Running Head: VALIDITY OF CSI

Validity of the Cognitive Skills Index On the Test of Cognitive Skills-Second Edition as a Measure of Intelligence with Special Education Children

Nicole C. LaPearl

Salisbury University

Submitted in partial fulfillment of requirements for the degree of

Master of Arts in Psychology

Salisbury UNIVERSITY

PSYCHOLOGY DEPARTMENT
1101 Camden Avenue
Salisbury, MD 21801-6860
410-543-6370

1-888-543-0148
TTY 410-543-6080
FAX 410-548-2056
www.salisbury.edu

COMPLETION OF THESIS

MEMO TO DIRECTOR OF GRADUATE STUDIES:

This is to certify that on May 12, 2003, Nicole C. LaPearl successfully completed the oral defense of her Thesis entitled Validity of the Cognitive Skills Index-Second Edition on the Test of Cognitive Skills as a Measure of Intelligence with Special Education Children presented in partial fulfillment of the requirements for the Master of Arts degree in Psychology.

Clarisse Chappell, Ph.D.
Thesis Committee Chairman

Thesis Committee Member

Thesis Committee Member

TABLE OF CONTENTS

List of Tables

Tables

Acknowledgments

In completing this research, I wish to express my appreciation to several individuals. Dr. Charisse Chappell, my committee chair, has provided valuable insight and help in preparing this document as well as direction in the actual data collection process. Committee members, Dr. Natalia Lion and Dr. Robert Graff, have also provided valuable assistance in this process. I would also like to thank Dr. George Stone and others in the school district for allowing me to access records in order to complete this project.

A special thank you to my family. My husband, JR, and children, Zachary and Maddison, have been a constant source of encouragement as this process has unfolded. My parents, Bruce and Barbara Culver, have also given me that gentle prodding needed to keep me going.

ABSTRACT

For many elementary and secondary school students, a standardized group ability test score may be found in their cumulative record. Such results are often used for placement in instructional programs. However, with relatively little research and conflicting results, it remains unclear if group ability test scores are appropriate to be used for any purpose. This study investigated the validity of using the Test of Cognitive Skills-Second Edition (TCS/2) as a measure of intelligence with special education students, as compared to the Wechsler Intelligence Scale for Children-Third Edition (WISC-III). To investigate the relationship between the group-administered TCS/2 and the individually administered WISC-III, scores of 66 students (grades 6 through 11) were compared. Pearson product-moment correlations revealed significant, positive relationships between the CSI and most scores of the WISC-III. However, despite significant relationships, the correlations are modest at best. T-tests revealed significant differences between the CSI and all WISC-III score means. CSI scores were significantly lower than WISC-III scores. Therefore, the CSI is not recommended as a valid source of information to be used in decision making regarding special education students. Further research is needed to determine the validity of the CSI with other student populations.

VALIDITY OF THE COGNITIVE SKILLS INDEX ON THE TEST OF COGNITIVE SKILLS-SECOND EDITION AS A MEASURE OF INTELLIGENCE WITH SPECIAL EDUCATION CHILDREN

Chapter I

Introduction

Test taking is a common facet of a child's educational experience in today's school environment. Various styles and types of tests are administered, usually either in a group setting or on an individual basis. One such type of test is an ability or intelligence test. Group ability or IQ tests, in addition to standardized achievement tests, are often administered by classroom teachers; and utilized by many school districts to both (a) determine students' readiness for various levels of instruction, and (b) to assist in the development of individualized instruction (Fields & Kumar, 1982). Individually administered intelligence tests are most often administered by school psychologists, when given in the academic setting, and are usually given for the purpose of qualifying students for special programs, most often special education. The question remains: Are group tests of ability an accurate measure of a student's intellectual ability as compared to the more well accepted individual measures?

Ability tests, also referred to as intelligence or IQ tests, are generally considered to be an estimate of scholastic aptitude or readiness to master a school curriculum. Such assessments often measure an individuals verbal reasoning skills (including verbal comprehension and the processing of language), nonverbal reasoning skills (including perceptual organization and visual-motor development), and memory. Ability tests stress the ability to apply information in new and different ways, and assess the amount of learning that occurs in a wide variety of life experiences. Ability tests are considered to be more valid measures of future performance than are achievement tests. Achievement tests assess more specific skills (i.e. reading and math), are heavily dependent on formal learning acquired in school or at home, and stress the mastery of factual information (Sattler, 1992).

Administration of ability and achievement tests can occur in either a group or an individual setting. Group-administered tests are administered to large numbers of individuals simultaneously. Paper-pencil formats and machine scorable answer sheets are common features of this administration format. Conversely, individually administered tests are given in a one-on-one setting. Those professionals administering tests in this one-on-one setting are most often well trained in both the procedures of test administration and the methods of test scoring. Requirements such as manipulating materials (i.e. blocks), timing of speed of performance, and oral presentation of words, sentences, or numbers while observing the performance to score it; makes intense familiarity with the tests a necessity. Unlike group-administered tests, individually administered tests do not require as much reading by the students because instructions and examples are presented orally. Significant advantages of individually administered tests include the child-tester interaction afforded in the one-to-one environment which may facilitate maintenance of a child's attention, and the tester's monitoring of the child's motivational levels and test behaviors (Walsh & Betz, 2001).

Scores gained from group intelligence tests are commonly reported in the cumulative files for individual students in the public school system. However, the purpose and use of these retained scores remain controversial. Although it is generally claimed that these tests are used to determine student readiness for and to determine the need for and type of individualized instruction, it is not known what specific information the group IQ test score actually provides (Fields & Kumar, 1982).

Group-administered tests have generally been criticized on a number of grounds (Wright & Piersel, 1987). Such criticisms have included: (a) tests generally utilize only one response mode, (b) the amount of qualitative data that can be gathered is limited, and (c) limitations of standardization samples and procedures. Group-administered tests generally utilize a response mode that is easily scored by machine, such as multiple choice. The problem with a multiple-choice test is that it only requires students to recognize the correct answer. Unlike an individually administered test, the nature of group-administered tests inhibits the amount of qualitative data that can be gathered. An examiner is unable to determine motivational factors, fatigue, anxiety, and other qualitative information during group-administered tests. Standardization samples of group-administered tests are not as representative or randomly selected as those used for individually administered tests. Group-administered tests are not standardized on the populations that are most likely to be in need of assessment. Also, group-administered ability tests are usually standardized by grade level, rather than age. Therefore, students in specific grades (which can include a very broad age range) take the same test and are grouped together for purposes of generating norms, rather than generating norms from very specific age groupings such as those used in the standardization of individually administered tests (Wright & Piersel, 1987).

Fields and Kumar (1982) attempted to find out exactly how the results of group IQ tests were utilized by classroom teachers. Ninety teachers, grades 2 through 6, were interviewed. Results of Fields and Kumar's study indicated that approximately 30% of teachers stated that they made little or no use of the group IQ test. The top reasons given for not using the results of the group IQ test were: tests were not fair, they preferred to make up their own test, or to go by what happens in the class. The majority of these teachers did not trust the reliability and validity of test scores. Some were also concerned that examining the test scores would affect their attitudes toward students; therefore they avoided the test scores.

Those teachers that stated they did use results of the group IQ test, indicated various reasons for doing so. The most popular use of the results was for discussion in parent conferences. Another frequent response was "to know a student's potential and/or determine ability-motivation discrepancy" (Fields & Kumar, 1982, p. 34). Using the test scores for referral purposes was another explanation given for using the IQ test scores. Approximately 33% actually stated that they used the scores for planning instruction.

Teachers who used the scores for developing instructional strategies reportedly treat high and low IQ scores very differently. With those students scoring high on the IQ tests, teachers stress intellectual work and assign challenging work to them. Those students with low IQ test scores are instructed using concrete examples or a step-by-step approach. The reasoning for these varied approaches appears to be that low IQ scores reflect difficulty with reasoning. The most frequent general response to how the scores are used for instructional planning was that it is used to plan instructional ability groups [i.e. high, high average, average, low average, and low]. These ability groups are most often utilized in the subjects of reading and math (Fields & Kumar, 1982).

According to Fields and Kumar (1982), teachers use group test results in a variety of ways, but how do they utilize individually administered test results? Generally, students are only administered an individually administered cognitive ability test, such as the WISC-III, if they have been referred as needing special services. Within the school system, school psychologists are the professionals licensed to conduct these evaluations. Following the evaluation, a meeting is held to discuss the results with parents and teachers to determine eligibility for programs. If eligibility is determined, those on the multidisciplinary team (including parents, teachers, and school psychologists) examine the strengths and weaknesses revealed to assist in developing an appropriate individualized educational program for that student. Teachers do not have these scores available for making general instructional decisions for all students, as these scores are not readily available as are group test scores.

Fields and Kumar were able to provide some insight as to how teachers choose to use or reasons for not using results from group IQ tests; however, it still remains a question how useful these results are for other professional personnel within the school system, particularly school psychologists. With the increasing utilization and acceptance of special education services, more strenuous demands have been placed on school psychologists. The school psychologist is required to conduct individual psychological examinations to assist in determining eligibility for those referred for special education services. A part of that evaluation process is usually an individually administered intelligence test, such as the Wechsler Intelligence Scale for Children-Third Edition (WISC-III). School psychologists rarely use results of group-administered tests of ability and achievement as a part of the evaluation process (Wright & Piersel, 1987). The continuous administration and interpretation of the individually administered WISC-III, takes an inordinate amount of time, leaving little time for them to pursue other intervention services (i.e. counseling). If results from the group-administered IQ tests (i.e. The Test of Cognitive Skills-Second Edition) were reliable and valid, could they be used in place of individually administered tests? Or might they be used as a screening device, to be more selective about who is administered an individual measure; therefore saving time, as results are already in student cumulative files?

Wright and Piersel (1987) conducted a study examining the usefulness of a group-administered ability test for decision making by educators. Group intelligence tests are generally used for one of two reasons. Often administered as a part of group achievement tests, they can function as a screening tool to identify students, who are sufficiently different (i.e. gifted, intellectually limited), warranting further assessment. The advantages of the group intelligence test include: it is administered quickly, inexpensive, and given to large numbers of students. A less frequent use of the group intelligence test is for use in making program decisions and instructional grouping assignments for individual students (Wright & Piersel, 1987).

Wright and Piersel (1987) investigated the relationship of the WISC-R and the

Educational Ability Series (EAS) [a group-administered ability test]. They found that "at predicting achievement, as measured by group or individual achievement tests or by teacher-assigned grades, the individually administered test (WISC-R) and the group-administered test (EAS) performed equally well" (p. 69). Although the two measures were comparable and had a shared variance of 53%, when used in making classification decisions based on ability scores, the WISC-R and the EAS could lead to very different decisions. The authors concluded that the EAS could not be recommended for use in classification decisions; however, as a general screening measure and its use in grouping students for instructional purposes, its use may be supported.

It was the purpose of this research to determine the validity of the Test of Cognitive Skills-Second Edition [TCS/2] (CTB/McGraw-Hill, 1993) as a measure of intelligence by comparing scores on the TCS/2 with scores on the individually administered Wechsler Intelligence Scale for Children-Third Edition [WISC-III] (Wechsler, 1991). No published research was found comparing these two newest versions of the respective tests and very little was found comparing the previous versions.

Test of Cognitive Skills-Second Edition

Overview. The Test of Cognitive Skills, Second Edition (TCS/2) is a group-administered cognitive abilities test designed to assess the academic aptitude of students in grades 2 through 12. Divided into four subtests, the TCS/2 is "intended to measure selected verbal, nonverbal, and memory abilities that can contribute greatly to students' success in an educational program" (CTB/McGraw-Hill, 1993, p.1).

For each subtest, scale scores are based on student performances on all the items in that test. Raw scores are converted to scale scores using conversion tables found in the *TCS/2 Norms* Book. The Total Test scale score is calculated by averaging the scale scores for the four subtests. The TCS/2 composite, the Cognitive Skills Index (CSI), is a normalized score with mean of 100 and a standard deviation of 16. It is generated from the total score using the distributions upon which the age percentile norms were based (CTB/McGraw-Hill, 1993).

Six levels of the TCS/2 assess students in grades 2 through 12:

Level 1	Grades 2 through 3
Level 2	Grades 4 through 5
Level 3	Grades 6 through 7
Level 4	Grades 8 through 9
Level 5	Grades 10 through 11
Level 6	Grades 11 through 12.

Each level includes four subtests: Sequences, Analogies, Memory, and Verbal Reasoning. The Sequences subtest is a nonverbal measure designed to assess the ability to comprehend a rule or principle implicit in a pattern or sequence of figures, letters, or numbers. Items include spatial relationships, ordered patterns, progressions, and combinations of parts that form a whole. Level 1 involves recognition of patterns or sequences of figures. Levels 2 through 6 look at letter and number patterns/sequences in addition to figures (CTB/McGraw-Hill, 1993).

The Analogies subtest is also nonverbal in nature and is designed to measure the ability to discern various relationships among picture pairs and then to infer parallel relationships between incomplete picture pairs. Items reflect such reasoning or problem-solving tasks as: comparing or contrasting, perception of the purpose or function of an object, understanding degree or proportion, and recognition of spatial relationships.

These types of reasoning abilities are important in reading comprehension and other basic skill areas (CTB/McGraw-Hill, 1993).

On the Memory subtest, the ability to recall previously presented material is measured. Recall of previously presented picture pairs comprise the items at Level 1. Levels 2 through 6 include items requiring recall of associations between nonsense words and their assigned definitions. The interval between the learning experience and the Memory test is approximately 15 minutes at Level 1 and 25 minutes for Levels 2 through 6 (CTB/McGraw-Hill, 1993).

Measurements of ability to solve verbal problems by reasoning deductively, analyzing category attributes, and discerning relationships and patterns are the purpose of the Verbal Reasoning subtest. Several item formats are used: identification of essential elements of objects or concepts, classification according to common attributes, inference of relationships between separate but related sets of words, and drawing logical conclusions from short passages (CTB/McGraw-Hill, 1993).

<ins>Revisions.</ins> The Test of Cognitive Skills, Second Edition [TCS/2] (CTB/McGraw-Hill, 1993) is a revision of the 1981 Test of Cognitive Skills (TCS). The original TCS was a revision of the Short-Form Test of Academic Aptitude (SFTAA). The TCS/2 kept the same basic format as the TCS, but substantial changes were made. Subtests, at all levels, contain all new items. The factor structure has been strengthened to yield scores for three cognitive factors – verbal, nonverbal, and memory ability. A new high school level, Level 6, was added to the TCS/2 despite its targeting the same overall range of students, grades 2 through 12, as the TCS. On the TCS/2, Memory items for Levels 2 through 6 consist of

nonsense words rather than obscure real words. This revision "makes the Memory subtest a better measure of recall ability because now students cannot answer items correctly by (1) using prior semantic knowledge, or (2) recognizing inflectional elements they learned independent of their exposure to the definitions in the subtests" (CTB/McGraw-Hill, 1993, p. 3). Verbal Reasoning items for Level 1 are now in textual formats like the other levels, as opposed to pictorial formats that were used in the 1981 edition. Despite the major revisions, the TCS/2 is still comparable in testing time to the TCS; it can normally be administered in less than an hour (CTB/McGraw-Hill, 1993).

Critique. Kamphaus, a professor of educational psychology, and author of numerous articles and texts on assessment, critiqued the TCS/2. Kamphaus' (1998) critique of the TCS/2 indicated that it was a carefully developed measure using state of the art test production methods. He found the administration procedures to be of the utmost quality, with clear and succinct guidelines. Another strength of the TCS/2, reported by Kamphaus, is the test stimuli. It is unambiguous and well organized in the response booklets. Kamphaus lauded the publishers on the unusual lengths to which they went to limit the influences of prior achievements on TCS/2 performance. Examples mentioned include: on the Memory subtest nonsense words are used as stimuli, on the Verbal Recognition subtest highly familiar English-language stimuli is used. He stated "the content blueprint for the four subtests was thoughtfully conceived to include only items that are conceptually consistent with the construct(s) measured by each subtest" (p.1027). Kamphaus found that the 150-page Technical Report was full of evidence of careful test development. TCS test development procedures benefit greatly from survey level achievement testing technologies. The items chosen made it through several screens before finally being selected. The screens included editorial reviews, extensive statistical and judgmental bias reviews, item factor analyses, distracter analyses, and item calibration using a three-parameter Item Response Theory (IRT) model. Item Response Theory is a statistical model that incorporates item difficulty, item discrimination, and student guessing in the development and selection of items. IRT is also applied to the computer scoring for the TCS and TCS/2 subtests for individual students. Kamphaus also praised the Technical Reports inclusion of often ignored issues such as the effects of speed (vs. power) on test performance, and the ability to differentiate samples of exceptional students.

Kamphaus (1998) raised concern about the precedence of data presentation over thorough discussion of the results. For example, less than a page is devoted to a discussion of item factor analysis results; however, approximately 80 pages are devoted to the presentation of intercorrelation matrices and factor analytic solutions. Another concern was the lack of interpretative information for the test user, provided in the Test Coordinator's Handbook. While it is thorough and well written, it lacks this added interpretative information that would be valuable when discussing the results with parents. Very little theoretical discussion of the constructs being measured was also an issue raised by Kamphaus. He was also disappointed that the KR-20 reliability coefficients for the subtests yielded more values in the .60 to .70 range than expected.

Kamphaus concluded:

> The overall value of the TCS/2, however, should not be lost in the details. The field of individually administered ability (intelligence) testing would progress

> significantly if it were to adopt many of the test development procedures used for the TCS/2. The TCS/2 sets a high standard with respect to diligent and thoughtful test development procedures for academic ability testing. (p. 1027)

Validity research. TCS/2 is a group-administered test of cognitive ability; however, it is not designed to measure all aspects of cognitive ability. Because it is designed for school use, greater emphasis is placed on those reasoning abilities deemed important for success in an educational program. "These include facility in dealing with verbal and nonverbal concepts and recalling previously encountered information" (CTB/McGraw-Hill, 1993, p. 5). Students have a variety of language experiences and cultural opportunities, which was taken into consideration during the development of the TCS/2. Items found on the TCS/2 measure abilities that can be developed through various channels available to everyone in our society, not just those abilities developed through formal school training.

The TCS/2 measures a construct that can be operationally distinguished from the achievement construct. Research was based on the "hypothesis that an academic aptitude test measures learning rate relative to an achievement test and the achievement test a measure of the amount of school material learned and used grade in school as a measure of learning time" (CTB/McGraw-Hill, 1993, p.5). Based on these criteria the aptitude test was successfully and consistently distinguished from the achievement measure. They postulated that these results provided general and indirect support for the construct validity of the TCS/2 (CTB/McGraw-Hill, 1993).

The TCS/2 was standardized jointly with the Comprehensive Test of Basic Skills, Fourth Edition (CTBS/4), a group-administered achievement test, during the spring of 1991. A sample of 87,797 students in grades 2 through 12 from public, Catholic, and other private schools served as the standardization sample. These students were drawn from 99 public school districts, 13 Catholic dioceses, and 68 private, non-Catholic schools. These two measures were standardized jointly so that they can be administered as a complete group assessment battery for those school districts choosing to utilize this type of standardized testing (CTB/McGraw-Hill, 1993).

Relatively little research has been published regarding the validity of the TCS or TCS/2. Excluding the critique by Kamphaus (1998) no other published research has been found on the TCS/2. In a personal communication with an executive from CTB/McGraw-Hill, Blood (1989) reported that the company did not complete any research correlating the TCS and any individually administered cognitive ability test.

McGiverin (1995) presented research that investigated the criterion related validity of the Test of Cognitive Skills. The TCS and WISC-R were compared to examine: differences between standard score means of CSI and each of the WISC-R IQs and factor scores, the relationships between standard scores of the TCS CSI and each of the WISC-R IQs and factor scores, and the range of confidence within which the WISC-R Full Scale IQ can be estimated from the CSI. This study focused on students receiving special education services. Subjects were 118 public school students, in grades one through eight, all whom qualified for special education services as learning disabled students.

Results of t-tests indicated that the TCS CSI significantly underestimated the cognitive abilities operationally defined by the WISC-R. When relationships between the

TCS CSI and WISC-R IQs and factor scores were examined, results indicated the standard scores of the CSI and the WISC-R were significantly related. The strongest relationship was found between the CSI and WISC-R FSIQ (r = .64). The standard error in estimating the WISC-R 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 95 percent probability of accuracy.

Post hoc analyses were also completed to further explore the comparability of the CSI and FSIQ. These analyses revealed that subjects tended to score lower on the CSI as FSIQ-CSI differences increased. Those subjects with reading skill weaknesses obtained significantly lower CSIs than those subjects with average reading skills. When FSIQs of these two groups were compared no significant differences were found. It was also found that students with attention/concentration weaknesses obtained significantly lower CSIs than those with average attention/concentration skills. Cognitive abilities of these two groups were not significantly different as measured by the WISC-R Verbal Comprehension Factor.

Overall, based on these results, McGiverin concluded that the TCS CSI cannot be recommended as an estimate of cognitive ability as operationally defined by the WISC-R for students with learning disabilities. Also, the use of the CSI for screening purposes is not advisable since it is likely to screen out students with learning disabilities.

Blood (1989) studied the comparability of the TCS and the Stanford Binet-Fourth Edition (SB-IV) for students enrolled in special education programs. Scores found in the cumulative school files of 75 special education students, grades 2 through 6, who participated in regular education classes for at least 50 percent of their school day, were used for this study. Results yielded substantial correlations between the CSI and Nonverbal Reasoning/Visualization Factor ($\underline{r}$ =. 667) and with the Composite ($\underline{r}$ =. 705). Moderate correlations were found between the Verbal Comprehension Factor and the CSI ($\underline{r}$ =. 543), and mild correlation between the Memory Factor and the CSI ($\underline{r}$ =. 270). Blood also found that mean differences were not significant between the CSI and Verbal Comprehension Factor, or between the CSI and the Memory Factor. However, significant differences were found in the means of the CSI and Composite and the CSI and the Nonverbal Reasoning/Visualization Factor.

The mild and moderate correlations between the CSI and the Memory Factor and the Verbal Comprehension Factor respectively, suggest that the CSI has little in common with either factor. Therefore there appears to be no practical value in using the CSI to predict scores on the Memory and Verbal Comprehension Factors. For referral/screening purposes there is greater value in predicting the SB-IV Composite than in predicting any of the SB-IV Factors.

Blood concluded that although it shouldn't be used as the only cognitive ability measure in making special education placement decisions, the CSI might make a worthwhile contribution to referral information. In the absence of individually administered standardized tests, the CSI may be used cautiously by speech pathologists to help determine eligibility for language therapy and by teachers for program planning.

McGiverin (1995) raised a number of interpretive concerns regarding Blood's study. First, although special education students served as the subject sample, no information regarding special service classification was reported. Thus, the collapse of data across categories may confuse interpretation of results. Another concern, while the 2.37 mean point difference was described as small (but statistically significant), the accuracy that an individuals SB-IV Composite can be estimated from a given CSI was not reported.

Robinson and Nagle (1992) investigated the comparability of the Test of Cognitive Skills (TCS) with the WISC-R and the Stanford Binet-Fourth Edition (SB-IV) in an effort to provide empirical evidence for the valid identification of gifted students. A sample of 75 gifted students enrolled in third, fifth, and eighth grades were used for the study. All students had previously been identified as gifted using a 100-point system. With 90 points or more qualifying students for services, 45 points are allotted for the performance on an aptitude/intelligence test, 45 points to the performance on a standardized academic achievement test, and 10 points accounted for by school grades and teacher recommendation.

Results revealed a mean TCS, Cognitive Skills Index (CSI) score of 130.23, mean WISC-R FSIQ of 124.65, and mean SB-IV Composite score of 121.39. Statistical analyses found that the TCS CSI scores were significantly higher than both the WISC-R FSIQ and SB-IV Composite score. It was also found that students scored significantly higher on the WISC-R (FSIQ) than on the SB-IV (Composite score). Absolute differences between individual scores were reported as follows: 44% of students scored within 5 points of their TCS CSI score on the WISC-R FSIQ, 28% scored within 6 to 10 points, 14.67% scored within 11 to 15 points, and 13.33% revealed a difference of 16 points or greater. The same comparisons were made with the SB-IV Composite scores and revealed the following: 33.33 percent fell within 5 points of their TCS CSI, 22.67 percent fell within 6 to 10 points, 22.67 percent fell within 11 to 15 points, and 21.33 percent were different by 16 points or more. Of those whose scores were different by 16 points or more, all scored lower on the SB-IV Composite score.

Robinson and Nagle (1992) also investigated the relationships among the test scores. Correlations between the TCS and both individually administered tests were significant: WISC-R ($\underline{r}$ =. 41), and SB-IV ($\underline{r}$ =. 51). The correlations between the CSI and WISC-R Verbal IQ were significant ($\underline{r}$ =. 33); however, the relationship between the CSI and WISC-R Performance IQ was not significant ($\underline{r}$ =. 21). When the SB-IV four area scores were compared with the CSI, all were found to be significant: Verbal Reasoning ($\underline{r}$ =. 49), Abstract/Visual Reasoning ($\underline{r}$ =. 41), Quantitative Reasoning ($\underline{r}$ =. 35), and Short-Term Memory ($\underline{r}$ =. 30).

According to Robinson and Nagle, these results suggest that many children enrolled in gifted programs would score higher on the TCS than on the WISC-R and SB-IV, thereby supporting findings of other researchers who have argued that group tests over identify students for gifted placement. Analysis of individual scores in this study indicated 28% of WISC-R Full Scale IQ scores and 44% of SB-IV Composite scores were more than 10 points different than the TCS CSI score, "indicating that a significant proportion of children will show substantial differences in scores when comparing this group test with the individually administered tests (Robinson & Nagle, 1992, p.111). They concluded that the TCS may be used as a screening device provided the results of the TCS were supported by an individual test of cognitive abilities.

Wechsler Intelligence Scale for Children-Third Edition

Overview. The Psychological Corporation published the Wechsler Intelligence Scale for Children-Third Edition (WISC-III), the latest version of the Wechsler scales for children, in 1991. It is an individually administered clinical instrument for assessing the intellectual ability of children aged 6 years, 0 months through 16 years, 11 months. The primary reason for revising the test was to update the norms. Wechsler developed the earliest version of the WISC-III in 1949 as a downward extension of the adult intelligence test, the Wechsler-Bellevue Intelligence Scale. This early version was known as the Wechsler Intelligence Scale for Children (Wechsler, 1991).

The WISC-III contains 13 subtests, six in the Verbal Scale and seven in the Performance Scale. Ten subtests comprise the standard battery, five subtests in each scale. In the Verbal Scale, the five standard subtests include: Information, Similarities, Arithmetic, Vocabulary, and Comprehension. The standard battery of the Performance Scale consists of: Picture Completion, Coding, Picture Arrangement, Block Design, and Object Assembly. The supplementary subtests include: Digit Span in the Verbal Scale and Symbol Search and Mazes in the Performance Scale. Approximately 73 percent of the WISC-R items were retained in the WISC-III either in the original or slightly modified form (excluding the Coding subtest). Symbol Search was the only new subtest added to this ability measure (Wechsler, 1991).

Deviation IQ (M=100, SD=15) is used for the Verbal, Performance, and Full Scale IQs; Index Scores (M=100, SD=15) for four factor scores, and scaled scores (M=10, SD=3) for the 13 individual subtests. An IQ is computed by comparing the examinee's scores with scores obtained by a representative sample of his or her age group. For individual subtests, raw scores are converted to scaled scores within the examinee's age group through use of tables found in the WISC-III manual. Percentile ranks for IQ and Index Scores as well as test age equivalents of raw scores are also provided on the WISC-III (Wechsler, 1991).

The table used to calculate IQ scores is based only on the 10 standard subtests. Supplementary subtests are excluded from the calculation of the IQ unless a standard subtest is spoiled or not given. However, when a supplementary subtest is used in place of a standard subtest, little is known about the reliability and validity of the IQs. None of the supplementary subtests were used in the construction of the tables used to generate IQs (Sattler, 1992).

Guidelines for the use of the three supplementary subtests are provided in the WISC-III manual. The guidelines state that Digit Span may substitute for any Verbal subtest, Mazes for any Performance subtest, and Symbol Search may substitute for Coding only. However, the manual fails to report how these recommendations were reached (Wechsler, 1991).

Factor Scores are also obtained from the WISC-III. Depending on the number of subtests administered, two to four factor scores can be calculated. These factor scores help in identifying meaningful psychological dimensions. The Verbal Comprehension factor measures verbal knowledge and understanding obtained through both informal and formal education and reflects the application of verbal skills to new situations. It consists of the following subtests: Information, Similarities, Vocabulary, and Comprehension. "The Perceptual Organization factor score, a nonverbal score, reflects the ability to interpret and organize visually perceived material within a time limit" (Sattler, 1992, p. 1049). It consists of: Picture Completion, Picture Arrangement, Block Design, and Object Assembly subtests. The Verbal Comprehension and the Perceptual Organization factors are obtained when the standard 10-subtest administration of the WISC-III is performed. The Processing Speed factor measures the ability to process visually perceived nonverbal information quickly. Concentration and rapid eye-hand coordination may be important components of the Processing Speed factor. This factor is comprised of the Coding and Symbol Search subtests. The Freedom from Distractibility Factor, comprised of the Arithmetic and Digit Span subtests, is not really a measure of distractibility or inattention but rather is better conceptualized as a working memory index. It measures the ability to hold information in mind temporarily while performing some operation or manipulation with that information, or engaging in an interfering task, then accurately reproducing information or correctly acting on it. It presumes attention and concentration and the ability to exert mental control.

To better understand this assessment instrument, a description of the subtests and their purpose is appropriate. The Verbal Scale consists of the following subtests: Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span. Information is a subtest that measures a child's range of factual knowledge and long-term memory. The child must answer a series of questions on a broad range of topics. Performance on this subtest "may be influenced by cultural opportunities, outside interests, richness of early environment, reading, and school learning" (Sattler, 1992, pp.1122-1123). The Similarities subtest requires a child to state how two things are alike and is a measure of verbal concept formation and long-term memory. Like Information, cultural opportunities, interests, reading habits, and school learning can influence performance on the Similarities subtest. Arithmetic is a subtest that measures a child's ability to mentally solve arithmetic problems. Various types of arithmetic problems (addition, subtraction, multiplication, division, and problem solving) make up this subtest. This subtest provides information about a child's numerical reasoning ability, concentration, attention, short-term memory and long-term memory. A child's attitude toward school and level of anxiety can effect performance on the Arithmetic subtest. Vocabulary is a subtest that provides information about the child's verbal skills, language development, and long-term memory. On this subtest the child is asked to give the definitions of various words of increasing difficulty. "Performance may be influenced by cultural opportunities, education, reading habits, and familiarity with English" (Sattler, 1992, p. 1123). The Comprehension subtest is a measure of a child's social judgment and common sense. The child must answer a series of questions about various problem situations. "Performance may be influenced by cultural opportunities, ability to evaluate and draw from past experiences, and moral sense" (p.1123). The final subtest of the Verbal Scale is the Digit Span subtest, which measures short-term memory, attention and concentration. The child is required to repeat a series of numbers given by the examiner; exactly as given on the first part and in reverse order on the second part (Sattler, 1992).

The Performance Scale consists of seven subtests, they are: Picture Completion, Coding, Picture Arrangement, Block Design, Object Assembly, Symbol Search, and Mazes. Picture Completion provides a measure of the child's alertness to detail and ability to differentiate essential from nonessential details. On this subtest the child must identify the missing part of a picture. Cultural experience and alertness to the environment may influence performance on this subtest. Coding assesses a child's ability to learn a code rapidly and provides an indication of speed and accuracy of eye-hand coordination, short-term memory, and attentional skills. The child is required to view a key where symbols are paired with other symbols and then must fill in the matching symbol when given one part of the pair. Motivation and rate of motor activity may influence performance on this subtest. Picture Arrangement assesses a child's ability to comprehend and evaluate social situations, ability to attend to details, alertness, planning ability, and visual sequencing. The child is given a series of pictures in the wrong order and they must arrange the pictures in the proper order so they tell a story that makes sense. Cultural opportunities may influence performance. Block Design provides a measure of a child's spatial visualization, nonverbal reasoning, and visual-motor coordination abilities. The child must reproduce a design given, using blocks. Rate of motor activity and degree of color vision may affect performance on this subtest. Object Assembly looks at a child's ability to synthesize concrete parts into meaningful wholes as well as visual-motor coordination. The child is asked to correctly assemble a puzzle. Rate of motor activity, persistence, and experience with part-whole relationships may influence performance. The Symbol Search subtest measures a child's visual discrimination and visual-perceptual scanning ability. The child is required to view target symbols then look at a second group of symbols and determine if the target symbol is in the other group of symbols. Rate of motor activity, motivation, and cognitive flexibility may be influential on a child's performance. The final subtest of the Performance Scale is Mazes. This subtest assesses a child's planning ability and perceptual organization ability. The child is required to draw a continuous line out of a maze without running into a blocked passage. Visual-motor organization ability and ability to delay actions may be factors that influence performance on this subtest (Sattler, 1992).

Critiques. Braden (1995) in a critique of the WISC-III, heralded the changes made from the WISC-R. He indicated that it would likely remain the test of choice for assessing children. Its predecessor, the WISC-R, is the most popular and widely researched test of children's intelligence. At the publication of Braden's critique, the WISC-III already had a substantial body of research supporting its use, not including related research already done on the WISC-R. The addition of index scores and factor analyses to support them are major improvements from the WISC-R. Depending on the procedure used, the factor structure of the WISC-III may or may not be supported. However, the Index model is still a better structure (closer to reality) than just the Verbal, Performance, and Full Scale approach of the WISC-R. Data suggest good convergent and divergent validity. Studies also support the ability of the WISC-III to predict relative outcomes.

The manual summarizes studies in which gifted, mentally deficient, Seriously Emotionally Disturbed, Learning Disabled, epileptic, Attention Deficit Disorder, hearing impaired/deaf, and language/speech impaired (L/S) children show atypical WISC-III scores or patterns. Results are consistent with expected values except: as expected L/S children have lower Verbal than Performance IQs but their Performance IQs are still significantly lower than average, and the Processing Speed Index was not lower than other Index scores in learning disabled children.

Another strength of the WISC-III is that the renorming took many steps to eliminate biases. Research suggests that the WISC-III is equally valid for native-English speaking children regardless of gender or ethnicity. The WISC-III also is an improvement over the WISC-R in its providing support for clinical diagnosis; however, at the publication of Braden's review, research suggested that the WISC-III was not terribly sensitive to abnormal conditions. An advantage that is rarely mentioned is the availability of a co-normed achievement battery, the Wechsler Individual Achievement Test (WIAT).

Sandoval (1995) also critiqued the WISC-III. "The Wechsler Intelligence Scale for Children (WISC) is undoubtedly the test of choice for tens of thousands of school and child clinical psychologists who need an appraisal of a child's intellectual functioning" (p.1103). Sandoval commended the modernizing of the WISC, finding materials to be well made and general appearance to be attractive. Stimuli are generally printed in color and various ethnic and racial groups are represented. Roughly ¾ of items were retained from the WISC-R. New items were added to the beginning and end of subtests so they may be used with more confidence. The Symbol Search subtest, of course, is an entirely new subtest. Transitioning from earlier versions of the WISC will be relatively easy for the experienced examiner. However, subtle differences should be studied prior to using the test. The manual is very good, containing technical information, directions for administration and scoring, and norms all in one volume. The tabs in the manual are also a noted improvement, making information more easily accessible. The norms for the revision are excellent, closely mirroring 1988 U.S. census data. The test has outstanding psychometric properties as well. Internal consistency and stability reliability coefficients are very high and are among the highest of any psychological measure. The exception is Mazes with an average stability coefficient of .57. Results of an interscorer agreement study were reported for the first time with very positive results. Useful tables are provided that aid in interpretation. However, there is not a table of subtest specificities, information valued by sophisticated examiners.

Validity research. Relatively little research has been published comparing group administered ability tests to individual measures such as the WISC-III. In contrast, the WISC-III has had a vast amount of research reported about its correlation with other individually administered tests, supporting the validity of the instrument.

The only research comparing the WISC-III to a group-administered test was a comparison with the Otis Lennon School Ability Test Sixth Edition (OLSAT). Wechsler (1991) reported results of the study of 65 children ages 6 to 16 years. Results indicated a strong correlation of .73 between the WISC-III FSIQ score and the OLSAT Total School Ability Index. Strong to moderate correlations of .69 and .59 were reported for the WISC-III VIQ and OLSAT Verbal School Ability Index and the WISC-III PIQ and the OLSAT Nonverbal School Ability Index respectively. No research has been published regarding the relationship between the TCS/2 and the WISC-III.

Studies comparing the WISC-III to other similar measures have found that it is indeed a valid instrument to use. Slate (1995) studied the relationship of the WISC-III and its predecessor the WISC-R. The mean WISC-III Full Scale, Verbal, and Performance IQs were 8, 5, and 7 points lower, respectively, than the mean WISC-R scores. T-tests revealed that these differences were significant. When corrected correlations were computed, correlations were high, sharing between 64 and 76 percent of common variance. Zimmerman and Woo-Sam (1997) also reported results of various studies comparing the WISC-III and the WISC-R in their review article on the criterion-related validity of the WISC-III. The average correlation of the WISC-R and WISC-III Full Scale IQs was .79. Results were similar in comparisons of Verbal and Performance IQs, where mean correlations were .77 and .72, respectively. Zimmerman and Woo-Sam reported that Full Scale IQs on the WISC-R were consistently higher than those on the WISC-III by an average of 6.17 IQ points.

Thompson and Sota (1998) studied the relationship between the WISC-III and the WAIS-R with a sample of 16-year-olds. Because either scale may be used with 16-year-olds, the authors thought it important to examine the relationship between the two. Results from the sample of 46, 16-year-olds revealed that mean summary IQs were slightly higher for the WAIS-R than the WISC-III; however, corresponding WAIS-R and WISC-III summary IQ means were more similar in this study than those found by Wechsler (1991). For subtests, five of the ten corresponding WAIS-R and WISC-III means were significantly different. The WAIS-R mean was significantly higher than the corresponding WISC-III mean for four of the subtests. Picture Arrangement had the lowest intertest correlation. Thompson and Sota were not surprised by this result because bonus points for speed are incorporated on the WISC-III version of this subtest, but not on the WAIS-

R. The WISC-III manual (1991) reports a similar study in which it was suggested that the two tests measured very similar constructs. On average, WAIS-R Verbal, Performance, and Full Scale IQs were about 2, 6, and 4 points above corresponding WISC-III IQs, respectively. Sattler (1992) reported that it is not unusual for individuals to score lower on newer tests than on older ones. Despite the score differences, Sattler indicated adequate concurrent and construct validity for the WISC-III.

Studies comparing the WISC-III and the Stanford-Binet IV have also been reported. Rust and Lindstrom (1996) reported the results of 57 children who were administered both the WISC-III and the Stanford-Binet IV. Average WISC-III Full Scale IQ scores were slightly higher than those for the Stanford-Binet IV, but the differences were not significantly different. For 29 children higher scores were obtained on the WISC-III Full Scale IQ than on the Stanford-Binet Composite; 24 students scored higher on the Stanford-Binet Composite. Overall the difference in the average IQs across all students was 1.35 IQ points. The correlation between the WISC-III Full Scale IQ and the Stanford-Binet IV Composite score was .81. These results further support the validity of the WISC-III.

Lavin (1996) also reported results comparing the WISC-III and the Stanford-Binet IV. Forty children were administered both tests, half were administered the WISC-III first and the other half the Stanford-Binet IV. No significant differences were found between the Stanford-Binet Composite and WISC-III Full Scale IQs, between WISC-III Verbal IQs and Stanford-Binet Verbal Reasoning Area scores, or between WISC-III Performance IQs and scores on Abstract/Visual Reasoning of the Stanford-Binet. Significant correlations were found between all Stanford-Binet and WISC-III IQs. "The correlations between Composite IQs of the Stanford-Binet and the Full Scale, Verbal, and Performance IQs of the WISC-III were .817, .789, and .609, respectively" (pp.493-494). Zimmerman and Woo-Sam (1997) reported average correlations of .77 for six studies comparing the WISC-III Full Scale IQ and the Stanford-Binet IV Composite Score. Results indicated that both tests measure similar constructs.

Rust and Yates (1997) studied the concurrent validity of the WISC-III and the Kaufman Assessment Battery for Children (K-ABC). A sample of 67 children ranging from 6 to 12 years, 6 months served as participants for this study. The correlation between the average Full Scale IQ and Mental Processing Composite was r = . 61. The reported average difference of the scores was .12 IQ points. However, it was noted that one participant had a WISC-III Full Scale IQ 37 points higher than the Mental Processing Composite while another student had a Mental Processing Composite score 30 points higher than the Full Scale IQ. While the test scores did not differ significantly, the data suggest that in some cases differences were very large. Overall, the findings were similar to studies reported in the K-ABC manual, comparing the K-ABC to the WISC-R.

Dumont, Cruse, Price, and Whelley (1996) compared the WISC-III with the Differential Ability Scales (DAS). Fifty-three students who, when initially evaluated for special education services, were administered a 12-subtest WISC-III served as the sample for this study. The DAS was administered during their required triennial reevaluation "to provide additional diagnostic information and to validate their intellectual functioning with previous scores on the WISC-III" (p. 206). The DAS General Conceptual Ability score

(GCA) correlates highly (.78) with the WISC-III Full Scale IQ. The verbal composites of both tests also correlated highly (.77) and the Nonverbal Reasoning and Spatial Ability clusters correlated fairly evenly with the WISC-III Performance IQ (.65 and .67 respectively). The authors concluded that the DAS may be used as an alternative to the Wechsler scales when conducting triennial evaluations for students with Specific Learning Disabilities.

Zimmerman and Woo-Sam (1997) also reported results of comparisons of the WISC-III and various other measures of intellectual ability. Overall, it was reported that strong criterion validity was found. Eleven different ability measures were compared with the WISC-III Full Scale IQ and an average correlation of .75 was reported. Lower WISC-III Full Scale IQ scores were characteristic of below average samples, while higher WISC-III Full Scale IQ scores were characteristic for samples of above average intelligence. Zimmerman and Woo-Sam also found that the WISC-III Full Scale IQ is substantially correlated with achievement measures, generally meeting the accepted criteria of .50 to .65.

Summary

For many students in today's schools, a standardized group ability test score may be found in their cumulative school record. Many have questioned the role the results of such tests play in a student's education. It has been suggested that the results be used as a screening measure in determining proper placement in instructional/educational programs. Many school districts already use such results as part of the requirements for gifted program placement. However, with relatively little research and conflicting results, it remains unclear if group ability test scores are appropriate to be used for any purpose. Additionally, because results are found in students' cumulative file, any school personnel with file access may view these results. Is it fair to allow teachers to form opinions about their students based on group test scores, particularly if such results are not valid?

This is an important issue because it carries implications for the accuracy of placement of children in special education. For a child to be diagnosed as learning disabled, a prescribed discrepancy between ability and achievement scores must be present. Each state has its own standards for determining discrepancy; Delaware utilizes a chart published in *The Administrative Manual for Special Education Services* (2000) to determine the existence of ability-achievement discrepancies. If the CSI is not a valid measure of cognitive ability, this would influence the ability-achievement discrepancy score necessary for diagnosis. For example, if the CSI underestimates ability, a discrepancy between ability and achievement scores might not be present, and thus an under identification of students for special education would exist. Conversely, if the CSI overestimates ability, an over identification of students for special education might occur because there would be a larger discrepancy score. Thus, it is important to investigate the validity of the CSI as a measure of intelligence with special education children.

It was the purpose of this research to investigate the relationship between the group-administered TCS/2 and the individually administered WISC-III. Specifically, we determined the validity of the TCS/2 as a measure of intelligence by comparing its scores with those of the WISC-III. Previous research has been conducted to determine the relationship between the previous versions of both these assessment instruments (TCS and

WISC-R). However, there were no research studies assessing the relationship of these current versions (TCS/2 and WISC-III). Similarly, no research relating the TCS/2 to any other individually administered intelligence test had been found. From research on previous versions of these tests, it was reasoned that a significant relationship existed between the WISC-III and the TCS/2.

Chapter II

Method

Subjects

Subjects of the study were selected from public school students in grades 6 through 12 that received special education services. All students resided in a unique bi-state school system in the Mid-Atlantic region. These students attend school in one state for elementary instruction (where the TCS/2 was administered) and a school in another adjoining state for middle and secondary instruction. Criteria for inclusion in the study were both a CSI score from the TCS/2, found in their cumulative record, and WISC-III scores within three years of the TCS/2 administration, found in their special education file.

Of the 122 students receiving special education services, 66 met criterion for inclusion in this study. Of the 66 students, whose records were utilized for this study, 14 were female and 52 were male. The majority of students were enrolled in grades 6 through 9 (6^{th} grade = 16, 7^{th} grade = 14, 8^{th} grade = 14, 9^{th} grade = 16, 10^{th} grade = 10, and 11^{th} grade = 1). The classification areas for which these students were eligible in the area of special education services were: learning disability ($\underline{n}$=55), physical impairment ($\underline{n}$=8), and educably mentally handicapped ($\underline{n}$=3).

Apparatus

TCS/2. The Test of Cognitive Skills, Second Edition (TCS/2) is a group-administered cognitive abilities test designed to assess the academic aptitude of students in grades 2 through 12. Divided into four subtests, the TCS/2 is "intended to measure selected verbal, nonverbal, and memory abilities that can contribute greatly to students' success in an educational program" (CTB/McGraw-Hill, 1993, p.1). The TCS/2 composite, the Cognitive Skills Index, is a normalized score that has a mean of 100 and a standard deviation of 16. Regular education teachers who have received in-service training regarding administration procedures administer this test.

WISC-III. The Wechsler Intelligence Scale for Children-Third Edition (WISC-III) is the latest version of the Wechsler scales for children published by The Psychological Corporation in 1991. It is an individually administered clinical instrument for assessing the intellectual ability of children ages 6 years, 0 months through 16 years, 11 months. For the purposes of this study, it was administered either by a certified school psychologist or an intern in the field of school psychology. The WISC-III contains 13 subtests that are divided into the Verbal and Performance Scales. Verbal, Performance, and Full Scale IQs are calculated as well as four Index scores all having a mean of 100 and a standard deviation of 15.

Procedure

Permission to obtain information from student's cumulative and special education files was granted from the superintendent of the school district. Information was gathered such that the identity of individual students remained confidential. Scores were only associated with a subject number and not with any personally identifying information. The following information was gathered: special education classification, WISC-III IQ and Index scores, date of WISC-III administration, TCS/2 CSI score, and date of TCS/2 administration.

Chapter III

Results

A descriptive summary of the means and standard deviations of test scores for the total subject sample is presented in Table 1. Means for the CSI, WISC-III IQs, Verbal Comprehension Index, and Freedom from Distractibility Index were within the low average range of the normative groups. The Perceptual Organization and Processing Speed Indices fell within the average range (see Table 1).

Relationship between CSI and WISC-III

The relationships between the CSI and each WISC-III score were examined through the computation of Pearson product-moment correlations which are displayed in Table 2. Results indicate that, for special education students, the standard scores of the CSI and the WISC-III IQs are significantly related. All Pearson coefficients for the CSI paired with each WISC-III score were positive and statistically significant, except for the CSI and Freedom from Distractibility relationship ($r = .30$, $p > .05$). The CSI and Full Scale IQ shared the strongest relationship ($r = .50$, $p < .05$).

Difference between CSI and WISC-III

Results of the tests of significance between the CSI and WISC-III scores are presented in Table 3. The t-test comparing subjects' CSI and Full Scale IQ yielded a t value of -2.74 ($df = 64$, $p = .008$). A comparison of the CSI and Verbal IQ yielded a t value of -2.20 ($df = 65$, $p = .032$). A t value of -4.20 ($df = 65$, $p = .000$) was yielded for the comparison of the CSI and Performance IQ. Comparison of the CSI and Verbal Comprehension Index yielded a t value of -2.87 ($df = 42$, $p = .006$). A t value of -5.26

(df = 42, p = .000) was yielded for the CSI and Perceptual Organization Index. T values of -2.41 (df = 36, p = .021) and -3.37 (df = 34, p = .002) were yielded for the comparisons of the CSI and Freedom from Distractibility and Processing Speed Indices, respectively. Results were all significant at the .05 level, which indicated significant differences between the means of the CSI and all WISC-III IQ and Index scores. Students WISC-III scores were significantly higher than their CSI score.

The standard error of estimate for the WISC-III Full Scale IQ, given the CSI, was 12.99. A 51-point range of standard scores is required in order to estimate a Full Scale IQ from the CSI within a 95% probability of accuracy.

Relationship between WISC-III scores

Results indicated that, for special education students, the WISC-III Full Scale IQ correlated significantly with all other IQ and Index scores (see Table 2). The highest correlations were between the Full Scale IQ and the Verbal and Performance Scale IQs. Correlations were also high between the Full Scale IQ and the Verbal Comprehension and Perceptual Organization Indices. However, the relationships between the Full Scale IQ and the Freedom from Distractibility and Processing Speed Indices were only moderate. Scores that were not significantly correlated included: Verbal IQ and Processing Speed Index (r = .20), Performance IQ and Freedom from Distractibility Index (r = .25), Verbal Comprehension Index and Processing Speed (r = .17), Perceptual Organization Index and Freedom from Distractibility Index (r = .07), and Freedom from Distractibility Index and Processing Speed Index (r = .29).

Chapter IV

Discussion

Results suggest that the two instruments studied, the TCS/2 and the WISC-III, do measure similar constructs. Significant, positive relationships were found between the TCS/2 CSI and the WISC-III IQs and Index scores, except for the CSI and Freedom from Distractibility Index relationship. That particular relationship was still positive, but not significant.

In general, the CSI is used or reported as the FSIQ is on the WISC-III. It is seen as an overall measure or indication of a student's cognitive ability. Therefore, it would be reasoned that relationship would be the strongest. That relationship, between the CSI and Full Scale IQ, was in fact the strongest relationship found in this study ($r = .50$, $p<.01$) between CSI and WISC-III scores. As the Verbal IQ and Performance IQ are combined in order to calculate the Full Scale IQ, the expectation that these relationships be the next strongest was met (both at $p<.01$). However, all of these relationships are modest at best.

Given the significant relationship between the CSI and Full Scale IQ, the standard error of estimate was calculated to determine the confidence range of scores within which the Full Scale IQ can be estimated from the CSI at an acceptable level of probability. This revealed a substantial error in estimating the Full Scale IQ from a given CSI for an individual student. Therefore, the confidence range of scores within which the Full Scale IQ can be estimated from the CSI was also substantial. A range of 51 standard score points was required to estimate a student's Full Scale IQ within a 95% probability of accuracy. In comparison, the standard error of estimate for the WISC-III Full Scale IQ is

3.07; thereby requiring a range of 12 standard score points to estimate the Full Scale IQ within a 95% probability of accuracy (Wechsler, 1991). Thus, although the relationship between the CSI and Full Scale IQ was statistically significant, an extreme range of scores was required to estimate the Full Scale IQ from the CSI at an acceptable confidence level for special education students. With such an extreme confidence range necessary to accurately estimate the WISC-III Full Scale IQ from the CSI, the CSI is not recommended for use as an estimate of cognitive ability.

Differences between the standard score means of the TCS/2 CSI and the WISC-III scores were all significant. The CSI mean was significantly lower than the means of all the WISC-III scores, indicating that the TCS/2 significantly underestimated the cognitive abilities assessed by the WISC-III. The significant differences found for this sample of special education students are consistent with prior studies that have examined the use of the TCS (predecessor of the TCS/2) with special populations. McGiverin (1995) found similar results with learning disabled students, using the TCS and WISC-R. Robinson and Nagle (1992) found CSI and FSIQ scores to be significantly different for gifted students; however, gifted students scored higher on the CSI than on the WISC-R. Similar results were also found in comparisons of the TCS and SB-IV. Blood (1989) found that special education students' CSI scores were significantly lower than the SB-IV Composite, whereas gifted students scored significantly higher on the CSI than the SB-IV Composite (Robinson & Nagle, 1992). Results of these studies indicate that the CSI typically underestimates the abilities of special education students, while overestimating the abilities of gifted students; further supporting that caution be used when reporting or relying on the

CSI as a measure of cognitive ability.

The CSI scores were significantly lower than the WISC-III scores; therefore underestimating a student's cognitive ability. This can under-identify students for special education because theoretically there is now a smaller difference between the student's ability and achievement scores. Therefore, a significant discrepancy between ability and achievement may not be found, thereby not identifying the student for special education services as a learning disabled student (AMSES, 2000). Unfortunately, these students are typically referred to as low achievers whom are working to their potential. For example, a student was administered the TCS/2 and received a CSI score of 80, when administered the WISC-III his Full Scale IQ was 97. At the same time, his achievement scores in the areas of reading, math, and written expression were all 75. If a placement decision was made based on the CSI score, using Delaware Department of Education discrepancy standards, the student would not be eligible for special education services as a learning disabled student because there is not a significant discrepancy between ability and achievement. However, if a placement decision was made based on the Full Scale IQ, using the same standards, the student would be identified as a learning disabled student in the areas of reading, math, and written expression and would be eligible to receive specialized instruction inside or outside the classroom (AMSES, 2000).

To determine the validity of the TCS/2 CSI based on its relationships with WISC-III scores, it is important to review the validity of the WISC-III by comparing the internal relationships found to ensure that results of this study coincide with validity research completed by Wechsler (1991). If our study revealed validity issues within the WISC-III, it would not be appropriate to determine the validity of the TCS/2 CSI based on its relationships with the WISC-III.

The Full Scale IQ was found to be significantly related to all scores reported on the WISC-III, thereby supporting the validity of this instrument. The strongest relationship reported was between the Full Scale IQ and the Performance IQ, with the Full Scale IQ and Verbal IQ relationship next strongest. These results mirror those found with the CSI and IQ relationships. The Full Scale IQ and Performance IQ/Verbal IQ relationships are very strong (r = .89 and r = .87, respectively). Strong relationships were also demonstrated between the Full Scale IQ and Perceptual Organization Index/Verbal Comprehension Index (r = .86 and r =.85, respectively). The Processing Speed and Freedom from Distractibility Index scores revealed the weakest relationships. This is supported by Wechsler who stated in the manual, "The Freedom from Distractibility and the Processing Speed scales are not as highly related to general intellectual ability as the Verbal Comprehension and Perceptual Organization scales" (Wechsler, 1991, p.210).

Although our results suggest that the two instruments in question are measuring similar constructs, the question remains whether the TCS/2 CSI is recommended as a valid measure of cognitive ability? Even though the relationships reported between the CSI and WISC-III scores were positive and significant (except for Freedom from Distractibility), the correlation between scores is moderate at best. The strongest relationship reported was only r = .50. It would be more acceptable if the correlation were at least r = .80. Therefore, it is recommended that extreme caution be used when regarding the CSI as a measure of cognitive ability for special education students.

A major implication of this study is that the TCS/2 is inappropriate for the screening of students for identification of need for special education services. The CSI significantly underestimated the cognitive abilities of many students in the sample. If used for screening purposes, the CSI would screen out many students identified for special education services. McGiverin (1995) and Blood (1989) also found the CSI to underestimate the cognitive abilities of special education students; however, Robinson and Nagle (1992) found that the CSI overestimated cognitive abilities of gifted students. Further research is needed to determine if the TCS/2 is an appropriate screening measure for specific groups of students.

With research suggesting that the CSI is not a valid measure of cognitive ability (when compared to the WISC-III) for special education students, teachers and other school personnel should not place emphasis on these scores when forming opinions about students and when considering educational placements. This study revealed that the CSI frequently underestimates the ability of students who would qualify for special education services, therefore under identifying those students. If placement were to rely solely on the CSI score, many students who need specialized instruction provided for by special education services, would not receive it because they were not identified. School districts that use the CSI score as a primary basis for entrance into specific educational programs, should take these results into consideration as this study indicates that the CSI underestimates cognitive ability for special education students. It is not practical to require individually administered ability tests for all program placements (as the demand is already high due to federally mandated use for special education placements); however, a more valid measure or process for screening and/or identification should be sought. The practicality of group administered tests, makes them a logical choice for screening purposes; however, more extensive research should be completed to determine those group-administered test(s) that would be more valid measures of cognitive ability.

As the research comparing these two instruments (particularly the newest versions) is very scarce, further studies comparing them are warranted. Future studies might include a larger special education sample, a gifted student sample, as well as a normative student sample. With larger sample sizes, it would also be helpful to disseminate the differences between different special education classifications. It would also be helpful to compare the TCS/2 with the validity of other group-administered tests to determine the most appropriate test to use for screening purposes, or to see what other purposes group tests might serve. More research would help to clarify whether the TCS/2 CSI is a valid measure of intelligence that could be used by school personnel as a factor in educational placement decisions.

References

AMSES - The administrative manual for special education services. (2000). Dover, DE: Delaware Department of Education.

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, 50, 2869.

Braden, J. P. (1995). Wechsler Intelligence Scale for Children-Third Edition. In J. C. Conoley & J. C. Impara (Eds.), The twelfth mental measurements yearbook (pp.1098-1103). Lincoln, NE: University of Nebraska.

CTB/McGraw-Hill. (1993). Test of Cognitive Skills-Second edition: Technical report. Monterey, CA: Author.

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

Dumont, R., Cruse, C. L., Price, L., & Whelley, P. (1996). The relationship between the Differential Ability Scales (DAS) and the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) for students with learning disabilities. Psychology in the Schools, 33(3), 203-209.

Fields, J. P., & Kumar, V. K. (1982). How teachers use the group I.Q. test scores. Journal of School Psychology, 20(1), 32-38.

Kamphaus, R. W. (1998). Test of Cognitive Skills, Second Edition. In J. C. Impara & B. S. Plake (Eds.), The thirteenth mental measurements yearbook (pp.1026-

1027). Lincoln, NE: University of Nebraska.

Lavin, C. (1996). The Wechsler Intelligence Scale for Children-Third Edition and the Stanford-Binet Intelligence Scale: Fourth Edition: A preliminary study of validity. Psychological Reports, 78(2), 491-496.

McGiverin, J. L. (1995). The comparison of the Test of Cognitive Skills and the Wechsler Intelligence Scale for Children-Revised for students with learning disabilities. (Doctoral dissertation, Indiana State University, 1995). Dissertation Abstracts International, 57, 3812.

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(2), 107-112.

Rust, J. O., & Lindstrom, A. (1996). Concurrent validity of the WISC-III and Stanford-Binet IV. Psychological Reports, 79(2), 618-620.

Rust, J. O., & Yates, A. G. (1997). Concurrent validity of the Wechsler Intelligence Scale for Children-Third Edition and the Kaufman Assessment Battery for Children. Psychological Reports, 80(1), 89-90.

Sandoval, J. (1995). Wechsler Intelligence Scale for Children-Third Edition. In J. C. Conoley & J. C. Impara (Eds.), The twelfth mental measurements yearbook (pp. 1103-1104). Lincoln, NE: University of Nebraska.

Sattler, J. M. (1992). Assessment of children (revised and updated 3rd ed.). San Diego, CA: Author.

Slate, J. R. (1995). Two investigations of the validity of the WISC-III. Psychological Reports, 76(1), 299-306.

Thompson, A. P., & Sota, D. D. (1998). Comparison of WAIS-R and WISC-III scores with a sample of 16-year-old youth. Psychological Reports, 82(3), 1339-1346.

Walsh, W. B. & Betz, N. E. (2001). Tests and assessment(4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale-Third Edition. San Antonio, TX: The Psychological Corporation.

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

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

Zimmerman, I. L., & Woo-Sam, J. M. (1997). Review of the criterion-related validity of the WISC-III: The first five years. Perceptual and Motor Skills, 85(2), 531-546.

Table 1. Means and Standard Deviations for CSI and WISC-III Standard Scores.

Test Score	Mean	Standard Deviation
Cognitive Skills Index	81.71	9.36
WISC-III Full Scale IQ	85.57	12.60
WISC-III Verbal IQ	84.92	12.18
WISC-III Performance IQ	88.61	14.83
WISC-III Verbal Comprehension Index	87.28	13.13
WISC-III Perceptual Organization Index	92.19	14.53
WISC-III Freedom from Distractibility Index	86.81	11.16
WISC-III Processing Speed Index	90.54	15.45

Table 2. Pearson Correlations for Total Sample.

	CSI	FSIQ	VIQ	PIQ	VCI	POI	FD
FSIQ	.50*						
VIQ	.42*	.87*					
PIQ	.47*	.89*	.54*				
VCI	.36*	.85*	.98*	.45*			
POI	.45*	.86*	.47*	.95*	.49*		
FD	.30	.50*	.62*	.25	.53*	.07	
PS	.35*	.52*	.20	.68*	.17	.53*	.29

*p<.05

Note.
CSI = Cognitive Skills Index
FSIQ = Full Scale IQ
VIQ = Verbal IQ
PIQ = Performance IQ
VCI = Verbal Comprehension Index
POI = Perceptual Organization Index
FD = Freedom from Distractibility
PS = Processing Speed

Table 3. Differences Between CSI and WISC-III Standard Score Means.

Scores	t	p
CSI - Full Scale IQ	-2.74	.008*
CSI - Verbal IQ	-2.20	.032*
CSI - Performance IQ	-4.20	.000*
CSI - Verbal Comprehension	-2.87	.006*
CSI - Perceptual Organization	-5.26	.000*
CSI - Freedom from Distractibility	-2.41	.021*
CSI - Processing Speed	-3.37	.002*

*p<.05