992; 1988) specifies the following:

A learning disability:

(a) is characterized by severe specific deficits in perceptual, integrative, or expressive processes that severely impair learning efficiency;

(b) includes conditions referred to, or previously referred to as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia;

(c) may be manifested in disorders of listening, thinking, talking, reading, writing, spelling, or arithmetic; and

(d) does not include learning problems due primarily to visual, hearing, or orthopedic impairments; mental or emotional handicaps; or environmental, cultural, or economic disadvantages.

Prior to the 1992 revision of Special Education Rules, Article 7 (Indiana State Board of Education), the definition of learning disability also indicated that eligible students were "chronic failures in the regular classroom" (Rule 7-1, Indiana State Board of Education, 1988, p. 45).

Test of Cognitive Skills

The Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1982) is a group administered test of cognitive ability: "a student's academic aptitude" (p. 1). Composed of four subtests, the TCS yields an age-based composite score, the Cognitive Skills Index (CSI), with a mean of 100 and standard deviation of 16. Five overlapping levels of the instrument assess students in grades 2 through 12. The TCS is further reviewed in Chapter 2.

Wechsler Intelligence Scale for Children--Revised

The Wechsler Intelligence Scale for Children--Revised (WISC-R) (Wechsler, 1974) is an individually administered,

standardized test of cognitive abilities for individuals between the ages of 6 years 0 months and 16 years 11 months. The test, which is further reviewed in Chapter 2, provides three deviation IQ scores as well as three empirically derived, factor scores. Described as follows, all standard scores have means of 100 and standard deviations of 15.

WISC-R IQ Scores. The WISC-R is composed of 12 subtests, with only 10 of the subtests mandatory. Raw score for each subtest are converted to scaled scores through the use of age-based tables separated by 4-month age spans. Scaled scores have means of 10 and standard deviations of 3 points. The Verbal Score and Performance Score are obtained by summing the scaled scores for subtests as indicated:

Verbal Score	Performance Score
Information (I)	Picture Completion (PC)
Similarities (S)	Picture Arrangement (PA)
Arithmetic (A)	Block Design (BD)
Vocabulary (V)	Object Assembly (OA)
Comprehension (C)	Coding (Cd)

The Digit Span (DS) and Mazes (Mz) may be given as supplementary subtests, and may be substituted, according to standardized instructions, for a Verbal or Performance subtest, respectively.

The Full Scale Score is computed by summing the 10 scaled scores for the Verbal and Performance subtests. The Full Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ) are obtained by entering conversion tables according

to computed sums: The three conversion tables apply to all ages (Wechsler, 1974).

WISC-R Factor Scores. Factor analyses of the WISC-R standardization sample yielded three factors: Verbal Comprehension (VC); Perceptual Organization (PO); and Freedom from Distractibility (FD) (Kaufmann, 1975). The factor scores are operationally defined using subtest scaled scores (Sattler, 1988, p. 816):

$$VC = 1.47\ (I + S + V + C) + 41.2$$
$$PO = 1.60\ (PC + PA + BD + OA) + 36.0$$
$$FD = 2.2\ (A + DS + Cd) + 34.0$$

According to research summarized by Sattler (1988), subsequent factor analyses with diverse subject samples, such as, African American, Native American, Mexican American, lower to middle class, mildly mentally handicapped, emotionally disturbed, delinquent, and learning disabled, have consistently supported the three-factor WISC-R structure.

Delimitations of the Study

The present study included the following delimitations which may affect the generalizability of results:

1. The sample is delimited to students enrolled in a single school corporation within the midwest region of the country.

2. The sample is delimited to students who met the eligibility and placement criteria for learning disability defined by the Indiana State Board of Education.

3. The study examined only those cognitive abilities measured by the Wechsler Intelligence Scale for Children--Revised and the Test of Cognitive Skills.

Organization of the Study

Chapter 1 has presented the statement of the need for the study of the problem, research questions, hypotheses, definition of terms, and delimitations of the investigation. Chapter 2 reviews the research literature relevant to the study. Chapter 3 details the research methods and procedures used in the study. Chapter 4 explains the results of statistical analyses conducted to test the research hypotheses. Chapter 5 provides a summary of the study, a discussion of the major findings of the study, and the implications of the study for direct practice and continued research.

Chapter 2

REVIEW OF RELATED RESEARCH

The following literature review provides: (a) provides an overview of the Wechsler Intelligence Scale for Children --Revised (WISC-R) and summarizes the validity research conducted to compare the WISC-R to several other individual measures of cognitive ability; (b) an account of the historical development of group tests of cognitive ability, the associated legal issues and critiques of these group tests, and summarizes several studies comparing the WISC-R to group tests of cognitive ability; and (c) a description, critiques, and validity research for the Test of Cognitive Skills (TCS).

The Wechsler Intelligence Scale for Children--Revised

Overview

The Wechsler Intelligence Scale for Children--Revised (WISC-R) (Wechsler, 1974) is an individually administered test of cognitive ability for individuals ages 6-0 to 16-11. Composed of 12 subtests, the WISC-R yields a global estimate of cognitive functioning, the Full Scale IQ (FSIQ), as well as two separate measures of Verbal IQ (VIQ) and Performance

IQ (PIQ). Factor analyses, regardless of statistical technique, age and ethnicity of subjects, or sample classification, have consistently produced a WISC-R structure composed of three factors: Verbal Comprehension (VC), Perceptual Organization (PO), and Freedom from Distractibility (FD) (Kaufman, Harrison, & Ittenbach, 1990; Sattler, 1988). Scores are reported as deviation IQs with means of 100 and standard deviations of 15. Subtests have means of 10 and standard deviations of 3.

A revision of the 1949 Wechsler Intelligence Scale for Children (WISC) (Wechsler, 1949), the WISC-R was standardized on a sample of 2,200 children enrolled in U.S. schools. The selection criteria and procedures were designed to yield a representative sample relative to 1970 census data; for example, racial/ethnic groups included African Americans, Native Americans, Asians, Puerto Ricans, and Mexican Americans. In his review, Sattler (1988) included these standardization measures among the WISC-R's assets. He also placed positive emphasis upon the WISC-R's: (a) high reliabilities (an internal consistency reliability average of .96, and standard errors less than 5 points for Full Scale, Verbal, and Performance IQs); (b) good administration procedures; (c) good quality test manual and materials; and (d) easy-to-follow scoring criteria.

Literature reviews indicate that the WISC-R has been the topic of a substantial number of empirical investigations. Indeed, Reynolds and Kaufman (1990)

reported that, since 1948, more than 1,100 studies of the WISC or WISC-R have been published with the majority of these articles focused upon the WISC-R. Results of a 1990 computer search produced over 500 articles on the WISC-R published since 1977 (Kaufman et al.). A summary of criterion-related validity research comparing the WISC-R to several other individual measures of cognitive ability follows.

Validity Research with the WISC-R and Individual Measures of Cognitive Ability

The WISC-R and the Woodcock-Johnson Psycho-Educational Battery. The Woodcock-Johnson Psycho-Educational Battery (W-J) (Woodcock & Johnson, 1977) is an individually administered set of 27 tests which assess cognitive ability, academic achievement, and interests for ages 3 years to adult. Part I of the battery (the Tests of Cognitive Ability, WJTCA) consists of 12 subtests including vocabulary, memory, quantitative concepts, concept formation, and spatial relations. Collectively, these 12 subtests yield a Broad Cognitive Ability (BCA) cluster score with a mean standard score of 100 and standard deviation of 15.

Several studies have attempted to investigate the concurrent and construct validity of the BCA as an estimate of intellectual ability by using the WISC-R FSIQ as a criterion measure. Results have tended to indicate a significant relationship between WISC-R and WJTCA for both

referred and nonreferred school-aged children; however, comparisons of the differences between WISC-R and WJTCA means for these subject groups have yielded somewhat varied outcomes.

Coleman and Harmer (1985) administered the WJTCA and the WISC-R to 54 children in grades 1 through 4 who, although referred for evaluation, failed to meet eligibility criteria as determined by multidisciplinary school teams. Both tests were administered 2 weeks apart. Data analyses indicated a significant .77 correlation between the WISC-R and WJTCA full scale scores. The mean difference of 5.65 standard score points between the WISC-R and WJTCA was also significant, with the WISC-R yielding the higher mean of 100.26. Results further indicated no significant effects for grade level (in comparisons of mean scores).

Estabrook (1984) conducted an investigation of the degree of overlap between the WISC-R and the WJTCA for a sample of 152 children referred by special education teams for suspected learning disabilities. The obtained .77 correlation coefficient between the full scale scores of the WISC-R and WJTCA was significant. The mean difference of 5.89 standard points, with a higher WISC-R, was not tested for statistical significance.

Thompson and Brassard (1984) compared the WISC-R and the WJTCA for a group of 60 third, fourth, and fifth grade students who were divided into three samples of 20 subjects each: a normal sample, a mild-moderate LD sample, and a

severe LD sample. Subject selection required an average range of estimated intellectual functioning; LD severity levels were defined by percent discrepancies between ability and achievement (30% and 45% for mild-moderate and severe LD, respectively). Tests were administered within a 2-month period with the exception of data accessed for 9 LD subjects whose scores were less than 1.5 years old. Significant correlations between the WISC-R and WJTCA full scale scores were found for all three samples: normal (r = .86), mild-moderate (r = .74), and severe (r = .93). Mean comparisons revealed a significant 9.5 point difference for the mild-moderate group and a significant 10.5 difference for the severe LD group, with both differences higher for the WISC-R. The difference between means for the normal sample was not significant. Thompson and Brassard thus suggested a linear trend of increasing WISC-R and WJTCA mean differences as a function of LD severity.

A significant relationship between the WISC-R and WJTCA composite scores was also reported by Bracken, Prasse, and Breen (1984) in their study of 136 elementary children. Significant WISC-R and WJTCA correlations for the 99 LD students and 37 normal students were .63 and .75, respectively. Consistent with previous studies cited, the WISC-R FSIQ mean was significantly higher than the WJTCA BCA mean for both LD and regular placement subject groups (by 9.1 and 6.2 standard points, respectively).

McGrew (1983) also found a significant .75 correlation between the WISC-R and WJTCA full scale scores for 52 children in grades 1 through 6 who had been referred for evaluation due to academic and socioemotional/behavioral concerns. Comparisons of full scale means revealed a nonsignificant difference of 0.4 standard score points. Further data analyses indicated that the differences between full scale means remained nonsignificant for those sample subjects who were identified as LD by child study teams. Thus, while providing support for the shared variance between cognitive ability estimates of the WISC-R and WJTCA, these results also stand in contrast to studies that indicate significant mean WISC-R and WJTCA full scale differences for both referred and normal subjects.

The WISC-R and the Kaufman Assessment Battery for Children. The Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983) is an individually administered test of cognitive ability and academic achievement for children ages 2-6 to 12-5 years. The Mental Processing Composite (MPC), an index of mental ability, is derived from the subtests intended to measure Sequential and Simultaneous processing styles. The MPC and has a mean of 100 and standard deviation of 15. Both the concurrent and construct validity of the K-ABC have been examined by studies that have compared the WISC-R FSIQ and the K-ABC MPC.

Zins and Barnett (1984) reported a significant .79 correlation between the WISC-R FSIQ and the K-ABC MPC for 40 volunteer, regular education children ages 6 through 12 years. Comparison of standard score means indicated a significant FSIQ-MPC difference with the FSIQ exceeding the MPC by 5.08 points.

A significant correlation of .80 between the WISC-R FSIQ and the K-ABC MPC was also found for 32 subjects selected from kindergarten through grade 5 classified as learning disabled (n =19) or educable mentally retarded (n = 13) (Obrzut, Obrzut, & Shaw, 1984). Subjects were separated by classification for the comparison of means. Results revealed a relatively small FSIQ-MPC difference (2.48) for the learning disabled group and a somewhat larger FSIQ-MPC difference (5.87) for the educable mentally retarded group. Differences were not tested for significance.

Naglieri (1985) found nonsignificant differences of less than 3 points between the WISC-R FSIQ and the K-ABC MPC after administering both tests, 1 week apart, to 3 groups of children ages 8-1 to 12-5: normal/nonreferred (n = 34), learning disabled (n = 34), and borderline mentally retarded (n = 33). The FSIQ-MPC correlation for the combined sample (r = .85) was significant.

Differences in WISC-R and K-ABC performances for children with serious emotional disturbance (SED) were examined in a study conducted by Pommer (1986). Subjects were 59 students, ages 6 to 11 years, who had been

previously classified as SED by multidisciplinary teams according to state and federal standards. All students attended private school programs for SED students. Tests were administered within a 4-month interval. A significant correlation of .58 was obtained between the FSIQ and MPC. The FSIQ mean was 9.48 greater than the MPC; the difference between means was not tested for significance.

Naglieri and Haddad (1984) sought to determine how the K-ABC and WISC-R compare in estimating the cognitive abilities for a sample of 33 children from grades kindergarten through 6 who had met state guidelines for LD services. The correlation between the FSIQ and MPC was a significant .78. The 5-point difference between the FSIQ and MPC means, with a higher FSIQ, was also significant.

Naglieri (1984) selected a sample of Native American children to examine the validity of the K-ABC using the WISC-R as a criterion. Both instruments were administered 1 week apart to 35 nonreferred Navajo children who resided on a reservation in the southwest United States. In contrast to several cited studies, the K-ABC MPC for this sample was significantly higher than the WISC-R FSIQ by 8.5 points. The .57 correlation between the MPC and FSIQ was also significant.

Forty-one, second through fourth grade students were tested by Klanderman, Devine, and Mollner (1985) for an investigation of the construct validity of the K-ABC using the WISC-R as a criterion. Administered within a 4-month

period, the K-ABC and WISC-R revealed a significant correlation of .56 between the MPC and FSIQ. The FSIQ was found to be higher than the MPC; the 3.58-point difference was not tested for significance.

Finally, McCallum, Karnes, and Edwards (1984) explored WISC-R and K-ABC differences for a group of gifted children. Forty-one children from grades 4 though 6 who had met state criteria for gifted placement, including an IQ of 130 or above on the WISC-R or Stanford-Binet, were selected and administered the K-ABC. WISC-R scores had been obtained within a three-year period. The .58 correlation between the FSIQ and MPC was not tested for significance. The FSIQ mean (132.37) was significantly higher than the MPC mean (119.24).

The WISC-R and the McCarthy Scales of Children's Abilities. The individually administered McCarthy Scales of Children's Abilities (McCarthy, 1972) assesses cognitive abilities for children ages 2-6 to 8-6. The General Cognitive Index (GCI) provides an index of intellectual functioning and is derived from a composite of the 15 subtests of the Verbal Scale, Perceptual-Performance Scale, and Quantitative Scale. The GCI has a mean of 100 and a standard deviation of 16. Concurrent validity studies of the GCI have yielded significant correlations with the WISC-R FSIQ.

Arinold (1982) administered the McCarthy and WISC-R to 20 children 7 to 8.5 years old. All children were in

regular classroom placement; half of the children were African American, and the other half were Caucasian. The correlation between the GCI and the FSIQ standard scores (r = .82) was significant. The FSIQ mean exceeded the GCI mean by 6.5 standard scale points; the FSIQ-GCI difference was not tested for significance.

Comparisons of FSIQ and GCI means were also completed by Reilly, Drudge, Rosen, Loew, and Fisher (1985). Twenty-six, first grade children in regular education classes were administered the WISC-R and McCarthy Scales as a component of a validity study. A significant .89 correlation between the GCI and FSIQ was reported. The difference between the GCI and FSIQ means was not significant.

Hynd, Quackenbush, Kramer, Conner, and Weed (1980) examined the GCI-FSIQ relationship for a group of 44 nonreferred, primary-grade Native American Navajo children. The GCI and FSIQ means did not differ significantly, and the .85 correlation between the means was significant. Researchers, however, cautioned against the use of either measure as an estimate of intelligence for Native American primary-grade children since their GCI and FSIQ means were significantly below normative means.

The WISC-R and the Peabody Picture Vocabulary Test--Revised. The Peabody Picture Vocabulary Test--Revised (PPVT-R) (Dunn & Dunn, 1981) is a nonverbal, multiple-choice instrument that assesses receptive vocabulary of individuals

ages 2-6 through adult. The single raw score is converted to a standard score with a mean of 100 and standard deviation of 15. In addition to vocabulary development, the PPVT-R score has also been interpreted as an estimate of verbal intelligence. Numerous investigations have been conducted to examine the validity of this interpretation using the WISC-R as a validity criterion for a variety of subject samples. Correlations between the PPVT-R and WISC-R Full Scale IQ have been consistently significant for diverse subject samples. Widespread discrepancies between the means, however, have prompted experts to challenge the validity of the PPVT-R score as an estimate of cognitive ability and to argue strongly against interpreting the standard score as an IQ: "It should not be used as a screening device for measuring intellectual level of functioning" (Sattler, 1988, p. 350).

For their PPVT-R and WISC-R comparison, Altpeter and Handal (1986) selected 208 children ages 6-0 through 11-11 referred for psychological evaluation due to various parent and physician concerns, such as developmental delay, academic difficulty, and medical symptoms. Administration of both instruments was completed in a single session. Statistical analyses revealed a significant correlation of .78 between the WISC-R FSIQ and the PPVT-R standard scores. The PPVT-R mean score underestimated the FSIQ mean by 3.68 points, a statistically significant difference for the total sample.

Further examination of differences between the FSIQ and PPVT-R means for classified subsamples indicated that the FSIQ mean was 14.7 points higher for the mentally deficient (PPVT-R ≤ 69) range, 3.8 to 3.0 points higher for low average and average ranges, and 5.6 to 9.14 points lower for the high average (110 ≤ PPVT-R ≤ 119) and superior (PPVT-R ≥ 120) ranges. Differences were not tested for significance.

Breen and Siewert (1983) administered the PPVT-R and WISC-R to 59 students from 6 to 15 years old who had been referred by multidisciplinary teams for assessment due to learning difficulties. Instruments constituted a component of a larger psychoeducational assessment. On the basis of this assessment, 30 students were found to meet state eligibility criteria for learning disability services, while the remaining 20 students did not meet criteria for special education placement. Correlations between the PPVT-R and FSIQ means were significant for both LD (r = .64) and regular placement (r = .77) groups with no significant difference between the correlations coefficients. Furthermore, PPVT-R means were significantly lower than FSIQ means for both LD and regular placement groups by 8.3 and 10.6 points, respectively.

In an earlier study, Breen (1981) selected a referred group of students from 6 to 15 years old as a subject sample for a comparison of the WISC-R and PPVT-R. Following referral and evaluation, mean FSIQ and PPVT-R scores for 32 students (LD = 17; ED = 4; Normal = 11) were compared. In

concert with subsequent investigations Breen found that the FSIQ mean exceeded the PPVT-R mean by a significant 9.81 points, and that the relationship between the FSIQ and PPVT-R (r = .71) was significant.

A significant relationship between the WISC-R FSIQ and the PPVT-R was also reported by Hollinger and Sarvis (1984) for a sample of 51 children. Drawn from elementary and middle schools (average age and grade not reported), these children had been referred for assessment on the basis of "characteristically diverse reasons" (p. 35) and were administered both instruments as part of a comprehensive evaluation. No indication of subsequent classification was reported. Standard score comparisons revealed a significant .82 correlation. Although researchers indicated that the PPVT-R overestimated the FSIQ for 21 of the 51 children and underestimated the FSIQ for the remaining 30, means were neither reported nor analyzed statistically.

In a related study, the WISC-R and PPVT-R were administered within a 1-week interval to 101 newly referred or reevaluated special education students ranging in age from 6-7 to 16-11 (Worthing, Phye, & Nunn, 1984). The reevaluated subsample was composed of 27 students with learning disabilities, 12 with mental disabilities, and 7 with emotional disturbances. Significant correlations were obtained (r = .67), and the .75-point difference between FSIQ and PPVT-R means was not significant.

A statistically significant relationship (r = .86) between the WISC-R FSIQ and PPVT-R was also obtained for a sample of incarcerated adolescents ranging in age from 13 to 16 years (Rosso, Falasco, & Koller, 1984). The FSIQ mean exceeded the PPVT-R mean by 7.66 points, a statistically significant difference.

Davis and Kramer (1985) investigated the relationship between the PPVT-T and the WISC-R for a sample of 40 nonreferred, second grade children. A significant .52 correlation between the PPVT-R and FSIQ was reported. PPVT-R mean was significantly lower than the FSIQ mean by 8.9 points.

The WISC-R and the Stanford-Binet Intelligence Scale: Fourth Edition. The Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE) (Thorndike, Hagen, & Sattler, 1986a) is an individually administered instrument composed of 15 subtests that provides an estimate of cognitive functioning for ages 2 through 23 years. The global SB:FE score, the Composite, has a mean of 100 and a standard deviation of 16. In examining the validity of this relatively new evaluation device, several studies have compared SB:FE to WISC-R performances for various subject samples.

In studies conducted by the SB:FE authors, 205 nonreferred subjects were administered the WISC-R and SB:FE (Thorndike, Hagen, & Sattler, 1986b). The FSIQ-Composite standard score correlation was .83, and the FSIQ mean

exceeded the Composite mean by 2.8 points. Neither result was tested for significance.

Rothlisberg (1987) also compared WISC-R and SB:FE performances for nonreferred children. He reported a significant relationship (r = .77) between the SB:FE Composite and the WISC-R FSIQ standard scores for 32 first and second grade children. The FSIQ mean was significantly higher than the Composite by 7.0 standard score points.

In additional to nonreferred students, students from special populations have been selected as subjects for studies that have examined the Composite-FSIQ relationship (Thorndike et al., 1986b). Gifted children (n = 19) produced scores that correlated 69 and means that differed by 1.7 points with a higher SB:FE. Standard scores for 90 children with learning disabilities yielded a correlation of .87 and a difference between means of 3.0 points, with the WISC-R exceeding the SB:FE. Sixty-one mentally retarded students were also administered both instruments. A .66 correlation between standard scores and a 3.2-point difference between standard score means (with a higher SB:FE) were found. Significance levels were not reported for these studies of exceptional children.

Group Intelligence Tests

Historical Development of Group Intelligence Tests

The history of group tests of intelligence can be traced to the entry of the United States into World War I in

1917. A collection of the country's leading psychologists, headed by APA president Robert Yerkes, concluded that Psychology's support for the war effort should be derived from the nascent field of psychological testing. This commitment to developing the means by which military recruits could be sorted, trained, and tracked was subsequently sanctioned by the United States Department of War (Cunningham, 1986).

Concluding that individual assessment was a pragmatic impossibility, the assembled committee determined that mass assessment through efficient means was necessary. Available instruments, however, required individual administration and were not readily adaptable for group administration or interpretation.

The solution to this dilemma was to come from a graduate student, Arthur Otis, then assigned as graduate assistant to Lewis Terman. Otis, as his dissertation research, had devised a paper and pencil form of the Stanford-Binet that required single responses to multiple choice items (Robertson, 1972). Given its established reliability, the committee judged this instrument to be their sine qua non--it was cheap, quick, and easily administered and scored for large groups.

With Otis accepted into the fold, the committee constructed the Army Alpha and Army Beta for literates and non-literates, respectively. These instruments rapidly gained acceptance within the military ranks as well as the

civilian population (French & Hale, 1990). Indeed, following the armistice, the country experienced an explosion of newly developed group tests of cognitive skills, including the Otis Group Intelligence Scale, the Otis Self-Administering Tests of Mental Ability, the Lorge-Thorndike and the California Short-Form Test of Mental Maturity (Cunningham, 1986). These tests, in concert with similar burgeoning instruments, eventually found their way into a broad spectrum of organizations, including the public school system.

Legal Regulation of Group Intelligence Tests

While mental tests were slowly gaining popularity over the years, civil rights activists and educators during the mid-century began to challenge the use of group intelligence tests for the purpose of tracking students into predetermined programs. Charging that these tests resulted in defacto discrimination, thereby defying desegregation mandates (e.g., Brown v. Board of Education, 1954), a spate of court cases questioned the validity and legal use of group tests of ability.

In perhaps the most potent of these suits, Hobson v. Hansen (1967), a federal district court condemned both the ability grouping system promulgated by the District of Columbia schools and the standardized group ability instruments that provided the data for such grouping decisions. A primary criticism of these tests derived from the court's conclusion that the tests were culturally

biased: "Because tests are standardized primarily on and are relevant to a white middle class group of students, they produce inaccurate and misleading test scores when given to lower class and Negro students" (p. 514). Indeed, according to evidence provided by the D.C. school system itself, 820 of 1272 students assigned to EMR tracks were found to be misclassified when they were reevaluated with an individually administered instrument (Bersoff & Hofer, 1990; Sandoval & Irvin, 1990). Thus, on the basis of the germinal Hobson v. Hanson ruling and subsequent legislation (e.g., IDEA, 1990), the identification and classification of students for special education services with the use of group intelligence instruments continues to be a prohibited practice.

Critiques of Group Intelligence Tests

Assessment experts have weighed the assets and limitations of group intelligence measures. Anastasi (1982), for example, has asserted that the advantages of group instruments include their ease of administration to large numbers of students, cost effectiveness, minimal examiner training requirements, and ease of scoring through clerical or machine methods. These administrative advantages, according to Anastasi, also combine to produce larger subject groups for standardization processes, thereby yielding better established norms than individual tests.

Group tests, however, share several common limitations for both administration and interpretation. In contrast to

individual administration, group administration precludes the establishment of such attributes as rapport, cooperation, interest, and persistence, as well as the identification of interfering circumstances such as fatigue, illness, and anxiety. Further, the single, multiple-choice mode employed by group tests, while providing scoring ease, requires only the recognition of a correct answer while obviating the opportunities for alternate response modes and elaboration of answers that are provided by individual intelligence tests (Anastasi, 1982; Cunningham, 1986; Salvia & Ysseldyke, 1988; 1991).

Prominent in critiques of group tests are concerns for test-taking and other requisite skills that are presumed to affect individual scores. Children, for example, who have reading difficulties, language problems, and/or attention deficits may produce scores that underestimate their cognitive abilities (Anastasi, 1982; Cunningham, 1986; Salvia & Ysseldyke, 1988; 1991): "One must be cautious in interpreting the results of intellectual-ability tests, particularly those of group tests administered to learning-disabled children, for they require reading competence and speed of response which constitutes handicapping barriers" (Robinson & Janos, 1987, p. 31).

Validity Research with the WISC-R and Group Tests of Cognitive Ability

The WISC-R and the Developing Cognitive Abilities Test.

The Developing Cognitive Abilities Test (DCAT) is a

component of the group administered Comprehensive Assessment Program (Wick, Smith, Beggs, & Mouw, 1980). Appropriate for grades 1 through 12, the DCAT assesses Verbal, Quantitative, Spatial, Applications, Critical Thinking, and Basic Cognitive abilities. Scores for the six areas as well as a Total score are reported as percentiles and percent correct.

Investigating the relationship between the WISC-R and DCAT, Karnes and Lee (1984) selected a sample of 77 subjects who were enrolled in grades 5 through 8 of a university residential program for the gifted. Criteria for the program were: (a) a WISC-R or Stanford-Binet score of at least 120, or (b) placement by a regional screening team. Utilizing percentile scores for both instruments, Pearson correlations indicated nonsignificant relationships between the WISC-R IQs and the DCAT Verbal and Quantitative areas. Significant correlational relationships were obtained for the WISC-R PIQ and DCAT Spatial Ability area (r = .22) and between the WISC-R FSIQ and Spatial Ability area (r = .32). Based on statistical results, researchers suggested that the WISC-R PIQ and DCAT Spatial Ability area measure similar constructs. With regard to the comparability of the two tests for gifted students, Karnes and Lee concluded, "clearly one measure could not be substituted for the other" (p. 374).

The WISC-R and the Otis Quick-Scoring Mental Ability Tests, New Edition. The use of group ability tests for both the screening and identification of children for gifted

programs was also examined by Hunter and Lowe (1980) in a comparison of the WISC-R and the Otis Quick-Scoring Mental Ability Tests, New Edition (Otis, 1954). The group measure, a revision of the Otis Self-Administering Tests of Mental Ability (Otis & Barrows, 1929), assesses the cognitive abilities for Grades 1.5 through college senior with three separate tests: Alpha Test for Grades 1.5-4.0; Beta Test for Grades 4-9; and Gamma Test for Grades 9-16. Total scores for each test are reported as IQs (M = 100, SD= 16).

Subjects for the Hunter and Lowe study were 34 fourth and fifth grade students ranging in age from 9 years, 1 month to 11 years, 3 months. Prior to subject selection, each student had achieved a standard score of at least 118 on the grade-level Otis Quick Scoring Mental Ability Test (M = 125.15) as well as a composite score of at least 1.5 years above grade level on the Iowa Test of Basic Skills (ITBS) (Riverside Publishing Co., 1965), a group achievement test.

The WISC-R administration yielded a mean FSIQ of 126.24.Given the initial identification criteria from the Otis and the ITBS, 76.47% of the subject sample obtained a FSIQ of 120 or above. The .25 correlation between the WISC-R and Otis was not significant.

Hunter and Lowe also conducted two multiple regression analyses to predict the WISC-R FSIQ. The first analysis utilized the composite scores of the Otis and the ITBS. In addition to these two group tests, the second analysis included the four scales of the Scale for Rating Behavioral

Characteristics of Superior Students (SRBCSS) (Renzulli & Hartman, 1971). Although the Otis and ITBS were able to account for about 12% (R = .34) of the WISC-R variance, the addition of the SRBCSS to the multiple regression equation increased the correlation (R = .65), thereby accounting for approximately 42% of the WISC-R variance.

Hunter and Lowe thus concluded, "that none of the group measures investigated are effective predictors of the WISC-R FSIQ, regardless of whether they are employed individually or in groups" (p. 64). They further stated that the use of the instruments for screening purposes would result in not only the misidentification of children as gifted but also the underidentification of many children for gifted programs. With regard to the financial concerns associated with individually administered measures of cognitive ability, researchers argued:

> There is little economy in administering large numbers of group tests in comparison to the cost of administering an individual test such as the WISC-R . . . the use of a short form of the WISC-R as a screening instrument might prove the most economical procedure in assuring that gifted children are identified. (p. 64)

As an alternative to group tests, and based on a cited study that yielded a .906 correlation between the FSIQ and the combined WISC-R Vocabulary and Block Design subtests, Hunter and Lowe advocated the use of such short forms as a more valid method for identifying gifted students.

The WISC-R and the Otis-Lennon Mental Ability Test. Covin (1976) explored the comparability of the Otis-Lennon Mental Ability Test (OLMAT) (Otis & Lennon, 1969) and the

WISC-R. A revision of the Otis Quick-Scoring Mental Ability Tests (Otis, 1954), six OLMAT levels measure the cognitive abilities of students enrolled in grades kindergarten through 12. All OLMAT levels provide a composite score, the Deviation Intelligence Quotient (DIQ), with a mean of 100 and standard deviation of 16.

The subjects of Covin's study, 119 first through third graders were administered the OLMAT Elementary I level (OLMAT I for grades 1.6-3.9). Prior to selection, all subjects had been referred for individual evaluation by their teachers who suspected that the children were mildly mentally handicapped. Composed of 53 African American and 66 Caucasian students, the sample's test results were also examined for racial differences in the relationships between the two measures.

Pearson correlations revealed significant relationships between the OLMAT I DIQ and each of the WISC-R IQs for the total group and the Caucasian children. For the African American children the correlations between the OLMAT I DIQ and the WISC-R FSIQ and VIQ were also significant, whereas the relationship between the OLMAT I DIQ and the WISC-R PIQ was not significant.

Although means were not reported, the differences between means for the OLMAT I DIQ and the WISC-R IQs were not significant for the Caucasian sample. Nonsignificant differences between means for the OLMAT I DIQ and the WISC-R FSIQ and Verbal IQ were also revealed for the African

American group. For the total sample, the differences between means for the OLMAT I DIQ and the WISC-R VIQ and PIQ were not significant. Significant differences were indicated for the total sample's OLMAT I DIQ and the three WISC-R IQs and for the African American group's OLMAT I DIQ and WISC-R PIQ.

With regard to comparability of measures for individual students, Coven indicated that the OLMAT I correctly placed the IQs of all children within the MiMH range or within a range of 55 to 79 standard points. Nevertheless, Coven stressed that a prediction range of ± 5.5 points would be required for a two-thirds level of confidence in estimating the WISC-R from the OLMAT I and noted that this range "may be a larger error than many examiners wish to accept" (p. 405). While cautioning that his results did not support the substitution of the WISC-R by the OLMAT I for placement considerations, Coven did support the use of the group instrument for screening purposes provided that further examination was conducted with the WISC-R.

Scores from the WISC-R and the Otis-Lennon Mental Ability Test, Elementary II Level (OLMAT II) (Otis & Lennon, 1969) were examined in a validity study conducted by Rasbury, Falgout, and Perry (1978). Subjects were 70 Caucasian students ages 7 to 8.25 years. In addition to comparing OLMAT II DIQ to the WISC-R IQs, the group ability results were also compared to WISC-R scores derived from a short-form (Yudin-type procedure) scoring format.

The following means (and ranges) for the WISC-R results were reported:

	WISC-R Long Form	WISC-R Short Form
FSIQ	121.02 (83-142)	124.60 (82-146)
VIQ	115.55 (72-143)	119.90 (73-145)
PIQ	122.27 (93-155)	124.15 (95-152)

The OLMAT II DIQ mean was 117.40 with a range of 78 to 150 standard score points. Differences between means for the WISC-R forms were not significant. Differences between WISC-R means and OLMAT II means were not tested for significance.

Pearson product-moment correlations were above .90 between corresponding FSIQs, VIQs, and PIQs of the WISC-R short and long forms. DIQ correlations with the WISC-R long form FSIQ, VIQ, and PIQ were .72, .64, and .50, respectively; DIQ correlations with the WISC-R short form IQs were .68, .58, and .53, respectively. Differences between corresponding correlations for DIQ and WISC-R IQs of short and long forms were not significant. (OLMAT II and WISC-R correlations were not tested for significance.) Based on their review of the group-individual correlation results, researchers concluded: "It is clear that the WISC-R and Otis-Lennon are not interchangeable measures of intellectual functioning" (p. 124).

The WISC-R and the Otis-Lennon School Ability Test. Avant and O'Neal (1986) examined the accuracy of prediction of the WISC-R Full Scale IQ from the Otis-Lennon School Ability Test (OLSAT) (Otis & Lennon, 1982) for a group of

referred children. A revision of the Otis-Lennon Mental Ability Test (Otis & Lennon, 1969), the OLSAT is a group administered test of ability (grades 1-12) that provides a composite School Ability Index (SAI) with a mean of 100 and standard deviation of 16. Based on pragmatic concerns, such as the consumption of financial and personnel resources, as well as the provision of enhanced student services, the primary purpose of the Avant and O'Neal study was to contribute to valid judgments regarding the use of the OLSAT as a efficient, prereferral screening device.

Avant and O'Neal selected a sample of 431 students, ages 6 to 16 years, who had been previously referred for individual evaluation. Case conference committees determined the following placement categories for these students: gifted = 23, learning disabled = 158, mildly mentally handicapped = 38, and no placement = 212. Pearson product-moment correlations for the WISC-R FSIQ and OLSAT SAI were computed for the total sample as well as for selected subgroups. Significant relationships between the two measures were obtained for the total group (r = .82) as well as for the learning disabled (r = .57), mildly mentally handicapped (r = .42), and no placement sample (r = .81). The relationship between measures for the gifted students (r = .10) was not significant. Mean WISC-R FSIQ and OLSAT SAI are summarized below. (Differences between means were not tested for significance.)

Prediction equations were developed utilizing a statistical method (Potthoff technique) which was implemented to detect possible biases in estimation.

Group	WISC-R FSIQ	OLSAT SAI
Total	91.40	90.02
Gifted	127.43	126.52
LD	92.92	88.66
MiMH	66.18	64.82
No Placement	90.87	91.58

(Standard errors of estimate were not reported.) Although biases were not found by placement groups, statistical analyses did reveal differential predictions for African American and Caucasian students for the total, LD, and no placement groups. Based on these results, researchers cautioned that school systems using the OLSAT SAI to predict the WISC-R FSIQ should be aware of this racial difference; they stated no conclusions or recommendations, however, regarding the prediction potential for the total sample or special education groups.

The WISC-R and the Educational Ability Series. The Educational Ability Series (EAS) (Thurstone, 1978) is a group administered test of cognitive ability that yields a composite standard score with a mean of 100 and a standard deviation of 15 for children in grades 1 through 12. Noting that few studies have contributed information to assist school psychologists and other school personnel in the valid utilization of group ability tests, Wright and Piersel (1987) sought to examine the comparability of the EAS and the WISC-R.

For the study, a sample of 151 nonreferred, volunteer children enrolled in third through sixth grade (ages 9-12 years) was selected. In addition to the WISC-R, students were also administered the Wide Range Achievement Test (WRAT) (Jastak & Jastak, 1978), and scores of the Science Research Associates (SRA) Achievement Series (Nasland, Thorpe, & Lefever, 1978) and teacher-assigned grades were collected for each student.

The relationship between the WISC-R and EAS was statistically analyzed through Pearson product-moment correlations. Coefficients were significant beyond the .001 level for the EAS and WISC-R Full Scale IQ (r = .73), Verbal IQ (r = .71), and Performance IQ (r = .57). Regression analysis for predicting the WISC-R Full Scale IQ from a given EAS yielded standard error of estimate of 10.2 standard score points.

Discrepancies between the WISC-R Full Scale IQ and the EAS were also examined and revealed a range of score differences from -26 to +22 full scale points with approximately one-third (39%) of the sample scoring at least 10 points (range = 10 to 26) higher on the EAS. (Nine percent of the sample had WISC-R scores at least 11 points higher than their EAS scores.) The majority of subjects (52%) had score differences of 10 points or less.

Wright and Piersel thus concluded that, although approximately 53% of variance was shared between the two measures, the large standard error of estimate, as well as

the wide range of score discrepancies for individual cases, indicated that, with regard to the grouping of students:

> The WISC-R and EAS could lead to very different decisions for individual children. . . . Its [EAS] use as an ability measure to answer questions about program placement cannot be recommended, since the high standard error of the estimate indicates a large proportion of classification errors would occur. (p. 70)

On the other hand, the researchers suggested that the EAS may be considered for use as a screening instrument, particularly in the event of extreme EAS score or unusually large differences between the EAS and achievement measures. Such results, according to Wright and Piersel, may be utilized to identify students in need of further attention.

The WISC-R and the Short Form Test of Academic Aptitude. As a component of a larger research project, Wikoff and Parolini (1982) required a measure of cognitive ability for each of their subjects. Although approximately half of the sample had a recent WISC-R score, the remaining subjects had not received individual evaluations. Given that each student's school records contained the results of the group administered Short Form Test of Academic Aptitude (SFTAA) (Sullivan, Clark, & Tiegs, 1974), Wikoff and Parolini sought to determine the extent to which WISC-R scores could be estimated from SFTAA results, thereby potentially providing a valid estimate of cognitive ability for all research subjects. Composed of five levels that assess children in grades 1.5 through 12, the SFTAA yields two scores, Language and Nonlanguage, with means of 100 and standard deviations of 16.

Data for analyses consisted of the standard scores for both tests obtained for 126 students enrolled in grades 7 through 9. Subjects had been classified as follows: 87 learning disabled, 19 "behaviorally impaired" (p. 388), 1 mildly mentally handicapped, and 19 regular education. The WISC-R Full Scale IQ standard score mean for the total sample was 97.96. The mean SFTAA Language score was 98.55, and the mean SFTAA Nonlanguage score was 93.99. Differences between means were not tested for significance.

Inter-instrument relationships were examined through both linear and multiple regression analyses. In order, Pearson correlations between the SFTAA Language score and the WISC-R Full Scale IQ, Verbal IQ, and Performance IQ were .61, .40, and .68. The relationships between the SFTAA Nonlanguage score and WISC-R IQs were .36, .56, and .56, respectively. Significance levels for these relationships were not reported. A multiple regression analysis utilizing both the SFTAA scores to predict the WISC-R Full Scale IQ yielded a multiple correlation coefficient of .73, thus accounting for approximately 53% of the WISC-R variability.

The Test of Cognitive Skills

Overview

The Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1981) is a group-administered test of cognitive ability "designed to assess a student's academic aptitude and thereby predict the student's level of success in school"

(p. 1). The TCS composite, the Cognitive Skills Index (CSI), is an age-based standard score that has a mean of 100 and standard deviation of 16.

Five overlapping levels of the TCS assess grades 2 through 12:

Level 1	Grades 2-3
Level 2	Grades 3-5
Level 3	Grades 5-7
Level 4	Grades 7-9
Level 5	Grades 9-12

Each level presents, in a multiple choice format, four subtests: Sequences, Analogies, Memory, and Verbal Reasoning. The Sequences subtest requires the student to select the figure, letter, or number that comes next in a pattern or sequence. Level 1 items are figures, and Level 2 through 5 items include figures, letters, or numbers.

The Analogies subtest for all test levels presents each item as a set of two pictures, A and B, requires the student to examine a third picture (C), and then to select a fourth picture (D) from a series of pictures such that the C:D and A:B relationships are maintained. Item contexts include people, objects, geometric figures, and scenes.

The Memory subtest requires the delayed recall/recognition of material administered prior to Sequences (Level 1) or Sequences and Analogies (Levels 2-5). Students at Level 1 are presented with 20 pairs of pictures, then asked to select the correct partner for single items 15 minutes later. At Levels 2 through 5, stimuli are 20

obscure words and their definitions; tests items present students the words alone 25 minutes later and require students to select the correct item definition.

The Verbal Reasoning subtest of Level 1 requires students to: (a) identify pictures that do not differ from others in a set, (b) recognize illogical pictures, and (c) identify items that depict logical solutions to orally presented situations. Verbal Reasoning items for the Level 2 through 5 tests require classification by common attributes (Level 2), recognition of word relationships (Levels 3-5), identification of essential object and concept features (Levels 2-5), and development of logical conclusions derived from short passages (Levels 2-5). Approximate administration time ranges from 47 minutes (Level 1) to 53 minutes (Levels 2-5).

A major revision of the Short-Form Test of Academic Aptitude (SFTAA) (Sullivan, Clark, & Tiegs, 1974), the TCS's item development and selection was based on the application of Item Response Theory (IRT), a three-parameter statistical model that incorporates (1) item difficulty, (2) item discrimination, and (3) student guessing. Minimal criteria for item selection included the following:

1. The location parameter of the item should fall within an appropriate range of scale scores. This parameter reflects the difficulty of the item.
2. The discrimination rating of the item should be acceptable. This rating reflects the item's ability to distinguish between students who score high on the test and students who score low.

3. The index of model fit for the item should be acceptable. This index indicates how well the item matches the item response model.

4. The item should meet minimal statistical requirements based on samples of black and Hispanic students. Nonbiased items should perform similarly across group, regardless of ethnicity. (CTB/McGraw-Hill, 1982, p. 9)

In addition to item selection, IRT is also applied to the computer scoring for the TCS subtests for individual students.

The TCS was standardized jointly with the Comprehensive Test of Basic Skills (CTBS) (CTB/McGraw-Hill, 1981) on 83,038 students in grades 2 through 12 enrolled in public and private schools across four geographic regions of the United States. Average ethnic proportions for region and community types were reported: African American = 13.2%; Hispanic = 6.5%; and other = 80.3%. The inclusion of special education students (EMR, ED, Hearing Impaired, LD, Orthopedically Handicapped, Speech Impaired, Other) in the standardization sample was based on the following directive to assessment coordinators: "Exclude only those special education students the school district does not include in any of its group achievement testing programs" (CTB/McGraw-Hill, 1983, p. 38). The numbers of special education students who were actually tested and subsequently included in the normative group are not reported.

Two years after the 1981 publication of the TCS, the Technical Report (CTB/McGraw-Hill, 1983) reported reliability and validity data. Subtest Kuder-Richardson 20

reliability coefficients computed from number-correct scores ranged from .72 (Level 5, Sequences; Level 1, Verbal Reasoning) to .90 (Level 2, Sequences). Reliabilities for the CSI were not reported. Statistical data related to validity were provided by product-moment intercorrelations for the TCS and the California Achievement Test, Form C (CAT/C) (CTB/McGraw-Hill, 1980). The TCS was scored using the IRT method. Coefficients between the CAT scores (Total Reading, Total Language, Total Math, Total Battery) and the TCS Total Scaled Score range from .50 (Total Language and TCS Level 1) to .86 (Total Battery and TCS Levels 4 and 5). No comparisons between the TCS and other measures of ability were reported: Blood (1989) stated that a personal communication with G. C. Glick, Project Manager in Product Development for CTB/McGraw-Hill, indicated that Glick knew of no such studies. A 1995 literature review yielded two criterion-related validity studies of the TCS. (These studies are reviewed in the research section.)

Critiques of the Test of Cognitive Skills

In his review of the TCS, Sternberg (1985) described several distinctive assets, including: (a) less emphasis upon accumulated knowledge, e.g., vocabulary and computation, and more emphasis upon reasoning and "ability to learn new information . . . measured via a Memory subtest" (p. 1556); (b) the provision of a Memory subtest that, relative to other such tests, is less cumbersome and has more face validity inasmuch as it "reflects the learning

of new vocabulary" (p. 1557); (c) sophistication in test development, for example, the use of Item Response Theory and procedures to minimize item bias; and (d) norms books that provide numerous methods for computing derived scores.

While asserting that the TCS is "a promising new instrument for assessing high-level mental abilities," (p. 1557) Sternberg also raised several concerns for the test's construction as well its psychometric properties. The Memory test, for example, at Levels 2 through 5 uses low-frequency English words; thus, in Sternberg's view, some examinees may benefit from prior knowledge of these items. Questioning the psycholinguistic soundness of the "essential" (p. 1557) features of words presented by one part of Verbal Reasoning, Sternberg also expressed doubts that the three subtest components measure a unitary ability. Furthermore, provided with only a preliminary report of technical information, Sternberg challenged the utility of the TCS on the basis of inadequate reliability and validity data.

In the absence of a sufficient technical information, Keith (1985) also criticized the psychometric soundness of the TCS. In addition to the lack of reliability and validity data, Keith argued that the TCS is further limited by its failure to provide: (a) inter-level continuity data, particularly since the Memory and Verbal Reasoning subtest demands are "drastically" (p. 1555) different at Level 1; (b) information concerning the relationship between IRT

computer scored and hand scored results; and (c) student standardization selection procedures, the absence of which may support the assumption that schools, rather than children, constituted the normative sample. Among the TCS's assets, Keith noted the use of IRT, freedom from bias procedures, emphasis on reasoning skills, and excellent manuals for both examiners and test coordinators.

Provided with the TCS's Technical Report (CTB/ McGraw-Hill, 1983), Troy (1985) deplored its failure to report group means. While supporting the assets of the examiner and norms manuals as well as the application of IRT for test development and scoring, Troy also challenged the assumption that unitary abilities are measured by the separate subtests. He judged the CSI to be a good predictor of academic achievement and recommended its use for student instructional grouping and for programs evaluation.

Validity Research with the Test of Cognitive Skills

Two studies have examined the comparability between the Test of Cognitive Skills (TCS) and two individually administered tests of ability, the Wechsler Intelligence Scale for Children--Revised (WISC-R) (Wechsler, 1974) and the Stanford-Binet: Fourth Edition (SB:FE) (Thorndike, Hagen, & Sattler, 1986a). In an effort to provide empirical evidence for the valid identification of gifted students, Robinson and Nagle (1992) compared the performance of a sample of 75 gifted students for the TCS, the SB:FE, and the WISC-R. Enrolled in third, fifth, and eighth grade subjects

had been previously identified as gifted through a state-mandated "weighted profile selection criterion" (p. 109). Based on a 100-point system (90 points required for eligibility), the selection criteria allotted 45 points to an aptitude/intelligence test, 45 points to a standardized academic achievement measure, and 10 points to school grades and teacher recommendation.

Results revealed composite tests means of 130.23, 124.65, and 121.39, respectively, for the TCS, WISC-R, and SB:FE. Statistical analyses indicated that students scored significantly higher on the TCS than on both the WISC-R and SB:FE; they also scored significantly higher on the WISC-R than the SB:FE. A summary of absolute differences between individual scores demonstrated that 44% of the sample scored within 5 points of their TCS on the WISC-R Full Scale IQ. Twenty-eight percent scored within 6 to 11 points, and 28% had a standard score difference of at least 11 points. Similar comparisons with the SB:FE showed that 33% of subjects scored within 5 points of their TCS, and 23% scored within 6 to 10 points. For 44% of students, TCS-SB:FE differences exceeded 10 points. Researchers also observed that all students with TCS-SB:FE differences of at least 16 points scored lower on the SB:FE.

The relationships among test scores were also investigated. Correlation coefficients between the TCS and both individually administered tests were significant: WISC-R ($\underline{r}$ = .41), and SB:FE ($\underline{r}$ = .51). The relationship

between the CSI and the WISC-R Verbal IQ was also significant (r = .33); the relationship between the CSI and WISC-R Performance IQ, was not significant (r = .21). For the SB:FE, correlations between the four area scores and the CSI were all significant: Verbal Reasoning (r = .49), Abstract/Visual Reasoning (r = .41), Quantitative Reasoning (r = .35), and Short-Term Memory (r = .30).

These results, according to Robinson and Nagle, suggested that many children enrolled in gifted programs would score higher on the TCS than on the WISC-R and SB:FE; further, results supported those of other researchers (cited by Robinson and Nagle) who have argued that group tests overidentify students for gifted placement. Emphasizing their own concerns, Robinson and Nagle stated:

> The disparity between scores obtained on individual and group tests would not be as critical if group tests were being used as screening devices, but . . . time demands of school psychologists make it extremely difficult to administer individual IQ tests, and many group tests are being used as the sole ability criterion for placement. (p. 111)

In light of their analyses of score discrepancies for individual students, researchers expressed more pronounced concern for the use of the TCS for gifted identification given that "a significant proportion of children will show substantial differences in scores when comparing this group test with the individually administered tests" (p. 111). Based on their integrated findings, Robinson and Nagle concluded that the TCS may be used as a screening device

provided that the TCS results were supported by an individual test of cognitive abilities.

The comparability of the TCS and SB:FE for students enrolled in special education programs was examined by Blood (1989). Scores from the two instruments were collected from the cumulative school files of 75 special education students, ages 7 to 12 years, grades 2 through 6, who participated in regular education classes for at least 50% of their school day. Statistical analyses yielded a significant .70 correlation between the TCS CSI and the SB:FE Composite. The 2.37-point difference between means, with a higher SB:FE Composite, was also significant.

Blood concluded that the data supported the use of the CSI as a "worthwhile contribution" (p. 64) to school psychologists as referral information, to speech pathologists for determining eligibility for language therapy, and to teachers for program planning. A number of interpretive concerns, however, may be raised for this study. For example, although special education students were selected as the subject sample, no information pertaining to special service classifications (e.g., learning disabled, mildly mentally handicapped, and/or emotional handicapped) was provided. The collapse of data across categories and variations in operational definitions, particularly with respect to cognitive ability criteria, may thus confuse the interpretation of statistical results.

Further, although the statistically significant 2.37-point difference between means was described as small for practical purposes, the accuracy within which an individual's SB:FE Composite can be estimated from a given TCS CSI was not investigated. The valid utility of the TCS for individual students, therefore, requires further study.

Chapter Summary

The validity of the WISC-R as an individually administered measure of cognitive abilities has been empirically examined by a sizeable number of investigations. On the basis on this body of research, the WISC-R has become an established criterion for validity studies of fledgling assessment devices.

In contrast to the WISC-R, the validity of the TCS is subject to several lines of inquiry. As a group administered test, the TCS shares the questionable validity status of other group tests of ability. More specifically, the TCS has received mixed reactions from test reviewers. While extolling the attributes of test development, experts have voiced concerns for the TCS's psychometric properties.

A review of research for empirical evidence revealed scant research with the TCS to support its criterion-related validity as an estimate of cognitive ability. Test manuals and technical reports provide validity data limited to the intercorrelations between the TCS and California Achievement Test. Two published studies comparing the TCS to individually administered measures of intelligence serve to

highlight the need for further research to support TCS criterion-related validity. For example, results of a study that compared the TCS with the WISC-R and SB:FE prompted researchers to issue a strong caution to school personnel when using the TCS as a screening instrument for gifted students. Moreover, for a second study, the omission of key indices of variation and critical descriptive sample information limit the interpretation of the comparison of the TCS with the SB:FE.

The psychometric soundness of the TCS, therefore, continues to be contingent on the investigation of its validity. Furthermore, the examination of the relationship between the CSI and an established criterion, such as the WISC-R, can contribute information to school personnel concerning the valid utility of the CSI in addressing concerns for both regular and special education students.

Chapter 3

METHODOLOGY AND PROCEDURES

This chapter presents the methodology and procedures of the study. The study is first summarized through a restatement of the purpose and null hypotheses of the investigation. Next, a description of the sample, which includes the source of the data, the basis for sample selection, and the procedures for data collection, is provided. A description of the assessment instruments, the Wechsler Intelligence Scale for Children--Revised (WISC-R) (Wechsler, 1974) and the Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1981), is then given and followed by an explanation of the treatment and statistical analyses of the data. Finally, conditions of the study that were beyond the researcher's control are discussed, and the assumptions of the study are stated.

Summary of the Study

Purpose of the Study

The purpose of the present study was to investigate the criterion-related validity of the Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1981). For students with learning

disabilities, the TCS and the Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) were compared to examine: (a) the differences between standard score means of the TCS Cognitive Skills Index (CSI) and each of the WISC-R IQs and factor scores, (b) the relationships between standard scores of the TCS CSI and each of the WISC-R IQs and factor scores, and (c) the range of confidence within which the WISC-R Full Scale IQ can be estimated from the CSI. Post-hoc analyses were also conducted to study CSI-FSIQ differences and to examine subject characteristics that may be associated with TCS performance.

Null Hypotheses

1. For students with learning disabilities, there are no significant differences between standard scores means of the TCS CSI and the WISC-R Full Scale IQ, Verbal IQ, Performance IQ, Verbal Comprehension factor, Perceptual Organization factor, and Freedom from Distractibility factor.

2. For students with learning disabilities, there are no significant relationships between standard scores of the TCS Cognitive Skills Index and the WISC-R Full Scale IQ, Verbal IQ, Performance IQ, Verbal Comprehension factor, Perceptual Organization factor, and Freedom from Distractibility factor.

Description of the Sample

Source of Data

Subjects of the study were selected from public school students in first through eighth grade who received learning disability special education services. All students resided in a single school corporation within the midwest region of the country. Demographic data presented in Table 1 indicate that the characteristics of the subjects' county of residence are representative of the statewide population.

Basis for Sample Selection

Subjects for the study met the following criteria documented by state and county forms: (a) placement in learning disability special services as determined by the Initial Case Conference Committee, (b) a CSI score obtained from Levels 2-5 of the Test of Cognitive Skills, and (c) WISC-R results obtained within 3 years of the TCS administration.

The Indiana State Board of Education (1988; 1992) defines learning disability by specifying the following:

A learning disability:

(a) is characterized by severe specific deficits in perceptual, integrative, or expressive processe that severely impair learning efficiency;

(b) includes conditions referred to, or previously referred to as perceptual handicaps, brain injury,

Table 1. 1989/1990 Demographic Data for Subjects' County and State.

	County	State
Income and Poverty Status		
Per capita income 1989	$10,538	$13,149
% below poverty		
Total population	11.8	10.7
Families	9.0	7.9
Children 5-17 years	12.8	13.0
Education for Persons 25 Years and Over		
% of high school graduates or higher	75.9	75.6
% of bachelor's degree or higher	9.8	15.6
Employment State for Persons 16 Years and Over		
% in labor force	58.7	65.9
% unemployed	6.0	5.7
Selected Social Characteristics for Family Households		
% of married couples with children under 18	46.6	47.3
% of female householder with children under 18	48.2	59.7

minimalbrain dysfunction, dyslexia, and developmental aphasia;

(c) may be manifested in disorders of listening, thinking, talking, reading, writing, spelling, or arithmetic; and

(d) does not include learning problems due primarily to visual, hearing, or orthopedic impairments; mental or emotional handicaps; or environmental, cultural, or economic disadvantages.

Prior to the 1992 revisions of Special Education Rules (Indiana State Board of Education, Article 7), the definition of learning disability also indicated that eligible students were "chronic failures in the regular classroom" (Indiana State Board of Education, Rule S-1, 1988, p. 45).

Placement in learning disability special services culminates the school corporation referral process implemented for each student. The process begins with a pre-referral stage initiated by a parent or teacher request for assistance in addressing academic, socioemotional, and/or behavioral concerns for a student. For most students, these concerns are first addressed by the school's Teacher Assistance Team (TAT), which meets with the teacher and/or parent for the purpose of clarifying the student's difficulties, developing intervention recommendations, delegating responsibilities for intervention implementation, providing supplemental support, and designating a schedule for case review. At case review, the TAT may determine that additional assistance is required; for example, assistance is frequently requested from the school psychologist who then receives the pre-referral information regarding teacher/parent concerns and their attempts to remediate the student's difficulties.

In the next referral stage, the school psychologist contacts the student's parents (or legal guardian) to provide the opportunity to discuss their concerns, to obtain

the student's developmental history, and to gather other relevant background information. The school psychologist also explains the assessment process, answers questions, and details the parent's legal rights regarding evaluation and special services. The parents are then given the option of consenting to an individual evaluation of the student.

Given signed parental consent, a multidisciplinary evaluation team is assembled. As mandated by the state, this team includes the child's teacher, the school psychologist, and a specialist in the area of suspected disability. The individual student evaluation for learning disability includes measures of both ability (intelligence) and academic achievement. The multidisciplinary team provides a written report to the Initial Case Conference Committee detailing the basis for learning disability eligibility decisions. The written report includes the documentation of a severe ability-achievement discrepancy, operationally defined by scores from individually administered instruments, and the exclusion of the effects of environmental, cultural, or economic disadvantage as the primary cause for the student's learning difficulties.

For the subjects of this investigation, the discrepancy between a student's ability and achievement was most frequently defined by a standard score difference between actual and predicted achievement. Actual achievement levels were assessed primarily through the individually administered Woodcock-Johnson Tests of Achievement (Woodcock

& Johnson, 1977) and the Woodcock-Johnson Tests of Achievement-Revised (Woodcock & Johnson, 1988). Predicted achievement levels were derived from the composite standard scores of an individually administered test of cognitive ability, most commonly, the Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974).

As a general policy, and, in concert with the recommendations of the Indiana Learning Disabilities Task Force (1982; 1987), the county has largely based predicted achievement upon a regressed IQ method. Using age-normed standard scores, this method accounts for the correlational relationship between ability-achievement measures, errors of measurement, and regression to the mean. The Task Force supported an 18-point ability-achievement difference as a requisite component of defining a severe discrepancy.

The actual-predicted achievement discrepancy must also be supported by a second instrument. The second instruments used for this subject sample included the Wide Range Achievement Test-Revised (WRAT-R) (Jastak & Wilkinson, 1984), Kaufman Test of Educational Achievement (K-TEA) (Kaufman & Kaufman, 1985), and the Basic Achievement Skills Individual Screener (BASIS) (The Psychological Corporation, 1983).

For each child found eligible for learning disability services, the Initial Case Conference designates appropriate placement. Four placement levels of learning disability services are operationally defined by the state according to

the percentof total instructional time the child receives special education services: Full-Time (over 50%); Part-Time (25% to 49%); Resource (up to 24%); and Consultation (0%).

Data Collection

The data were collected by the researcher. A list of potential subjects was first compiled from the classroom rosters of special education teachers throughout the school corporation. For each student, a data sheet was completed. All identifying information was omitted.

Results of the Test of Cognitive Skills (TCS) for potential subjects were located in individual cumulative folders in their home schools. Within the cumulative folder for each student, TCS and achievement results of the ISTEP (Indiana Statewide Testing for Educational Progress) battery are affixed to a designated index card. Inasmuch as the ISTEP battery is administered to grades 2, 3, 6, 8, 9, and 11, several sets of ISTEP results were frequently included on the index card. All data were photocopied within the school building and returned to the files.

Data collection was continued by accessing individual special education files housed in the main special services office. Each file was examined to determine the date of initial placement, to verify both learning disability eligibility and placement decisions, and to record initial assessment results.

Description of the Instruments

Wechsler Intelligence Scale for Children--Revised

The Wechsler Intelligence Scale for Children--Revised (WISC-R) (Wechsler, 1974) is an individually administered test of cognitive ability for individuals ages 6-0 to 16-11. Based upon 10 of 12 subtests, the WISC-R Full Scale IQ provides a global estimate of cognitive functioning. Supplemental deviation scores may be computed from subtest scaled scores: Verbal IQ; Performance IQ; Verbal Comprehension factor; Perceptual Organization factor; and Freedom from Distractibility factor. All deviation scores have a mean of 100 and a standard deviation of 15.

A revision of the Wechsler Intelligence Scale for Children (Wechsler, 1949), the WISC-R has been lauded for numerous strengths. Standardization procedures, for example, were designed to yield a sample representative of the United States according to 1970 census data: Demographic variables represented by the sample include gender, race, and socioeconomic status (Sattler, 1988; Wechsler, 1974).

Psychometric properties of the WISC-R constitute a second area of strength. Wechsler (1974) reported average split-half correlations for 11 age groups as .96, .94, and .90, respectively, for the Full Scale IQ, Verbal IQ, and Performance IQ. These high reliability coefficients extend across the age groups. Average split-half correlations for the age groups for the 12 subtests range from .77 to .86 for

Verbal subtests and from .70 to .85 for Performance subtests. For separate age groups, 126 subtest reliabilities equal or exceed .65. Of the 6 coefficients falling below .65, only one is less than .60 (r=.57 for the Mazes subtest at age 16.5 years).

High test-retest reliabilities were also found for the retesting of 303 children from the standardization sample following an interval of approximately one month. Stability coefficients were reported as follows: Full Scale IQ, .95; Verbal IQ, .93; Performance IQ, .90.

A function of test score reliability and variability, the standard errors of measurement provide an indication of confidence in making judgments regarding a child's true ability. Such confidence, operationalized statistically by bands of error around a test score, is a decided strength for the WISC-R (Sattler, 1988). Standard errors of measurement, averages across 11 age groups, for the Full Scale IQ, Verbal IQ, and Performance IQ are 3.19, 3.60, and 4.66, respectively. Varying somewhat for age, standard errors of measurements for the Full Scale IQ ranges from .95 to .96 for all age groups. The Verbal IQ standard error of measurements varies from a low of .91 for age 6.5 to .96 for age 12.5. The standard error of measurement ranges from .89 to .91 for all age groups for the Performance IQ. These results also increase confidence in utilizing the WISC-R as a measure of cognitive ability by narrowing the size of difference between Verbal IQ and Performance IQ required for

statistical significance: a computation critical to valid test interpretation.

Wechsler (1974) supported the validity of the WISC-R through intercorrelations with three other measures of cognitive abilities. For the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967), the Full Scale IQ's correlated .82, and the WPPSI mean exceeded the WISC-R mean 2.5 points. WISC-R comparisons with the Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 1955) yielded a mean difference of 6.2, with a higher WAIS, and a Full Scale IQ correlation of .95. Based on the 1972 norms, the Stanford-Binet (Form L-M) (Terman & Merrill, 1960) IQ exceeded the WISC-R Full Scale IQ by an average of 1 point and reached a validity coefficient of .73.

In addition to Wechsler's research, a multitude of investigations have examined the criterion-related validity of the WISC-R by comparing the Full Scale IQ with composite scores of other individually administered measures of cognitive ability. Concurrent validity coefficients of these studies are reported in the range of .56 for the Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman, 1983) Mental Processing Composite (MPC) for a group of nonreferred fourth-graders (Klanderman, Devine, & Moller, 1985) to .89 for the General Cognitive Index (GCI) of the McCarthy Scales for Children (McCarthy, 1972) for first grade children in regular education (Reilly, Drudge, Rosen, Loew, & Fisher, 1985).

Sattler (1988) also supported the criterion-related validity of the WISC-R with summaries of research that, in addition to including cognitive ability measures, explored other related variables. He reported median correlations of .56 to .71 between the WISC-R Full Scale IQ and individual achievement instruments, and a median correlation of .39 between the WISC-R Full Scale IQ and school grades. (Additional descriptive information and validity research for the WISC-R are provided in Chapter 2.)

Test of Cognitive Skills

The Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1981) is a group-administered test of cognitive ability. Developed to assess students in grades 2 through 12, five overlapping levels of TCS present a multiple choice format of four subtests: Sequences, Analogies, Memory, and Verbal Reasoning. The four subtests combine to yield a composite, the Cognitive Skills Index (CSI), an age-based standard score with a mean of 100 and standard deviation of 16.

The TCS is a major revision of the Short-Form Test of Academic Aptitude (SFTAA) (Sullivan, Clark, & Tiegs, 1974). Strengths of the TCS include: (a) the application of Item Response Theory for both development and computer scoring, (b) less emphasis on accumulated knowledge and more emphasis on the ability to learn new material and reasoning skills, (c) the inclusion of a Memory subtest, and (d) well-written examiner manuals (Keith, 1985; Sternberg, 1985; Troy, 1983).

The TCS was standardized jointly with an achievement battery, the Comprehensive Test of Basic Skills (CTBS) (CTB/McGraw-Hill, 1981), on 83,038 students enrolled in public and private schools selected from four geographic areas of the United States. Keith (1985) faulted the TCS standardization procedures that selected schools, rather than children, for the normative sample. The ethnic proportions of the sample are described in the Technical Report (1983); however, the extent to which special education students may have been included in the sample is unclear inasmuch as student selection was left to the discretion of assessment coordinators: "Exclude only those special education students the school district does not include in any of its group achievement testing programs" (CTB/McGraw-Hill, 1983, p. 38).

Test reviewers have questioned the psychometric properties of the TCS. Reliability information is limited to Kuder-Richardson 20 coefficients for number-correct subtest scores. These coefficients range from .72 (Level 5, Sequences; Level 1, Verbal Reasoning) to .90 (Level 2, Sequences). No reliability data are reported for the CSI standard score.

Pearson product-moment correlations between the TCS total scaled scores and the California Achievement Test, Form C (CAT/C) (CTB/McGraw-Hill, 1980) total scores for Reading, Language, Math, and Battery Composite are reported as validity data for the TCS. Coefficients range from .50

(Total Language and TCS Level 1) to .86 (Total Battery and TCS Levels 4 and 5).

Robinson and Nagle (1992) compared the TCS to both the WISC-R and SB:FE for a sample of 75 gifted students in third, fifth, and eighth grade. Results indicated that students scored significantly higher on the TCS than on the WISC-R and the SB:FE. Correlation coefficients between the TCS CSI and the composite scores of the individually administered tests were significant. A summary of absolute differences between test scores for individual students indicated that 44% of students scored within 5 points of their TCS CSI on the WISC-R FSIQ, and 28% scored within 6 to 10 points. The remaining 28% of subjects had a CSI-FSIQ difference of a least 11 points. Comparisons with the SB:FE composite indicated that 33% of students scored within 5 points of their CSI, and 23% scored within 6 to 10 points. Differences between the CSI and SB:FE composite exceeded 10 points for 44% of the students. For those with TCS-SB:FE differences exceeding 15 points, all subjects scored lower on the SB:FE.

Blood (1989) reported a significant correlation of .70 between the CSI and the Composite Score of the SB:FE for a sample of 75 special education students enrolled for at least 50% of their school day in regular education classes in grades 2 through 6. Information regarding area of services (for example, mildly mentally handicapped, hearing impaired, orthopedically handicapped, emotionally

handicapped) was not given. (The TCS is further discussed in Chapter 2.)

Design and Statistical Treatment of Data

Computer analyses of the data were conducted by the university computer center through the use of the Statistical Analysis System (SAS) (SAS Institute, 1982). Differences between standard score means from the TCS and WISC-R were examined through two-tailed, *t*-tests for correlated samples ($p < .05$). Relationships between standard scores for each CSI and WISC-R pair were determined through the computation of Pearson product-moment correlation coefficients which were tested for significance ($p < .05$). The range of confidence within which the WISC-R Full Scale IQ may be estimated from the CSI was determined through the application of a linear regression equation.

Limitations

The term "learning disability" was a source of limitation for this study both theoretically and pragmatically. The construct of learning disability (e.g., aspects of etiology, the nature of processing deficits or differences, and the classification of subtypes) continues to be the focus of debate and general controversy. The concomitants of such controversy are the means by which researchers and practitioners operationally define a learning disability.

Pertinent to this study, the requisite discrepancy between ability and achievement was not specified by state mandates and thus left to the parameters designated by the multidisciplinary team. The regressed IQ discrepancy method, advocated by the state-level Learning Disabilities Task Force (1982, 1987) was predominantly utilized in the subjects' school corporation. Nevertheless, all eligibility and placement decisions, operational definitions and evaluation results notwithstanding, were ultimately left to the judgment of the Initial Case Conference Committee.

Assumptions

1. All tests were administered and scored according to standardized procedures.

2. Achievement tests yielded valid estimates of reading, writing, and mathematics skills.

3. In determining eligibility for each subject, a multidisciplinary team operationally defined learning disability according to the stated criteria, thereby augmenting consistency in eligibility (diagnosis) for the sample.

4. The Level 2, 3, and 4 forms of the Test of Cognitive Skills (TCS) assess the same cognitive abilities; therefore, scores from the three forms may be combined for data analyses.

Chapter 4

RESULTS

This chapter presents the results of the statistical analyses of the data relevant to the two main research questions. The study is first summarized, and the subject sample is described. Data analyses conducted in order to answer the research questions are then presented. Finally, the results of several post-hoc analyses are given.

Summary of the Study

The purpose of the present study was to examine the criterion-related validity of the Test of Cognitive SKills (TCS) (CTB/McGraw-Hill, 1981), a group administered test of cognitive ability. For students with learning disabilities, standard scores from the TCS and the individually administered Wechsler Intelligence Scale for Children-Revised (WISC-R) were compared in order to answer two research questions:

Research Question One: For students with learning disabilities, are there a significant differences between the standard score means of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ, Verbal IQ, Performance

IQ, Verbal Comprehension factor, Perceptual Organization Factor, and Freedom from Distractibility factor?

Research Question Two: For students with learning disabilities, do significant relationships exist between the standard scores of the TCS CSI and the WISC-R Full Scale IQ, Verbal IQ, Performance IQ, Verbal Comprehension factor, Perceptual Organization Factor, and Freedom from Distractibility factor? Given that a significant relationship exists between the TCS CSI and the WISC-R Full Scale IQ, within what range of confidence can the WISC-R Full Scale IQ be estimated from the TCS CSI for individual students?

Data for the study were the WISC-R scores from the subjects' initial evaluations for special education services and the TCS Composite scores, the Cognitive Skills Index (CSI), from the computer-scored outcome of group testing. The time between testing for individual subjects was limited to a three-year interval.

Statistical analysis of data was conducted by the university computer center through Statistical Analysis System (SAS) (SAS Institute, 1982) computer software.

Subjects

Subjects of the study were 118 (males = 86; females = 32) public school students in first through eighth grade (M = 3.68), ages 81 to 162 months (M = 116.96), who received learning disability special education services. Only those students found eligible for and placed in learning

disability services by the Initial Case Conference were considered for subject selection. Enrolled in 10 schools with 12 learning disability teachers, all subjects resided in a single school corporation within the midwest region of the country. (Demographic data for the region are presented in Chapter 3.)

Subjects of the study met the following selection criteria: (a) placement in learning disability special services, (b) a CSI score obtained from Levels 2-5 of the Test of Cognitive Skills, and (c) WISC-R results obtained with three years of the CSI.From a preliminary compilation of 309 students enrolled in learning disability services, 46 students were eliminated from the sample because their home school records lacked the requisite CSI data from TCS Level 2, 3, or 4. Five additional students were excluded on the basis of errors in recorded birthdates that resulted in invalid CSI scores. Following the review of special education files, an additional 140 students were eliminated from the subject group on the basis on one or more of the following reasons:

(a) The Initial Case Conference determined the student eligible for special education services in an area other than learning disability, for example, mildly mentally handicapped, traumatic brain injury;

(b) The Initial Case Conference determined the student ineligible for special education services;

(c) The WISC-R was not administered for initial assessment;

(d) The file data was incomplete for students who transferred from learning disability programs outside the school corporation, for example, no WISC-R scores and/or Initial Case Conference summary results were received through exchanged information;

(e) The time interval between the WISC-R and TCS administration exceeded 3 years.

Table 2 presents a descriptive summary for the total sample with regard to number, gender, age and grade at initial placement, and placement status. Of the total sample, the proportion of students in the four placement areas were as follows: Full-Time (6.78%), Part-Time (43.22%), Resource (42.37%), and Consultation (7.63%).

Table 2. Mean Age (in months) and Grade (in tenths of year) for Subjects at Initial LD Placement.

	Gender n			Mean Age	Mean Grade
	n	F	M	(range)	(range)
Total Sample	118	32	86	116.96 (81-162)	3.68 (1.0-8.1)
Placement Groups					
Full-Time	8	3	5	124.13 (92-148)	4.20 (1.2-7.0)
Part-Time	51	14	37	116.45 (86-162)	3.67 (1.0-8.1)
Resource	50	11	39	115.34 (81-161)	3.57 (1.4-7.5)
Consultation	9	4	5	122.44 (92-154)	4.00 (1.7-7.0)

Table 3 displays the content areas in which subjects were found eligible for learning disability services. Review

of the domains of Reading, Written Expression, and Mathematics indicates that the most prevalent area of eligibility was Written Expression for which 78% of students received services. Special services for Reading and Mathematics disability were received by 75% and 25% of the sample, respectively.

Table 3. Content Areas of LD Eligibility for Subjects at Initial Placement.

	n	%
Reading	13	11.0
Mathematics	9	7.6
Written Expression	17	14.4
Reading, Mathematics	4	3.4
Reading, Written Expression	58	49.2
Mathematics, Written Expression	4	3.4
Reading, Mathematics, Written Expression	13	11.0

A primary criterion considered by the Initial Case Conference for the determination of learning disability eligibility is a discrepancy between the student's actual and expected level of academic achievement. For this sample, 83% of students met or exceeded the 18-point difference between actual achievement scores and the expected (or projected) achievement scores estimated through the regressed IQ method applying an inter-instrument correlation of .65. The average difference between

estimated and actual achievement was 21.10 standard score points (SD = 7.72).

In addition to the receipt of learning disability services, subject selection for the study was based on the availability of CSI scores obtained from Levels 2-5 of the Test of Cognitive Skills. The number of students in the final sample with CSIs (selected for data analyses) from the different TCS levels were: Level 2 = 86 students (72.88%); Level 3 = 28 students (23.73%); Level 4 = 4 students (3.39%). The Test of Cognitive Skills was administered to these subjects by school faculty and staff who had received training in administering this test. All subjects had been tested in the regular education classroom along with their grade peers.

Selection was also based on the results of a WISC-R administered within three years of the TCS. The mean time between test administration was 14.62 months (SD = 9.05). The WISC-R had been individually administered to subjects by a licensed school psychologist. Twenty-seven percent of students had been administered the TCS prior to the WISC-R.

Analysis of Research Question One

For students with learning disabilities, are there significant differences between the standard score means of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ, Verbal IQ, Performance IQ, Verbal Comprehension factor, Perceptual Organization Factor, and Freedom from Distractibility factor? To address this inquiry, the

differences between standard score means for each of the six WISC-R and the TCS CSI were analyzed through six, two-tailed t-tests for correlated samples.

Table 4 displays the means and standard deviations of test scores for the total subject sample. Means for the CSI, WISC-R IQs, Verbal Comprehension, and Perceptual Organization were within the average range of the normative groups; the Freedom from Distractibility mean was within the low average range.

Table 4. Means and Standard Deviations for WISC-R and CSI Standard Scores.

Test Score	Mean	Standard Deviation
Cognitive Skills Index	92.47	11.13
WISC-R IQs		
Full Scale	95.58	10.62
Verbal	94.75	11.71
Performance	98.10	12.20
WISC-R Factors		
Verbal Comprehension	96.72	12.17
Perceptual Organization	100.81	13.38
Freedom From Distractibility	86.84	10.61

CSI data were also available for both the subjects' county and state (See Table 5). CSI means for both the county (110.0) and the subject sample (94.47) were within an average range; the mean for the learning disability group, however, was a standard deviation (15.53 standard score points) below the mean for the county.

Results of the tests of significance between the CSI and six WISC-R scores are presented in Table 6. The t-test

Table 5. CSI Means for the State and School Corporation.

1992-1993

Mean CSI for all state school corporations	109.4
Mean CSI for subjects' school corporation	110.0

1987-1993 CSI Grade Means for Subjects' School Corporation

	Grade				
Year	2	3	6	8	9
92-93	107.6	113.6	110.3	110.0	108.4
91-92	108.5	111.7	111.8	110.6	108.0
90-91	107.9	111.7	109.5	109.4	106.1
89-90	107.9	110.7	108.7	108.6	105.8
88-89	109.0	110.5	107.1	106.6	105.6
87-88	106.4	108.9	106.6	108.4	106.3

comparing subjects' CSI and Full Scale IQ yielded a t value of -3.65 (df = 117, p = .0001) which was significant at the .05 level.

Table 6. Differences Between CSI and WISC-R Standard Score Means.

Scores	t	p
CSI-Full Scale IQ	-3.65	.0001*
CSI-Verbal IQ	-2.28	.024*
CSI-Performance IQ	-4.71	.0001*
CSI-Verbal Comprehension	-4.00	.0001*
CSI-Perceptual Organization	-7.05	.0001*
CSI-Freedom from Distractibility	5.25	.0001*

* Significant at $p < .05$

The *t*-test comparing subjects' CSI and Verbal IQ yielded a *t* value of -2.28 (*df* = 117, *p* = .024), which was significant at the .05 level.

The *t*-test comparing subjects' CSI and Performance IQ yielded a *t* value of -4.71, (*df* = 117, *p* = .0001), which was significant at the .05 level.

The *t*-test comparing subjects' CSI and Verbal Comprehension factor yielded a *t* value of -4.00 (*df* = 117, *p* = .0001), which was significant at the .05 level.

The *t*-test comparing subjects' CSI and Perceptual Organization factor yielded a *t* value of -7.05 (*df* = 117, *p* = .0001), which was significant at the .05 level.

The *t*-test comparing subjects' CSI and Freedom from Distractibility factor yielded a *t* value of 5.25 (*df* = 117, *p* = .0001), which was significant at the .05 level.

Results indicated significant differences between the standard score means of the TCS Cognitive Skills Index and the WISC-R Full Scale IQ, Verbal IQ, Performance IQ, Verbal Comprehension factor, Perceptual Organization Factor, and Freedom from Distractibility factor for students with learning disabilities. Examination of the data also revealed that, relative to their CSI performances, students scored significantly higher on the WISC-R Full Scale IQ, Verbal IQ, Performance IQ, Verbal Comprehension factor, and Perceptual Organization factor. Subjects' Freedom from Distractibility factor mean, on the other hand, was significantly lower than their CSI mean. (The Freedom from

Distractibility factor and the CSI are explored further in the post hoc analyses.)

Analysis of Research Question Two

For students with learning disabilities, do significant relationships exist between the standard scores of the TCS CSI and the WISC-R Full Scale IQ, Verbal IQ, Performance IQ, Verbal Comprehension factor, Perceptual Organization Factor, and Freedom from Distractibility factor? Given that a significant relationship exists between the TCS CSI and WISC-R Full Scale IQ, within what range of confidence can the WISC-R Full Scale IQ be estimated from the TCS CSI for an individual student? The relationships between the CSI and each WISC-R score were examined through the computation of Pearson product-moment correlations which are displayed in Table 7.

Results indicated that, for students with learning disabilities, the standard scores of the CSI and the WISC-R are significantly related: All Pearson coefficients for the CSI paired with each WISC-R score were positive and significantly different from zero. The strongest relationship was found between the CSI and WISC-R Full Scale IQ with approximately 41% of shared variance between the two tests.

Given the significant relationship between the TCS CSI and WISC-R Full Scale IQ (r =.6406, p < .05), the second component of Research Question Two was pursued through

Table 7. Pearson Correlations for Total Sample.

	CSI	FSIQ	VIQ	PIQ	VC	PO
FSIQ	.6406*					
VIQ	.5508*	.7639*				
PIQ	.4296*	.7513*	.1905*			
VC	.5144*	.7137*	.9597*	.1571		
PO	.4609*	.7636*	.2494*	.9533*	.2120*	
FD	.4254*	.6085*	.4444*	.4569*	.3065*	.3390*

* $p < .05$

regression analysis. Regression of the Full Scale IQ (Y) on the CSI (X) yielded the following equation:

$$\hat{Y} = .6115X + 39.0281, \text{ where } \hat{Y} = \text{estimated } Y$$

Estimated Full Scale IQs were then computed by substituting each student's CSI standard scores for X in the regression equation. The standard deviation of the error, that is, the standard deviation of differences between the computed and actual Full Scale IQs, was 8.1934. (Table 8 provides the standard errors of estimate for the six WISC-R standard scores.)

Through the application of the standard error of estimate, the ranges of confidence within which the WISC-R Full Scale IQ can be estimated from the TCS CSI for an individual student were quantified by the following statements:

Confidence Level	Range for Estimated Full Scale IQ
90%	Computed FSIQ +/- 13.5191
95%	Computed FSIQ +/- 16.0591
99%	Computed FSIQ +/- 21.1490

Table 8. Standard Errors of Estimate for WISC-R Scores Given the TCS CSI.

	Standard Error of Estimate
WISC-R IQs	
Full Scale	8.1934
Verbal	9.8172
Performance	11.7759
WISC-R Factors	
Verbal Comprehension	10.4846
Perceptual Organization	11.8868
Freedom from Distractibility	9.6446

These statements indicate that, given the CSI for an individual student, the estimated FSIQ will range across approximately 28 standard score points in order to achieve a 90% probability of accuracy. A 33-point range of standard scores, or more than two standard deviations, is required in order to estimate a FSIQ from the CSI within a 95% probability of accuracy. Estimation of the FSIQ with an accuracy of 99% probability will require a range of 43 standard score points.

Post-Hoc Analyses

The purpose of the post-hoc analyses was two-fold: (a) to further examine the comparability of the CSI and FSIQ from several pragmatic perspectives; and (b) to explore

subject characteristics that may be associated with CSI results.

Comparability of the CSI and FSIQ

Pertinent to the use of the CSI as an estimate of general cognitive ability, the substantial confidence intervals for estimating the WISC-R Full Scale IQ from the TCS CSI suggested that a wide range of CSI-FSIQ discrepancy existed for individual subjects in the sample. The distribution of absolute values of CSI-FSIQ differences summarized in Table 9, revealed that 53% of the students in the subject sample produced CSI scores within five points of their FSIQs. Further review of the data showed that about one of every five students in the total sample (22.88%) had a standard score difference of at least 11 points between the CSI and FSIQ. Approximately one out of every 10 students (11.02%) showed a CSI-FSIQ discrepancy exceeding 16 standard score points.

Table 9. Distribution of Absolute CSI-FSIQ Differences.

Range	n	% Subjects	n Negative	n Positive
0- 5	62	52.54	30	32
6-10	29	24.58	18	11
11-15	14	11.01	9	5
16-20	8	6.78	8	0
21-25	3	2.54	2	1
26-30	1	0.85	1	0
31-35	0	0.00	0	0
36-40	1	0.85	1	0

The CSI-FSIQ distribution also reflected a direct relationship between the size and direction of the standard

score discrepancy: As the range of discrepancies increased, the relative frequency of negative CSI-FSIQ differences increased. For example, the group of students with CSI-FSIQ differences of 5 points or less was approximately equal in proportion of positive (52%) and negative (48%) differences. Review of students with CSI-FSIQ differences of 10 points or less showed an increased proportion of negative (68%) versus positive (32%) differences.

Examination of wider ranges between the CSI and FSIQ indicated that about 78% of students with differences exceeding 10 points scored lower on the CSI. Stated another way, about 18% of the total sample scored more than 10 points lower on the CSI than the FSIQ. Of those students with differences exceeding 15 points, 92.31% scored lower on the CSI. In other words, one of every 10 students in the total sample scored 16 or more points lower on the CSI than the FSIQ.

The CSI-FSIQ comparability was also examined by grouping students according to WISC-R FSIQ ranges as displayed in Table 10. Specifically, the data were organized to observe relative variations in discrepancies for scores as they approach the mean.

Review of this distribution indicated that, rather than the tendency of scores close to the mean on one variable to yield estimated scores close to the mean on the second variable, substantial CSI-FSIQ differences were observed across ability ranges. For example, approximately one-half

(47%) of the subject group that had WISC-R FSIQ scores in the average range showed discrepancies exceeding five standard score points. Moreover, this average ability group, as operationalized by the WISC-R, represented a CSI-FSIQ range of 42 points (- 28 to +13).

Table 10. Distribution of FSIQ WISC-R Scores and CSI-FSIQ Ranges.

FSIQ Ranges	n	|CSI-FSIQ| Mean	CSI-FSIQ Range	|CSI-FSIQ| > 5 n
130+	0			
120-129	4	18.00	-36 to -3	3
110-119	10	8.10	-22 to +3	5
90-109	68	6.56	-28 to +13	32
80- 89	34	6.71	-12 to +23	16
70- 79	2	3.00	0 to +6	1
68 and Below	0			

Applying conventional ability descriptors, CSI-FSIQ differences were observed from a classification perspective; that is, students were placed in ability groups based upon both the FSIQ and CSI in order to examine the extent to which classification of individuals might change as a result of substituting the CSI for the FSIQ. (See Table 11.)

Results indicated that approximately one-half (47%) of all students would be reclassified by using the CSI; about two out of every three (66%) of these students would be shifted to a lower ability range. Examination of these classification changes revealed that three of the four Superior students were reclassified as Average and one as High Average. Seven of the 10 High Average students were

Table 11. FSIQ and CSI Distribution and Changes in Ability Classifications.

WISC-R Ability Classifications	Classification Changes With CSI				
	n	n	%	% Higher	% Lower
Superior	4	4	100	0	100
Above Average	10	7	70	0	100
Average	68	23	34	13	87
Low Average	34	21	62	71	29
Borderline Deficient	2	1	50	100	0

reclassified as Average on the basis of the CSI. Of the 68 students within the Average range of cognitive ability (again, operationalized by the WISC-R FSIQ) 23 were reclassified into the following categories: Superior (1), High Average (2), Low Average (15), Borderline Deficient (3), and Mildly Mentally Handicapped (2). Fifteen of the students in the Low Average range were shifted to the Average group, whereas six students were reclassified as Borderline Deficient in their abilities. Finally, for the Borderline Deficient group, the classification of one student was changed to Low Average on the basis of the CSI.

Factors Related to CSI Performance

The examination of characteristics of the subjects for the investigation stemmed, in part, from the data analyses conducted to address the main research questions. For example, the six *t*-tests of differences between the CSI and WISC-R means were significant. A distinction, however, emerged from these tests; that is, the Freedom from Distractibility (FD) factor was the single WISC-R mean that

was significantly lower than the CSI. Further investigation was conducted to examine the extent to which, if any, the subjects' FD scores may be associated with their TCS performance.

The FD factor, discussed in both Chapters 1 and 2, is an empirically derived score that is computed from the scaled scores of three WISC-R subtests: Arithmetic, Coding, and Digit Span. As implied, the FD factor subtests may be sensitive to the processes involved in sustained attention. Cautioning that the Freedom from Distractibility term may be overly simplistic, Sattler (1988) also emphasized that FD subtest performance may also be adversely influenced by "anxiety, short-term retention deficits, encoding deficits, poor rehearsal strategies, difficulty in rapidly shifting mental operations on symbolic material, and inadequate self-monitoring skills" (p. 174).

A data analysis was conducted in order to answer the question: Is there a significant difference between CSI means for students with below average FD scores and students with at least average FD scores? Two groups of students were first identified in the subjects sample. The "Average FD" group (n = 38) was composed of students with FD scores at least within the average range (FD ≥ 90). The "Below Average FD" (n = 40) group required the student to have an FD score at least one standard deviation below average (FD < 85) and also below both the student's individual Verbal Comprehension (VC) and Perceptual Organization (PO) factor

scores (at least one FD-VC or FD-PO difference being significant at the .05 level).

Inasmuch as the scaled score for two of the FD subtests are included in the calculation of the WISC-R Full Scale IQ, the VC factor, which does not include the FD subtests, was used as an estimate of general cognitive ability for the two groups. The VC was chosen rather than the PO since processes underlying the FD subtests may also be required by several of the PO subtests (Sattler, 1988).

The CSI and VC standard score means for each group were computed and tested for significance through two-tailed *t*-tests for independent samples. Results are displayed in Table 12 and Table 13.

The *t*-test comparing the CSI mean of students with average FD scores to the CSI mean of students with below average FD scores yielded a *t* value of 3.33 (*df* = 76, *p* = .0001), which was significant at the .05 level.

The *t*-test comparing the VC mean of students with average FD scores to the VC mean of students with below average FD scores yielded a *t* value of 1.92 (*df* = 76, *p* = .055), which was not significant.

Table 12. Difference Between CSI Means for Students with Average and Below Average Freedom from Distractibility (FD) Factor Scores.

Group	*n*	Mean	SD	*t*	*p*
Average FD	38	98.00	10.77	3.33	.0001*
Below Average FD	40	89.80	10.76		

*$p < .05$

Table 13. Difference Between WISC-R Verbal Comprehension (VC). Factor Means for Students with Average and Below Average Freedom from Distractibility (FD) Factor Scores.

Group	n	Mean	SD	t	p
Average FD	38	100.79	15.20	1.92	.055*
Below Average FD	40	95.85	7.00		

*$p \geq .05$

These results indicated that the difference between CSI means for students with average FD scores and for students with below average FD scores was significant: Students with average FD scores had significantly higher CSI means than the students with below average FD scores. Results also demonstrated that the WISC-R Verbal Comprehension factor score means for these two groups were not significantly different.

A second line of inquiry regarding attributes that may be associated with CSI scores also emerged from both the literature review and the data analyses. As cited in Chapter 2, for example, test experts such as Anastasi (1982) have cautioned that group tests may yield scores that underestimate the true cognitive abilities for students with reading disabilities. Providing some support to this view, the present study revealed that subjects, 75% of whom were enrolled in reading disability services, scored significantly lower on the CSI than the WISC-R Full Scale IQ. Thus, a second question was developed for the post-hoc investigation: Is there a significant difference between

CSI means for students with average reading achievement and students with below average reading achievement?

In order to address this question, two groups of students from the subject sample were first identified on the basis of their reading achievement. In this case, reading achievement was defined by the Word Identification or Word Recognition standard score based on age norms of an individually administered achievement instrument. Students were assigned to one of two groups: (a) the "Average" reading achievement group defined by a standard score equal to or exceeding 90 points; or (b) the "Below Average" reading achievement group defined by a standard score below 85, that is, at least one standard deviation below the mean. On the basis of these criteria, 21 students were classified as "Average," and 62 students were classified as "Below Average" in reading achievement. In addition to CSI means, the Full Scale IQ means for the two groups were also tested for significance through two-tailed, *t*-tests for independent samples. Results are displayed in Table 14 and Table 15.

Table 14. Difference Between CSI Means for Students with Average and Below Average Reading Achievement.

Group	*n*	Mean	SD	*t*	*p*
Average Reading	21	97.95	10.61	2.26	.024*
Below Average Reading	61	91.29	13.89		

*$p < .05$

The t-test comparing the CSI mean of students with average reading achievement scores to the CSI mean of students with below average reading achievement scores

Table 15. Difference Between WISC-R Full Scale IQ Means for Students with Average and Below Average Reading Achievement.

Group	n	Mean	SD	t	p
Average Reading	21	100.29	10.46	1.70	.089
Below Average Reading	62	95.68	10.84		

*p ≥ .05

yielded a t value of 2.26 (df = 81, p = .024), which was significant at the .05 level.

The t-test comparing the FSIQ mean of students with average reading achievement scores to the FSIQ mean of students with below average reading achievement scores yielded a t value of 1.70 (df = 81, p = .089), which was not significant at the .05 level.

These results showed a significant difference between the CSI means of students with average reading achievement and students with below average reading achievement. Further analysis revealed that students with average reading achievement scored significantly higher on the CSI than students with below average reading achievement. Results also demonstrated that the mean Full Scale IQs of these two groups were not significantly different.

Chapter 5

SUMMARY, DISCUSSION, AND IMPLICATIONS

This chapter is composed of three sections. The first section summarizes the statement of the problem, the procedures for gathering and analyzing the data, and the research questions addressed by the investigation. The second section discusses the major findings of the study, and the third section suggests implications of the study for both direct practice and continued research.

Summary

Statement of the Problem

The Test of Cognitive Skills (TCS) is a group test of "academic aptitude" (TCS Handbook, 1982, p. 1) that may assist teachers and schoool psychologists in meeting the educational needs of their students. The soundness of decisions that are based on the TCS, however, is contingent on research that supports its validity. More specifically, the validity of the TCS Cognitive Skills Index (CSI) as an estimate of cognitive ability requires investigations that compare the CSI to an established criterion.

The assessment of intelligence through group tests has become a routine practice in the public schools. Although test experts (e.g., Anastasi, 1982; Cunningham, 1986; Salvia & Ysseldyke, 1988; 1991) question the validity of group tests, a majority of classroom teachers hold that group test scores have substantial value when identifying individual student needs for instruction, grouping students of similar ability for instructional purposes, and referring students who may be in need of special education services (Dusek & Joseph, 1985; Fields & Kumar, 1982). Furthermore, teachers' knowledge of ability scores significantly influences their acceptance of behavioral intervention (Martens & Meller, 1989), judgment of children from different social classes (McCombs & Gay, 1988), and perception of students with respect to both emotional and behavioral attributes (Prawat & Jarvis, 1980). Thus, the score from a group ability test has become a critical component in the instruction and treatment of an individual child in many schools.

School psychologists, in contrast to classroom teachers, rarely consider group test results in addressing referral concerns for individual students (Elliott, Piersel, & Galvin, 1983). In addition to concurring with issues raised by test experts regarding the general utility of group ability tests, school psychologists have also voiced more specific validity concerns about using group tests for screening purposes (Hartsough, Elias, & Wheeler, 1983).

While generally disregarding group test data, school psychologists rely heavily on the use of individual assessment instruments to estimate cognitive abilities. Moreover, federal and state laws also require school psychologists to use individual instruments for both initial assessment and three-year, re-evaluations (e.g., IDEA, 1990); these mandates are based, in part, on legal challenges to the validity of group tests and the prior misuse of group test results to place students in separate educational programs (e.g., Hobson v. Hanson, 1967).

As a consequence of their professional preference and in compliance with special education law, school psychologists devote a substantial portion of time to administering and interpreting individual measures of intelligence. Surveys indicate that school psychologists administer an average of seventeen intelligence tests each month, or about one test each workday (Reschly, Genshaft, & Binder, 1987). If the time consumed by individual intelligence assessment were reduced, school psychologists could deliver a wider range of services to their schools, such as counseling and consultation. If valid, group test scores may contribute evaluation data, thereby potentially providing the school psychologists the opportunity to expand their professional practices.

For both classroom teachers and school psychologists, the valid utility of group ability tests is contingent upon "the appropriateness of the inferences that can be made on

the basis of test results" (Salvia & Ysseldyke, 1991, p. 145). The appropriateness of such inferences for all students must be founded on the results of studies that examine the criterion-related validity, both concurrent and predictive, of group measures. Research must also investigate the validity of group ability tests for students whose learning difficulties (e.g., in reading and/or concentration) may compromise their test performance (Anastasi, 1982; Cunningham, 1986; Salvia & Ysseldyke, 1988; 1991).

The purpose of the present study was to investigate the criterion-related validity of the group administered Test of Cognitive Skills (TCS) (CTB/McGraw-Hill, 1981). For students with learning disabilities, the TCS and the individually administered Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974) were compared to examine: (a) the differences between standard score means of the TCS Cognitive Skills Index (CSI) and each of the WISC-R IQs and factor scores, (b) the relationships between standard scores of the TCS CSI and each of the WISC-R IQs and factor scores, and (c) the range of confidence within which the WISC-R Full Scale IQ can be estimated from the CSI. Post-hoc analyses were also conducted to study CSI-FSIQ differences and to examine subject characteristics that may be associated with TCS performance.

Statement of Procedures

Data collection. Data collection was conducted by the researcher. A list of students receiving learning disability services was compiled from the classroom rosters of special education teachers throughout the school corporation. For each student, a data sheet was completed, and all personal identification information was omitted.

Results of the TCS were located in individual cumulative folders which are housed in students' home schools. All data were photocopied at the school, and folders were returned to the school files in the school's main or guidance office.

Special education files, housed in the main special services office, were examined to determine the date of initial placement, to verify both learning disability eligibility and placement decisions, and to gain initial assessment results for individual students.

Subject selection. Subjects of the study were 118 public school students (males = 86; females = 32), in first through eighth grade (M = 3.68), ages 81 to 162 months (M = 116.96), who received learning disability special education services. All students resided in a single school corporation within the midwest region of the country. Subjects for the study met the following selection criteria: (a) placement in learning disability special services as determined by the Initial Case Conference Committee, (b) a

CSI score obtained from TCS Levels 2-5, and (c) WISC-R results obtained within 3 years of the TCS administration.

Design and Statistical Treatment of Data. Computer analyses of data for the study were conducted by the university computer center through the use of the Statistical Analysis System (SAS) (SAS Institute, 1982). Differences between standard score means from the TCS and WISC-R were examined through two-tailed, t-tests for correlated samples. The degree of relationship between standard scores for each CSI and WISC-R pair was assessed by Pearson product-moment correlation coefficients which were tested for significance. The range of confidence within which the WISC-R Full Scale IQ may be estimated from the CSI was determined through the application of a linear regression equation.

Several post-hoc analyses were implemented to examine individual CSI-FSIQ differences and to investigate the extent to which CSI test performance may be associated with reading and/or concentration skills. In order to examine individual CSI-FSIQ differences, data were organized and analyzed according to: (a) absolute differences between the CSI and WISC-R FSIQ for individual students, and (b) conventional WISC-R categories of ability levels. With regard to factors that may be related to CSI test performance, two-tailed, t-tests for independent samples were conducted to compare the CSI means of: (a) students with average FD scores to students with below average FD

scores, and (b) students with average reading achievement scores to students with below average reading achievement scores.

Research Questions

The following research questions were investigated:

Question 1. For students with learning disabilities, are there significant differences between the standard score means of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Verbal Comprehension factor (VC), Perceptual Organization factor (PO), and Freedom from Distractibility (FD) factor?

Question 2. For students with learning disabilities, do significant relationships exist between the standard scores of the TCS Cognitive Skills Index (CSI) and the WISC-R Full Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Verbal Comprehension factor (VC), Perceptual Organization factor (PO), and Freedom from Distractibility (FD) factor? Given a significant relationship, what is the range of confidence within which an individual's FSIQ may be estimated from the CSI?

Discussion

Data analyses conducted to address the first research question revealed that differences between the standard score means of the TCS CSI and the WISC-R FSIQ, VIQ, PIQ, VC, PO, and FD were significant for students with learning disabilities. The standard score mean of the TCS CSI was

significantly lower than the standard score means of the WISC-R FSIQ, VIQ, PIQ, VC, and PO. On the other hand, the standard score mean of the TCS CSI was significantly higher than the standard score mean of the WISC-R FD.

These results indicate that the TCS significantly underestimated the cognitive abilities assessed by the WISC-R; further, the CSI underestimated the general cognitive abilities (FSIQ) of students with learning disabilities as well as their verbal (VIQ, VC) and performance/perceptual (PIQ, PO) abilities. In contrast, although the subject group demonstrated a significant weakness in the abilities assessed by the WISC-R FD (e.g., the ability to focus and sustain attention), this weakness was not demonstrated by the TCS CSI.

The significant difference between the CSI and FSIQ means for this sample of students with learning disabilities is consistent with prior studies that have examined the valid use of the TCS with other special populations. For example, CSI and FSIQ standard score means were also significantly different for students identified as gifted (Robinson and Nagle, 1992). Gifted students, however, scored higher on the CSI than the WISC-R. This contrast in score comparisons for subject groups has also been found in comparisons of the TCS and the Stanford-Binet: Fourth Edition (SB:FE) (Thorndike, Hagen, & Sattler, 1986a); special education students scored significantly lower on the TCS CSI than the SB:FE Composite (Blood, 1989), whereas

gifted students scored significantly higher on the TCS CSI than the SB:FE Composite (Robinson and Nagle, 1992).

The investigation of the second research question yielded significant, positive relationships between scores of the TCS CSI and each of the WISC-R scores (FSIQ, VIQ, PIQ, VC, PO, FD) for students with learning disabilities. The strongest relationship was found between the CSI and FSIQ with approximately 41% of shared variance between the two tests. The CSI and FD shared 18% of variance, the weakest relationship among the test scores.

In a prior study the CSI was also significantly and positively related to the FSIQ and VIQ for gifted children (Robinson & Nagle, 1992); however, for gifted children, the CSI-PIQ relationship was not significant. Further, although the present study demonstrated that the CSI and FSIQ shared about 41% of variance for students with learning disabilities, Robinson and Nagle reported that the CSI and FSIQ shared only about 17% of variance for gifted students. Similar results have also been reported in validity studies comparing the CSI to the SB:FE Composite. According to Blood (1989), for example, the CSI-Composite relationship for special education students was significant with approximately 49% shared variance. Robinson and Nagle (1992) also found a significant relationship between the CSI and SB:FE Composite for gifted students; shared variance between measures was about 25%.

Results of the present investigation of the relationship between the TCS and the WISC-R compare favorably with validity research with the Otis-Lennon School Ability Test (OLSAT) and the (Otis & Lennon, 1982) WISC-R. For students with learning disabilities, Avant and O'Neal (1982) reported a significant relationship between the OLSAT composite (DIQ) and the FSIQ. They also found a significant DIQ-FSIQ relationship for students in mildly mentally handicapped placement and for students who were referred but determined ineligible for special education. In contrast to the significant CSI-FSIQ relationship reported by Robinson and Nagle (1992) for gifted students, Avant and O'Neal found that the DIQ-FSIQ relationship was not significant for their sample of gifted students.

Given the significant relationship between the CSI and FSIQ, the second research question was further examined through the computation of a linear regression equation. Applying the equation resulted in substantial error in estimating the FSIQ from a given CSI for an individual student. Therefore, the confidence range of scores within which the FSIQ can be estimated from the CSI at an acceptable level of probability was also substantial. For example, a range of 29 standard score points was required to estimate a student's FSIQ within a 90% probability of accuracy. Estimating the FSIQ within a 95% and 99% probability of accuracy required confidence ranges of 33 and 43 points, respectively. Thus, although the relationship

between the CSI and FSIQ was statistically significant, the CSI accounted for less than one-half of the FSIQ variance. One consequence of the 59% of FSIQ variance not explained by the CSI was the extreme range of scores required to estimate the FSIQ from the CSI at an acceptable level of confidence for students with learning disabilities.

Wright and Piersel (1987) also found that a wide range of scores was required for nonreferred students in order to estimate the WISC-R FSIQ from another group ability test, Educational Ability Series (EAS). In this case, estimating a FSIQ from a given EAS at a 90% probability of accuracy required a confidence band of $\pm$ 17, or a range of 35 standard score points. Although the 90% confidence level for estimating the FSIQ from the CSI was somewhat narrower at 29 points, the standard errors for both the EAS and TCS are unacceptable for estimating the WISC-R FSIQ.

Through post-hoc analyses, CSI-FSIQ differences for individual students were examined by the present study. The distribution of absolute CSI-FSIQ differences revealed that 53% of total sample had CSI-FSIQ differences of 5 points or less, 25% had differences of 6 to 10 points, and 22% had differences greater than 11 points. In addition to students with learning disabilities, students identified as gifted showed a similar distribution of absolute CSI-FSIQ differences in a prior study (Robinson & Nagle, 1992): 44% had differences of 5 points or less, 28% had differences of 6 to 10 points, and 28% had differences exceeding 11 points.

Results revealed a direct relationship between the size and direction of CSI-FSIQ standard score differences for individual students; that is, with increasing score differences, subjects showed a strong tendency to score lower on the CSI. For example, 53% of the subjects had CSI-FSIQ differences of 5 points or less with an approximately even distribution between positive and negative differences. For subjects with larger CSI-FSIQ differences of 6 to 10 points (25% of the total sample), a majority (62%) scored lower on the CSI. The proportion of negative CSI-FSIQ differences continued to increase as 78% of students with differences exceeding 10 points (23% of the total sample) scored lower on the CSI.

Review of the total distribution of CSI-FSIQ differences indicated that about 1 of every 10 (11.02%) students scored 16 or more points lower on the CSI than the WISC-R Full Scale IQ; 18% scored 11 or more points lower on the CSI than the WISC-R Full Scale IQ. On the other hand, 1 student in the total sample scored 16 points or more higher on the CSI, and 6 (5%) students scored at least 11 points higher on the CSI.

A distribution of absolute CSI-FSIQ differences was also constructed according to 10-point ranges for FSIQs from 70 to 129 in order to determine the extent to which, if any, standard score differences may be attributable to the tendency of scores close to the mean on one variable to yield estimated scores close to the mean on the second

variable. Rather than a regression to the mean effect, a relatively consistent distribution of CSI-FSIQ differences was observed across all ability ranges. Furthermore, comparisons across ranges revealed comparable proportions of differences greater than 5 points and differences 5 points or less. Consistent with the proportions of other FSIQ ranges, for example, the average range (90-109) (n = 68) was composed of 32 (47%) CSI-FSIQ differences exceeding 5 points and 36 (53%) differences of 5 points or less. Moreover, within this average ability group, CSI-FSIQ differences ranged widely from -28 to +13 standard score points. Thus, the observed differences between the CSI and FSIQ scores were not a primary result of a simple regression to the mean.

The validity of using the CSI to group or classify students was investigated by comparing the classifications of subjects in conventional ability categories according to their CSIs and FSIQs. Results revealed that, relative to their FSIQ classifications, 47% of students would be reclassified if CSIs were used instead of FSIQs; 66% of these students would be reclassified in lower ability groups. Thus, decisions for and attitudes towards an individual student may be vastly different when they are based on TCS or WISC-R results.

Post-hoc analyses also yielded two additional findings with regard to subject characteristics that may be associated with CSI outcomes. Inasmuch as the FD mean was

the only WISC-R score that was significantly lower than the TCS CSI mean, a closer inspection of differences between CSI means for subjects with average and below average FD standard scores was conducted. Data analyses indicated that students with average FD factors scored significantly higher on the CSI than students with below average FD factors. Given that the verbal comprehension abilities were equally well developed (and within the average range) for both FD groups, results suggest that the CSI performance of students with learning disabilties may be sensitive to the abilities operationally defined by the FD. This finding is bolstered by observing that the CSI mean of the average FD subject group not only exceeded the CSI mean for the total study sample, but also approached the CSI mean for TCS standardization group.

The extent to which reading skills levels may be associated with differential CSI performance was also investigated through post-hoc analyses. Results indicated that students with average reading skills scored significantly higher on the CSI than students with below average reading skills. In contrast, the two groups did not differ in their general cognitive abilities as estimated by the WISC-R FSIQ. These findings, therefore, add empirical support to the concerns of test experts, such as those expressed by Anastasi (1989), who caution that group tests may underestimate the true cognitive abilities of students with concentration and/or reading difficulties.

Implications and Recommendations

The integrated results of the investigation indicate that the CSI cannot be recommended as an estimate of the cognitive abilities, as measured by the WISC-R, for students with learning disabilities. The WISC-R FSIQ could not be estimated from the CSI within an acceptable range of accuracy. Moreover, the CSI significantly underestimated the cognitive abilities of many students in the sample; about 1 out of every 10 subjects scored at least 16 points lower on the CSI than the FSIQ.

Thus, a major implication of the study is that the TCS is inappropriate for the screening of students with learning disabilities; indeed, given that the CSI tends to underestimate their cognitive abilities, the TCS is likely to screen out many students with learning disabilities. Moreover, teachers should be wary in using the CSI to address their concerns for individual students who demonstrate learning difficulties in the regular classroom. Caution in using the CSI should be exercised particularly when teachers suspect that students may have a learning disability; for example, a lower than expected CSI should not deter teachers from considering a referral for individual evaluation.

Recommendations for Future Studies

1. The extent to which the CSI may be a valid index of cognitive ability for nonreferred children is unknown; thus, studies that compare the CSI to an established criterion are

strongly recommended to examine the validity of the TCS. In the absence of validity data for a nonreferred sample, and in light of research results with special populations, the validity of the CSI as a measure of "academic aptitude" (TCS Handbook, 1982, p. 1) remains unknown. Thus, the use of the CSI for grouping students of "similar ability" for instruction or for identifying individual students needs for instruction also appears to be questionable at this time.

2. Reliability studies are also imperative to examine the validity of the TCS. Review of test manuals and research reports indicates that test-retest reliability for the CSI has not been established. Further, the reliability of CSI scores from level to level requires investigation. The present study, for example, excluded the Level 1 test due to its distinct difference from higher test levels. Further, preliminary review of data for the subjects of the present study suggests considerable variability among level scores for individual students. Test levels also overlap children in grades 3, 5, 7, and 9; thus score comparisons between these levels should also be conducted to investigate TCS reliability.

3. A productive line of inquiry may be developed by reconceptualizing the CSI as an index of academic achievement (actual) rather than an estimate of academic ability (expected). That is, although the TCS subtests are not intended to assess achievement, the skills required for classroom success (e.g., following directions, attending and

concentrating, remembering, completing paper-pencil tasks, reading) may also be required for TCS success. On the other hand, conditions such as hearing or visual weaknesses, attention difficulties, learning disabilities, socioemotional problems (e.g., anxiety, depression), and/or visual-motor difficulties are likely to interfere with both TCS and classroom performance, thereby resulting in similar achievement data.

Thus, studies that compare that CSI to achievement scores for both nonreferred students and special populations may yield results that support both the criterion-related and construct validity of the TCS. For example, given that the CSI is operationally defined as a measure of achievement, one appropriate inference may be that students who score below the average TCS range merit further attention. Similarly, this revised understanding of the CSI as an index of achievement rather than aptitude/ability may promote the valid use of the TCS in screening or identifying students for gifted programs.

4. Finally, studies should be conducted to compare the TCS with the WISC-III (Wechsler, 1991). The WISC-R was selected for the present investigation inasmuch as a plethora of studies provide a firm foundation for the WISC-R as a criterion measure. With its revised format, materials, Symbol Search subtest, and factor structure, the WISC-III may yield somewhat different results from those of its predecessor.

REFERENCES

Altpeter, T., & Handal, P. J. (1986). Use of the PPVT-R for intellectual screening with school-aged children: A caution. Journal of Psychoeducational Assessment, 4, 145-154.

Anastasi, A. (1982). Psychological testing. NY: Macmillan Publishing.

Arinoldo, C. G. (1982). Concurrent validity of McCarthy's Scales. Perceptual and Motor Skills, 54, 1343-1346.

Avant, A. H., & O'Neal, M. R. (1986, November). Investigation of the Otis-Lennon School Ability Test to predict WISC-R Full Scale IQ for referred children. Paper presented at the annual meeting of the Mid-South Educational Research Association. Memphic, TN. (ERIC Document Reproduction Service No. ED 286 883)

Bersoff, D. N., & Hofer, P. T. (1990). The legal regulation of school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (pp. 937-961). New York: Wiley.

Blood, B. A. (1989). The relationship between achievement on the Test of Cognitive Skills and the Stanford-Binet Intelligence Scale: Fourth Edition for elementary school students. (Doctoral dissertation, Ball State University, 1989). Dissertation Abstracts International.

Bracken. B. A., Prasse, D. P., & Breen, M. J. (1984). Concurrent validity of the Woodcock Johnson Psycho-Educational Battery with regular and learning disabled students. Journal of School Psychology, 22, 185-192.

Breen, M. J. (1981). Comparison of the Wechsler Intelligence Scale for Children--Revised and Peabody Picture Vocabulary Test--Revised for a referred population. Psychological Reports, 49, 717-718.

Breen, H. J., & Siewert, J. C. (1983). Comparison of the Peabody Picture Vocabulary Test--Revised and the Wechsler Intelligence Scale for Children--Revised for learning disabled and referred students. Journal of Psychoeducational Assessment, 1, 95-99.

Brown v. Board of Education, 347 U.S. 483 (1954).

Carvajal, H., & Weyand, K. (1986). Relationships between scores on Stanford-Binet IV and Wechsler Intelligence Scale for Children-Revised. Psychological Reports, 59, 963-966.

Coleman, M. C., & Harmer, W. R. (1985). The WISC-R and Woodcock-Johnson Tests of Cognitive Ability: A comparative study. Psychology in the Schools, 22, 127-132.

Covin, T. M. (1976). Comparison of Otis-Lennon Mental Ability Test, Elementary I Level and WISC-R IQs among suspected mental retardates. Psychological Reports, 38, 403-406.

CTB/McGraw-Hill. (1980). California Achievement Test. Monterey, CA: Author.

CTB/McGraw-Hill. (1981). Comprehensive Test of Basic Skills. Monterey, CA: Author.

CTB/McGraw-Hill. (1981). Test of Cognitive Skills. Monterey, CA: Author.

CTB/McGraw-Hill. (1982). Test of Cognitive Skills: Test coordinator's handbook and guide to interpretation. Monterey, CA: Author

CTB/McGraw-Hill. (1983). Test of Cognitive Skills: Technical report. Monterey, CA: Author

Cunningham, G. K. (1986). Educational and psychological measurement. NY: Macmillan Publishing Company.

Davis, S. E., & Kramer, J. J. (1985). Comparison of the PPVT-R and WISC-R: A validation study with second-grade students. Psychology in the Schools, 22, 265-268.

Dunn, L. M., & Dunn, L. M. (1981). Peabody Picture Vocabulary Test--Revised. Circle Pines, MN: American Guidance Service.

Dusek, J. B., & Joseph, G. (1985). The bases of teacher expectancies. In J. B. Dusek, V. C. Hall, & Meyer, W. J. (Eds.), Teacher expectancies (pp. 229-250). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Elliott, S. N., Piersel, W. C., & Galvin, G. A. (1983). Psychological re-evaluations: A survey of practices

and perceptions of school psychologists. Journal of School Psychology, 21, 99-108.

Estabrook, G. E. (1984). A canonical correlation analysis of the Wechsler Intelligence Scale for Children--Revised and the Woodcock Johnson Tests of Cognitive Ability in a sample referred for suspected learning disabilities. Journal of Educational Psychology, 76, 1170-1177.

Fields, J. P., & Kumar, V. K. (1982). How teachers use the group IQ test scores. Journal of School Psychology, 20, 32-38.

French, J. L., & Hale, R. L. (1990). A history of the development of psychological and educational testing. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Intelligence and achievement (pp. 3-28). NY: The Guilford Press.

Hartsough, C. S., Elias, P., & Wheeler, P. (1983). Evaluation of a non-intellectual assessment procedure for early screening of exceptionality. Journal of School Psychology, 21, 133-142.

Haywood, H. C., Brown, A. L., & Wingenfeld, S. (1990). Dynamic approaches to psychoeducational assessment. School Psychology Review, 19, 411-422.

Hobson v. Hansen, 269 F Supp 401 (D. D.C. 1967).

Hollinger, C. L. , & Sarvis, P. H. (1984). Interpretation of the PPVT-R: A pure measure of verbal comprehension? Psychology in the Schools, 21, 34-41.

Hunter, J. A., & Lowe, J. D. (1980). The use of the WISC-R, Otis, Iowa, and SRBCSS in identifying gifted elementary students. The Southern Journal of Eductional Research, 12, 59-65.

Hynd, G. W., Quackenbush, R., Kramer, R., Conner, R., & Weed, W. (1980). Concurrent validity of the McCarthy Scales of Children's Ability with Native American primary-grade children. Measurement and Evaluation in Guidance, 13, 29-34.

Indiana Learning Disabilities Task Force. (1982). Learning Disability Definition and Eligibility Criteria: Guidelines for Implementation. Indianapolis: Indiana State Department of Education.

Indiana Learning Disabilities Task Force. (1987). Questions and Responses to Definition and Eligibility Criteria:

Guidelines for Implementation. Indianapolis: Indiana State Department of Education.

Indiana State Board of Education. (1988). Rules, Regulations and Policies Adopted and Promulgated by the Indiana State Board of Education State of Indiana (Rule S-1), 511 IAC 7-1. Indianapolis: Author.

Indiana State Board of Education. (1992). Indiana Special Education Rules (Article 7, Rules 3-16), Title 511. Indianapolis: Author.

Individuals with Disabilities Education Act (IDEA), P. L. 101-476. (1990). Washington, DC: U.S. Government Printing Office.

Jastak, J. F., & Wilkinson, G S. (1984). Wide Range Achievement Test--Revised. Wilmington, DE: Jastak Associates.

Karnes, F. A., & Lee, L. A. (1984). Correlations of scores on the Developing Cognitive Abilities Test and the Wechsler Intelligence Scale for Children-Revised for intellectually gifted students. Psychological Reports, 54, 373-374.

Kaufman, A. S., Harrison, P. L., & Ittenbach, R. F. (1990). Intelligence testing in the schools. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology (pp. 289-327). New York: Wiley.

Kaufman, A. S., & Kaufman, N. L. (1983). Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

Keith, T. Z. (1985). Test of Cognitive Skills. In J. V. Mitchell, Jr. (Ed.). The ninth mental measurements yearbook. (Vol. 2, pp. 1555-1556). Lincoln, NE: University of Nebraska.

Klanderman, J., Devine, J., & Mollner, C. (1985). The K-ABC: A construct validity study with the WISC-R and Stanford-Binet. Journal of Clinical Psychology, 41, 273-281

McCallum, R. S., Karnes, F. A., & Edwards, R. P. (1984). The test of choice for assessment of gifted children: A comparison of the K-ABC, WISC-R, and Stanford-Binet. Journal of Psychoeducational Assessment, 2, 57-63.

McCarthy, D. (1972). The McCarthy Scales of Children's Abilities. NY: Psychological Corporation.

McCombs, R. C., and Gay, J. (1988). Effects of race, class, and IQ information on judgments of parochial grade school teachers. Journal of Social Psychology, 128, 647-652.

McGrew, K. S. (1983). Comparison of the WISC-R and Woodcock-Johnson tests of cognitive ability. Journal of School Psychology, 21, 271-276.

Martens, B. K. , Meller, P. J. (1989). Influence of child and classroom characteristics on acceptability of interventions. Journal of School Psychology, 27, 237-245.

Naglieri, J. A. (1984). Concurrent and predictive validity of the Kaufman Assessment Battery for children with a Navajo sample. Journal of School Psychology, 22, 373-379.

Naglieri, J. A. (1985). Use of the WISC-R and K-ABC with learning disabled, borderline mentally retarded, and normal children. Psychology in the Schools, 22, 133-141.

Naglieri, J. A., Das, J. P., & Jarman, R. F. (1990). Planning, attention, simultaneous, and successive, cognitive processes as a model for assessment. School Psychology Review, 19, 423-442.

Naglieri, J. A., & Haddad, F. A. (1984). Learning disabled children's performance on the Kaufman Assessment Battery for children: A concurrent validity study. Journal of Psychoeducational Assessment, 2, 49-56.

Nasland, R. A., Thorpe, L. P., & Lefever, D. W. (1978). SRA Achievement series. Chicago: Science Research Associates.

Obrzut, A., Obrzut, J. E. , & Shaw, D. (1984). Construct validity of the Kaufman Assessment Battery for Children with learning disabled and mentally retarded. Psychology in the Schools, 21, 417-424.

Otis, A. S. (1929). Otis Self-Administering Tests of Mental Ability. NY: World Book Co.

Otis, A. S. (1954). Otis Quick-Scoring Mental Ability Tests, New Edition. NY: World Book Co.

Otis, A. S., & Lennon, R. T. (1969). Otis-Lennon Mental Ability Test, Elementary Level I. NY: Harcourt Brace Jovanovich, Inc.

Otis, A. S., & Lennon, R. T. (1969). Otis-Lennon Mental Ability Test, Elementary, Level II. NY: Harcourt Brace Jovanovich, Inc.

Otis, A. S., & Lennon, R. T. (1982). Otis-Lennon School Ability Test. NY: The Psychological Corporation.

Pommer, L. T. (1986). Seriously emotionally disturbed children's performance on the Kaufman Assessment Battery for Children: A concurrent validity study. Journal of Psychoeducational Assessment, 4, 155-162.

Prawat, R. S., & Jarvis, R. (1980). Gender difference as a factor in teachers' perceptions of students. Journal of Educational Psychology, 72, 743-749.

Psychological Corporation. (1983). Basic Achievement Skills Individual Screener. San Antonia, TX: Author.

Rasbury, W. C., Falgout, J. C., & Perry, N. W. (1978). A Yudin-type form of the WISC-R: Two aspects of validation. Journal of Clinical Psychology, 34, 120-126.

Reilly, T. P., Drudge, O. W., Rosen, J. C., Loew, D. G., & Fischer, M. (1985). Concurrent and predictive validity of the WISC-R, McCarthy Scales, Woodcock-Johnson, and academic achievement. Psychology in the Schools, 22, 380-382.

Renzulli, J. S., & Hartman, R. K. (1971). Out of the classroom: Scale for rating behavioral characteristics of superior students. Exceptional Children, 38, 243-248.

Reschly, D. J., Genshaft, J., & Binder, M. S. (1987). The NASP survey: Comparison of practitioners, NASP leadership, and university faculty on key issues. Washington, DC: National Association of School Psychologists.

Reschly, D. J., & Grimes, J. P. (1990). Best practices in intellectual assessment. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology--II (pp. 425-439). Washington, DC: National Association of School Psychologists.

Reschly, D. J., & Wilson, M. S. (1990). Cognitive processing versus traditional intelligence: Diagnostic utility, intervention implications, and treatment validity. School Psychology Review, 19, 443-458.

Reynolds, C. R., & Kaufman, A. S. (1990). Assessment of children's intelligence with the Wechsler Intelligence

Scale for Children-Revised (WISC-R). In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Intelligence and achievement (pp. 127-165). NY: The Guilford Press.

Robertson, G. J. (1972). The development of the first Group Mental Ability Test. In G. H. Bracht & K. D. Hopkins (Eds.), Perspectives in educational and psychological measurement. Englewood Cliffs. NJ: Prentice-Hall.

Robinson, E. L., & Nagle, R. J. (1992). The comparability of the Test of Cognitive Skills with the Wechsler Intelligence Scale for Children-Revised and the Stanford-Binet: Fourth Edition with gifted children. Psychology in the Schools, 29, 107-112.

Robinson, N. M., & Janos, P. M. (1987). The contribution of intelligence tests to the understanding of special children. In J. D. Day & J. G. Borkowski (Eds.), Intelligence and exceptionality: New directions for theory, assessment, and instructional practices (pp. 21-56). Norwok, NJ: Ablex Publishing Corporation.

Rosso, M., Falasco, S. L., & Koller, J. R. (1984). Investigations into the relationship of the PPVT-R and the WISC-R with incarcerated delinquents. Journal of Clinical Psychology, 40, 588-591.

Rothlisberg, B. A. (1987). Comparing the Stanford-Binet: Fourth Edition to the WISC-R: A concurrent validity study. Journal of School Psychology, 25, 193-196.

Riverside Publishing Co. (1965). Iowa Test of Basic Skills. Princeton, NJ: Author.

Salvia, J., & Ysseldyke, J. E. (1988). Assessment in special and remedial education (4th ed.). Boston: Houghton Mifflin Company.

Salvia, J., & Ysseldyke, J. E. (1991). Assessment (5th ed.). Boston: Houghton Mifflin Company.

Sandoval, J., & Irvin, M. G. (1990). Legal and ethical issues in the assessment of children. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Intelligence and Achievement (pp. 86-104). NY: The Guilford Press.

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

Statistical Analysis System Institute. (1982). Statistical Analysis System. NY: Author.

Sternberg, R. J. (1985). Test of Cognitive Skills. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook. (Vol. 2, pp. 1556-1557). Lincoln, NE: University of Nebraska.

Sullivan, E. J., Clark, W. W., & Tiegs, E. W. (1974). Short Form Test of Academic Aptitude. Monterey, CA: CTB/McGraw-Hill.

Terman, L. M., & Merrill, M. A. (1960). Stanford-Binet Intelligence Scale. Boston: Houghton Mifflin.

Thompson, P. L., & Brassard, M. R. (1984). Validity of the Woodcock-Johnson Tests of Cognitive Ability: A comparison with the WISC-R in LD and normal elementary students. Journal of School Psychology, 22, 201-208.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986a). Stanford-Binet Intelligence Scale: Fourth Edition. Chicago: The Riverside Publishing Company.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986b). Stanford-Binet Intelligence Scale: Fourth Edition: Technical manual. Chicago: The Riverside Publishing Company.

Thurstone, T. (1978). Educational Ability Series. Chicago: Science Research Associates.

Troy, M. E. (1985). Test of Cognitive Skills. In D. J. Keyser,& R. C. Sweetland (Eds.). Test critiques (pp. 780-786). Kansas City, MO: Test Corporation of America.

Wechsler, D. (1949). Wechsler Intelligence Scale for Children. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1955). Wechsler Adult Intelligence Scale. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1967). Wechsler Preschool and Primary Scale of Intelligence. San Antonio: The Psychological Corporation.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children--Revised. NY: The Psychological Corporation.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children--3rd ed. San Antonio, TX: The Psychological Corporation.

Wick, J. W., Smith, J. K., Beggs, D. L., & Mouw, J. T. (1980). Developing Cognitive Abilities Test. Glenview, IL: Scott, Foresman.

Wikoff, R. L., & Parolini, R. J. (1982). Prediction of the WISC-R Full Scale IQ from the SFTAA. Journal of Clinical Psychology, 38, 387-388.

Woodcock, R. W., & Johnson, M. B. (1977). Woodcock-Johnson Psycho-Educational Battery. Boston: DLM Teaching Resources.

Woodcock, R. W., & Johnson, M. B. (1988). Woodcock-Johnson Psycho-Educational Battery--Revised. Boston: DLM Teaching Resources.

Worthing, R. J., Phye, G. D. & Nunn, G. D. (1984). Equivalence and concurrent validity of PPVT-R forms L and M for school-age children with special needs. Psychology in the Schools, 21, 296-299.

Wright, D., & Piersel, W. C. (1987). Usefulness of a group-administered ability test for decision making. Journal of School Psychology, 25, 63-71.

Zins, J. E., & Barnett, D. W. (1984). A validity study of the K-ABC, the WISC-R, and the Stanford-Binet with nonreferred children. Journal of School Psychology, 22, 369-371.