2006

Is the GAI a good short form of the WISC-IV?

Kimberly A. Scott

Is the GAI a good short form of the WISC-IV?

Thesis submitted to
the Graduate College of
Marshall University

In partial fulfillment of the
requirements for the degree of
Education Specialist
in School Psychology

By
Kimberly A. Scott

Dr. Pete Prewett, Committee Chairperson
Dr. F. J. Kreig
Dr. Goudy

Marshall University
Mary 2006

Abstract

A comparison was done between the standard scores of the GAI and the FSIQ of the WISC-IV using 31 subjects. The mean difference between the GAI/ FSIQ standard score is 3.74. *T tests of significance* show that there is not a significant difference between the scores of the GAI and the Full Scale IQ. The *Pearson r correlation* (.963 @ .01 level) suggests there is a strong positive correlation between the GAI and the Full Scale IQ. In summary, the GAI is a good predictor of the FSIQ of the WISC-IV. More data is required to determine if a statistical difference between GAI and FSIQ scores exist with a bigger sample size.

Is the GAI a good short form of the WISC-IV?

The use of short forms for intelligence testing deserves research and should be incorporated into professional psychology training and continuing professional education (Thompson, 2004). According to NASP and APA ethical standards, school psychologists must use instruments which are valid and reliable and have up-to-date standardization (Jacob and Hartshorne, 2003).

Short Forms of IQ Tests

A survey of psychologists in Australia, Canada, the United Kingdom, and the United States found that the majority have administered 10 or fewer short-form tests in the past year (54% in Australia, 59% in Canada, 53% in the United Kingdom, and 61% in the United States). The most frequently reported approach was a selection of subtests from the Wechsler scales. Wechsler subtest short forms had been used by 77% to 90% of respondents who reported information about their procedures. These included combinations of two to eight subtests that have been recommended in the literature. Several psychologists surveyed administered one less verbal or performance subtest and used an accepted prorating procedure. Other combinations reported were ad hoc selections of subtests. Approximately 50% to 60% of the Wechsler subtest combinations reported in Australia, the United Kingdom and the United States and approximately 75% of those listed in Canada were recommended by the literature for estimating IQs. These results raise some concern that a substantial number of short-form combinations were not based on the short-form literature (Thompson, 2004).

Seventy percent of the psychologists surveyed used an IQ score as the means of metric obtained by the short form. Others used scaled scores, factor scores, or the qualitative information provided by the subtests. Australia (27%), Canada (12%), the United Kingdom (31%), and the United States (15%) who used selected Wechsler subtests, referred to published articles, tables, or formulas for relating sums of scaled scores to IQ estimates. Others used a prorating procedure (22% in Australia, 8% in Canada, 31% in

the United Kingdom, and 15% in the United States). The remaining psychologists surveyed either did not estimate IQs or did not provide sufficient information to explain their procedures (Thompson, 2004). Prorating is not recommended because it does not take into account subtest reliability (Sattler, 1992), and prorating may result in errors up to 1 standard deviation (Crawford et al., 1992).

The median score estimated by the psychologists surveyed for administration of a short-form IQ test was 30 minutes (Thompson, 2004). Psychologists reported that short-forms were frequently used to save time and accommodate client characteristics (e.g. limited attention, poor concentration, lack of perseverance) and other psychological and physical conditions that limited time required for full IQ assessments. Short-forms were also frequently used for screening purposes and to obtain a global estimate of intelligence or ability. Psychologists (70% in Australia, 60% in Canada, 82% in the United Kingdom, and 60% in the United States) considered following up with further intelligence testing when using the short-forms. Some routinely administered the cognitive tests, but the majority made a clinical decision based on the results of the short form and the capacity of the client for additional testing (Thompson, 2004).

Sattler (2004) recommends that short forms of the WISC-IV may be used for: 1) screening purposes, when the short form may be followed by administration of the rest of the test, 2) research purposes, or 3) obtaining an estimate of the child's intellectual functioning when a precise IQ is not required. The short form you use must have acceptable reliability and validity, be appropriate for the child's abilities, be able to provide useful information, and be administered when all 10 core subtests cannot be administered. Using short forms greatly amplifies errors in test administration and gives more weight to each subtest used in the short form. Short forms should not be used to provide a classification for a clinical or psycho educational purposed or for programming decisions. If a short form is used, "Estimate" must be added to the record form and psychological report. Table A-7 in Appendix A of Sattler (2004), provides combinations

of subtests for the WISC-IV and their reliability and validity. The more subtests used in the short form, the higher the reliability and validity of the estimated IQ. The scores obtained from a short form will need to be converted to an estimated FSIQ. Simple proration and regression procedures are not applicable since they do not deal with subtest reliability (Tellegen & Briggs, 1967). The procedure for converting the short-form scores into IQs can be found in Sattler, 2001, *Assessment of Children: Cognitive Applications*. Conversion Tables to obtain IQs from short form scores can be found in Sattler (2004) Appendix A.7. These IQ scores were obtained using the Tellegen & Briggs procedure (Sattler, 2004). Appendix A.8 in Sattler (2003) reports whether the observed scatter (the highest scaled score minus the lowest scaled score) on all the short forms is unusual or not (Sattler, 2004).

The primary criterion for evaluating short forms is the validity coefficient (the correlation between the short form and the total test). Different short forms can be ranked according to the validity coefficient and the short form with the highest validity is considered the best (Silverstein, 1999). Reliability of the short form is also important. Cyr and Brookover recommend using the averaged validity and reliability coefficient as the criterion for evaluating short forms (Cyr & Brookover, 1984).

Taking into consideration that the majority of reporting psychologists in the Thompson study used short forms of the WISC-III in the past year and used subtests of the Wechsler Scales for their short form, this study wished to examine the correlation between the GAI and the WISC-IV Full Scale Intelligence Quotient (FSIQ) of the WISC-IV.

Overview of WISC-IV and GAI

The WISC-IV provides the FSIQ and four index scales: The Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), the Working Memory Index (WMI), and the Processing Speed (PSI). The FSIQ is a measure of general intelligence and best represents “g” or general intellectual functioning. The FSIQ is the most reliable score on

the WISC-IV. It is derived from the sums of the scaled scores for the VCI, PRI, WMI, and PSI. The VCI is composed of three core subtests (Similarities, Vocabulary, and Comprehension) and two supplemental subtests (Information and Reasoning). The VCI measures verbal knowledge and understanding obtained through both formal and informal education and reflects the use of verbal skills to new situations. The PRI is composed of three core subtests (Block Design, Picture Concepts, and Matrix Reasoning) and one supplemental subtest (Picture Completion). The PRI measures the ability to interpret and organize visual material and to produce and test hypotheses related to problem solving. The WMI is composed of two core subtest (Digit Span and Letter-Number Sequencing) and one supplemental subtest (Arithmetic). The WMI measures immediate memory and the ability to concentrate, sustain attention, and exert mental control. The PSI is composed of two core subtest (Coding and Symbol Search) and one supplemental subtest (Cancellation). The PSI measures the ability to process visually perceived material quickly, with concentration and eye-hand coordination (Wechsler, 2003).

The WISC-IV contains 15 subtests divided into 10 core and five supplemental subtests. The 10 core subtests are administered when composite scores are needed. The five supplemental subtests provide additional cognitive information, clinical information, and discrepancy analysis. If needed, supplemental subtests can be used as substitutes for core subtests. The decision to substitute should be based on clinical need, rather than examiner preference. For example, if a child has difficulties with fine motor skills, you may choose to use Cancellation in place of Coding or Picture Completion in place of Block Design. If a subtest is invalidated for whatever reason, a substitution may be needed. At least one supplemental subtest is available for each index score (Wechsler, 2003).

The standard administrative order of subtest must be followed, even when you expect to substitute for a core subtest (Wechsler, 2003).

If two of the three Verbal Comprehension core subtests are valid or if two of the three Perceptual Reasoning core subtests are valid, you can prorate the sum of the scaled score using the Table A.7 on page 241 of the WISC-IV Administration Manual. This table was developed by multiplying the sum of the scaled score by 3/2. You cannot use proration for Working Memory or Processing Speed if only one of the two subtests is valid. Proration requires two valid subtests. Proration is similar to using a short form of the test. Proration should be avoided whenever possible since it violates the standard test procedure and introduces unknown measurement error. If proration is used, write "Estimate" by the Index scores and FSIQ on the Record Form and in the psychological report. The Tellegen and Briggs procedure may also be used to give Estimation IQ scores (Sattler, 2004).

The revision goals for the WISC-IV, as cited in the Wechsler Manual (2003), are as follows:

1) To update theoretical foundations of the instrument; several theories of intelligence signify the importance of fluid reasoning. The WISC-IV added three subtests to measure fluid reasoning ability: Matrix Reasoning, Picture Concepts, and Word Reasoning. Picture Concepts and Word Reasoning are adapted form the WPPSI-III and Matrix Reasoning is adapted from both the WAIS-III and WPPSI-III. Research has shown that working memory is an important component of fluid reasoning and other higher order cognitive processes and is closely related to achievement and learning (Fry & Hale, 1996; Perlow, Jattuso, & Moore, 1997, Swanson, 1996). The WISC-IV was designed to more efficiently measure working memory. Arithmetic was revised to be more age appropriate and better measure working memory. Research shows that there is a greater demand on working memory for Digit Span Backward than Digit Span Forward, so separate process scores were developed for these tasks (Reynolds, 1997). Research has shown that the speed of information processing is related to cognitive functioning (Kail & Salthouse, 1994), reading performance (Kail & Hall, 1994), reasoning by the use of cognitive

resources, and the efficient use of working memory for higher order fluid tasks (Fry & Hale, 1996; Kail, 2000). Processing speed may be related to conditions such as epilepsy, Attention-Deficient/Hyperactivity Disorder, and traumatic brain injury (Donders, 1997). Processing speed is also related to neurological development (Berthier, DeBloi, Poirier, Novak, & Clifton, 2000; Cepeda, Kramer, & Gonzalez DeSather, 2001). Research suggests a relationship between working memory, processing speed, and reasoning (Carpenter, Just & Shell, 1990; Fry & Hale, 1996; Kail & Salthouse, 1994). Better processing speed may decrease demands on working memory and help reasoning. Cancellation was developed to provide a supplemental Processing Speed subtest (Wechsler, 2003). Sixty percent of the core subtests in the WISC-IV measure crystallized knowledge and fluid reasoning ability, while 40% measure auditory rote memory and visuomotor processing speed. Ninety percent of the WISC-III measured crystallized knowledge and fluid reasoning ability, while 10% measured visuomotor processing speed. Therefore, consideration will need to be taken when comparing test-retest scores form the WISC-IV and WISC-III. For example, a child with relative strengths in verbal and non verbal reasoning and relative weaknesses in memory and processing speed would probably have a higher FSIQ on the WISC-III than the WISC-IV (Sattler, 2004). 2) To enhance clinical utility: 16 special groups were sampled during the WISC-IV standardization. These were children who were identified as cognitively gifted, mild or moderate mental retardation, Learning Disorders (LD), Attention-deficit/hyperactivity disorders (AD/HD, LD/AD/HD, Expressive and mixed Receptive-Expressive language Disorders, Autism, Asperger's Disorder, open or Closed Head Injury, and motor impairment. A total of 550 children were administered the WISC-IV and Wechsler Individual Achievement Test- 2nd edition (WIAT-II) as a linking study. (Wechsler, 2003). 3) To improve Psychometric properties: As cited in the Wechsler Manual, research suggests that older norms produce inflated scores on intelligence measures (Flynn, 1998). The FSIQ is likely to be lower on the WISC-IV than on the WISC-III by approximately

three points (Sattler, 2004). The WISC-IV normative data was derived form a sample from August 2001 to October 2002 with 2,200 children, except for Arithmetic which used 1,100 children. The sample involved diverse ages (6-16 years), sex, race/ethnicity, parent education level, and geographic region. Several concurrent studies were conducted to provide evidence of the WISC-IV reliability and validity (Wechsler, 2004). Studies correlating the WISC-IV with the WISC-III, WPPSI-III, WAIS-III, WASI, and measures of achievement, memory, emotional intelligence, and adaptive behavior indicate the WISC-IV has satisfactory criterion validity. The WISC-IV also has good reliability. The 4 Composites and the Full Scale have internal consistency reliability coefficients of .81 or above over the entire age range used in the standardization sample. Internal consistency reliability coefficients for the 11 age groups are as follows: Verbal Comprehension= .91 to .95; Perceptual Reasoning= .90 to .93; Working Memory= .90 to .93; Processing speed= .81 to .90; and Full Scale= .96 to .97 (Sattler, 2004). The average standard errors of measurement (SEM) in standard score points are as follows: Verbal comprehension= 3.78; Perceptual Reasoning= 4.25; Working Memory= 5.21; Processing Speed= 5.21; and Full Scale = 2.68 (Sattler, 2004). The WISC-IV subtest form 3 g-related clusters. Good measures of g are Vocabulary, Information, Similarities, Arithmetic, Word Reasoning, and Comprehension. Fair measures of g are Block Design, Matrix Reasoning, Picture Completion, Letter-Number Sequencing, Symbol Search, Picture Concepts, Digit Span, and Coding. Cancellation is a poor measure of g. The Verbal Comprehension subtests have the highest g loading s in the WISC-IV. (Sattler, 2004). Floors and ceiling were improved to cover a wide range of cognitive abilities from mental retardation to giftedness at varying ages (Wechsler, 2003). As cited in the Wechsler Manual (2003), item bias was analyzed using the Mantel-Haenszel bias analysis (Holland & Thayer, 1988) and item response theory and bias analysis (Hambletone, 1993). Experts in cross-cultural research and intelligence testing reviewed the scale (Wechsler, 2003).

The correlations between the FSIQ and the 10 core subtests range from .46 to .72 and the 4 index scores correlate between .70 and .86 with the FSIQ. The VCI and the PRI correlate the most strongly with the FSIQ (VCI= .85 PRI=.86; WMI= .76; and PS=.70) as shown in table 5.1 of the Technical Manual (Wechsler, 2004). The FSIQ is important in identifying children who are mentally retarded, intellectually gifted, learning disabled, or low achievers. It is a useful score when there is little subtest or index score variability. The FSIQ incorporates verbal and nonverbal reasoning, working memory and processing speed that is considered important to overall intelligence. However, if there are significant differences between index scores or between some index scores and the FSIQ, then the FSIQ cannot and should not stand alone as the overall summary of a child's cognitive abilities. The interpretation of the FSIQ must include a description of the differences in the abilities of the child. The base rate must be considered when reporting significant discrepancies between scores to describe the practicality of the results. Tables for base rates discrepancy comparisons are found in the WISC-IV Manual Appendix B.1-B.6. The psychologist may elect not to report the FSIQ if it is not the best representation of the child's diverse abilities. The FSIQ is important in the ability- achievement discrepancy that was written into law (1997) under the Individuals with Disabilities Education Act (IDEA) and adopted by most school districts in the United States. This ability-achievement discrepancy criterion measures a significant discrepancy between the FSIQ and a standardized measure of achievement in areas such as reading and math to identify children with learning disabilities for assistance in special education (Prifitera, 2005).

Various sources for the GAI tables are available, but were generated by different methods. The GAI tables developed by Prifitera et al. (2005) are the only sources supported by Harcourt Assessment, Inc. These tables were developed using the original WISC-IV standardization sample (2200 subjects). GAI tables created by Kaufman and Flanagan (2004) and Dumont and Willis (2004) were developed using a statistical

equation called the Tellegen and Briggs formula (1967) which correlated the VCI and the PRI. The tables created by Dumont & Willis and Kaufman & Flanagan were developed while waiting for the sample data to be standardized for the GAI. The early tables by Kaufman & Flanagan and Dumont & Willis should no longer be used (Raiford, Weiss, Rolfhus, and Coalson, 2005).

Summary of Literature Review

Most psychologists reported they are using short forms of the Wechsler Scales (WISC-III) to obtain an IQ score in order to make a clinical decision. Some of these short form combinations were not research based. (Thompson, 2004). The GAI of the WISC-IV provides scores for six subtests, giving more data and less chance of error than other shorter version short forms. The GAI has a higher loading of g, is composed of subtests which are more interrelated, and has a higher correlation with the FSIQ than the WMI and PSI. It has been noted that subtest variability between the VCI and PRI with the WMI and PS may provide a significant difference between the GAI score and the FSIQ (Wechsler, 2003).

Statement of the Problem

The purpose of this study is to examine if the GAI (used as a short form) is a good short form of the WISC-IV.

Research Questions

Is the GAI (used as a short form) a good short form of the WISC-IV?

1. Does the WISC-IV FSIQ correlate with the WISC-IV GAI score?
2. Is there a significant difference between the mean GAI Score and the mean Full Scale IQ score?

Hypothesis

1. There is a strong positive correlation between the GAI score and the FSIQ score of the WISC-IV.
2. There is not a significant difference between the GAI short form score and the Full Scale IQ score of the WISC-IV.

Subjects

This study used data that was previously collected. The WISC-IV was administered by one of two certified school psychologists employed within public school systems during the 2003-2004 school year. The students that were administered these tests had been referred for a psychoeducational evaluation to help determine if the students qualified for special education services. Informed consent for testing was obtained per state of Ohio Special Education Procedures. A total of thirty one students were administered these tests. Approximately half of the students attended a rural school district and the other half attended an urban school district. The student's age, grade, gender, and test scores were entered into an Excel spreadsheet by the school psychologists that administered the tests. An ID number replaced the students' names on these spreadsheets. The researchers for this project do not have access to the original test protocols or the students' names. (See table 1).

Table 1 **Ages and Gender of Students**

Gender	n	Age (Mean)*	S.D.	Range
Males	18	81.7	13.7	73-168
Females	13	110.6	16.9	92-145
Total	31	114	23.9	73-168

*Age in months

Procedures

Means, Standard Deviations, *T tests of significance*, and *Pearson r correlations* were computed to compare the GAI scores/ FSIQ scores for the 31 subjects.

Results

Hypothesis 1: Does the WISC-IV FSIQ correlate positively with the WISC-IV GAI score? The *Pearson r correlation* ($r = .96$, $p < .01$) suggests there is a strong positive correlation between the GAI and the Full Scale IQ. The overlapping variance is 92%.

. Hypothesis 2: Is there a significant difference between the mean GAI Score and the mean FSIQ? (See table 2). Results show an average difference of 3.74 on the standard score between the GAI (86) and the Full Scale IQ (82.77). The range of difference between the GAI standard score and the FSIQ standard score is -5 to 13. *T tests of significance* (t (30) = 1.0 $p < .01$) show that there is not a significant difference between the scores of the GAI and the Full Scale IQ (accept the null hypothesis).

Table 2 **Means, S. D., and Ranges of Test Scores**

Test	**Mean**	**S. D.**	**Range**
GAI*	86	14.27	53-108
FSIQ**	82.7	13.1	56-110

*GAI= General Ability Index **FSIQ= Full Scale IQ

Discussion/ Recommendations

The GAI as a short form is a good predictor for the Full Scale IQ of the WISC-IV. The GAI standard score was essentially the same score as the Full Scale IQ on the WISC-IV. It is recommended that the GAI be used to predict the Full Scale IQ or make a decision to do further testing, but should not be used to make a decision regarding labeling/ educational service for a child.

This study used a small sample size from a two school systems and should be repeated using a large sample across other regions of the country. It also used a single administration of the WISC-IV to retrieve both the Full Scale IQ score and the GAI score. Future research should examine the predictive validity of the GAI short form when the short form is administered first and then the total test at a later date. More research could provide information regarding the GAI as a short form for different exceptionalities by comparing – FSIQ scores with subjects labeled as gifted, MMI, SLD, and AD/HD.

References

Berthier, N. E. DeBlois, S., Poirier, C. R., Novak, M. A., & Clifton, R. K. (2000). Where's the ball? Two- and three- year-olds reason about unseen events. *Developmental Psychology,* 36(3), 394-401.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, 97(3), 404-431.

Crawford, J. R., Allan, K. M. & Jack, A. M. (1992). Short-forms of the UK WAIS-R: Regression equations and their predictive validity in a general population sample. *British Journal of clinical Psychology*, 31, 191-202.

Cyr, J., and Brooker, B. (1984) Use of appropriate formulas for selecting WAIS-R short forms. *Journal of Consulting and Clinical Psychology*, 52, 903-905.

Donders, J., & Warschausky, S. (1997). WISC-III factor index score pattern after traumatic head injury in children. *Child Neuropsychology*, 3(1), 71-78

Flynn, J. R. (1998). WAIS-III and WISC-III IQ gains in the United States from 1972 to 1995: How to compensate for obsolete norms. *Perceptual and Motor Skills*, 86, 1231-1239.

Fry, A. F., & Hale, S. (1996). Processing Speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science* 7(4), 237-24`.

Hambleton, R. K. (1993). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Assessing adolescent and adult intelligence* (pp. 533-585). Boston: Allyn & Bacon.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Jacob, Susan and Hartshorne, Timothy (2003). *Ethics and Law for School Psychologists. 4th edition.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Kail, R. (2000). Speed of information processing: Developmental change and links to intelligence. *Journal of School Psychology*, 38(1), 51-61.

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology*, 30(6), 949-954.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity, *Acta Psychologica*, 86, 199-225.

Perlow, R., Jattuso, M., & Moore, D. d. (1997). Role of verbal working memory in complex skill acquisition. *Human performance*, 10(3), 283-302.

Prifitera, A., Saklofske, D., Weiss, L. (2005). *WISC-IV: Clinical Use and Interpretation-Scientist-Practioner Perspectives.* Amsterdam: Elsevier Academic press.

Raiford, S., Weiss, L., Rolhus, E., & Coalson, D. (2005). *WISC-IV General Ability Index* (Technical Report No. 4) Retrieved June 1, 2005 from http://harcourtassessment.com/hai/Images?pdf/wisciv/WISCIVTechReport4.pdf

Reynolds, C. R., & Kamphaus, R. W. (1990). *Handbook of psychological & educational assessment of children: Intelligence & achievement.* New York: Guilford press

Sattler, J. M. (1992b). *Assessment of children. WISC-III and WPPSI-R supplement.* San Diego,CA: Author

Sattler. J. M. (2004). *Assessment of Children: WISC-IV and WPPSI-III Supplement.* San Diego: Jerome M. Sattler, Publisher, Inc.

Silverstein, A. B. (1990). Short forms of individual intelligence tests. *Psychological Assessment: a Journal of Consulting and Clinical Psychology,* 2, 3-11.

Swanson, H. L. (1996), Individual and age-related differences in children's working memory. *Memory & Cognition*, 24(1), 70-82.

Thompson, Anthony P., LoBello, Stephen G., Atkinson, Leslie, Chisholm, Vivienne, and Ryan, Joseph J. (2004). Brief Intelligence Testing in Australia, Canada, the United Kingdom, and the United States. *Professional Psychology: Research and Practice*, Vol. 35, Issue 3, 0735- 7028.

Tulsky, D. Saklofske, D., Wilkins, C., Weiss, Cepeda, N. J., Kramer, A. F., & Gonzalez deSather, J. C.M. (2001). Changes in executive control across the life span: examination of task- switching performance. *Developmental Psychology*, 37(5), 715-730

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