Ψ Psychology Press
Taylor & Francis Group

Significant Factor Score Variability and the Validity of the WISC-III Full Scale IQ in Predicting Later Academic Achievement

Miranda E. Freberg and Beverly J. Vandiver

The Pennsylvania State University, University Park, Pennsylvania

Marley W. Watkins

Arizona State University, Tempe, Arizona

Gary L. Canivez

Eastern Illinois University, Charleston, Illinois

The purpose of this study was to investigate the validity of the WISC-III (Wechsler, 1991) Full Scale IQ (FSIQ) scores in predicting later academic achievement given significant variability among any of the four WISC-III factor scores. Taken from an archival data set, the sample was composed of 6- to 13-year-old students who were twice evaluated for special education eligibility over approximately a three-year retest interval. Participants were separated into two groups based on the presence or absence of significant factor score variability and then matched across groups on disability, FSIQ, age, sex, and ethnicity. The results of hierarchical multiple regression analyses indicated that the FSIQ was a valid predictor of academic achievement scores even in the presence of significant factor score variability.

Key words: academic achievement, factor score variability, intelligence, test interpretation

Researchers and practitioners have differing opinions about which IQ scores should be interpreted on intelligence tests as well as which IQ score—global, factor, or subtest scores—most reliably and accurately predicts academic achievement (Sattler, 2001). According to Sattler, psychologists generally agree that the global ability score is the most parsimonious, valid predictor of academic achievement. However, some practitioners have contended that factor and subtest scores provide additional clinically useful information, including profiles of individuals' specific strengths and weaknesses (Donders, 1996; Kaufman, 1994; Kaufman & Lichtenberger, 2000).

One survey of 354 nationally certified school psychologists revealed that 89% of respondents interpreted factor and subtest scores to gain additional information above and beyond the global IQ (Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000).

Despite the popularity of subtest score analysis, current research fails to support this practice (McDermott & Glutting, 1997; Watkins & Canivez, 2004; Watkins & Kush, 1994). Nevertheless, the debate about which scores to interpret continues, particularly in the prediction of academic achievement scores. The current consensus is that the global IQ is the best predictor of academic achievement in the absence of underlying discrepancies among factor and subtest scores (Sattler, 2001). However, scholars disagree about which IQ scores to interpret when significant factor or subtest

Address correspondence to Miranda Freberg, The Pennsylvania State University, 226 CEDAR Building, University Park, PA 16802. E-mail: mee138@psu.edu

discrepancies occur. Some experts (e.g., Kaufman, 1994) argue that in the presence of discrepancy, factor or subtest scores are more accurate representations of an individual's cognitive ability and should be used in the prediction of achievement instead of the global IQ score.

In fact, it has been contended that significant factor or subtest score discrepancies indicate that the global IQ does not accurately represent an individual's overall intellectual functioning and its use leads to inaccurate predictions of achievement (Hale & Fiorello, 2001; Weiss, Saklofske, & Prifitera, 2003). Thus, proponents of this view have argued against the use of the global IQ when factor or subtest variability exists. Sattler (2001) stated that in the presence of such discrepancy "the Full Scale IQ may represent a forced average of rather disparate primary skills" (p. 321). Weiss et al. agreed that the global IQ may not be the best measure of a child's cognitive ability if underlying factor scores differ. For example, they posited that a 37-point discrepancy between factor scores on the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991) may indicate that the overall Full Scale IQ (FSIQ) is less meaningful as a summary of an individual's intellectual ability. Under these circumstances, Weiss et al. recommended that the FSIQ not be reported or only reported with caution. Although accepted by many experts (e.g., Kaufman, 1994; Sattler, 2001), this assumption has not been extensively empirically validated.

SUPPORT FOR THE GLOBAL IQ SCORE

Ryan, Kreiner, and Burton (2002) examined whether significant intersubtest scatter on the Wechsler Adult Intelligence Scale–Third Edition (WAIS-III; Wechsler, 1997a) compromised the predictive validity of the FSIQ in predicting the eight indexes on the Wechsler Memory Scale–Third Edition (WMS-III; Wechsler, 1997b) in a sample of veteran men suffering from a variety of medical and psychiatric disorders with high versus low intersubtest scatter. High intersubtest scatter was quantified as a nine-point difference or greater between the highest and lowest subtest scaled scores whereas low intersubtest scatter was set at less than a nine-point difference. The two groups were matched on FSIQ within one point, with no significant differences found between the two groups on FSIQ, age, years of education, distribution of ethnicities, or diagnoses. Separate regression analyses for the low- and high-scatter groups were used to predict the WMS-III indexes from the WAIS-III FSIQs. Using the Potthoff method to compare intercepts and slopes, no significant differences were found between the low- and high-scatter groups. That is, the FSIQ was found to be a valid predictor of WMS-III index scores even in the presence of intersubtest scatter (Ryan et al., 2002).

Kahana, Youngstrom, and Glutting (2002) used scores obtained on the Differential Ability Scales (DAS; Elliott, 1990) from a national representative sample of 6- to 17-year-old youth to examine (a) the overall frequency of significant discrepancies between factor and subtest scores, and (b) the predictive utility of the factor and subtest discrepancies in forecasting academic achievement. Multiple regression analyses were used to test the incremental validity of the factor and subtest scores. The General Conceptual Ability (GCA) was entered as the first block for all three achievement criteria. The second block consisted of the factor scores, a discrepancy status variable indicating the presence or absence of a significant discrepancy between the factor scores, and the interaction of the factor score and the discrepancy status variable. The third block of each regression model contained the subtest scores, the significant subtest discrepancies, and the interaction between subtest and discrepancy (Kahana et al., 2002). Results indicated that the GCA was the only variable in all of the models tested that contributed a significant amount of unique variance in predicting academic achievement. The discrepancy and interaction variables contributed minimal additional variance (1 to 3.7%) to the prediction. Kahana et al. concluded that the GCA was the most parsimonious and robust predictor of all three achievement criteria measured by the DAS.

Glutting, Youngstrom, Ward, Ward, and Hale (1997) investigated whether the Verbal IQ (VIQ), the Performance IQ (PIQ), or any of the four factor scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed) on the WISC-III improved the prediction of academic achievement above and beyond the FSIQ score among a nationally representative sample of 283 nonreferred children and a sample of 636 children referred for an evaluation. A series of hierarchical regressions was conducted to determine the relative contributions of the FSIQ, the VIQ, the PIQ, and the four underlying factor scores in predicting achievement (Reading, Math, Writing, and Language), as measured by the Wechsler Individual Achievement Test (WIAT; Wechsler, 1992). FSIQ was always entered into the regression equation first, followed either by the four factor scores or by the VIQ and PIQ. In every regression equation, the FSIQ contributed most substantially to the prediction of academic achievement. As a group, the four factor scores accounted for an additional 5–16% of the variance beyond what the FSIQ contributed to the prediction of the achievement criteria. Separately each factor score uniquely contributed 0–5% of the variance. Together, the VIQ and PIQ accounted for 2.1–4.6% of the variance above and beyond the FSIQ, and neither the VIQ nor the PIQ accounted for any significant unique variance. The researchers concluded

that the FSIQ was the most parsimonious and powerful predictor of academic achievement and that factor scores produced negligible increases in the prediction of all achievement criteria.

SUPPORT FOR FACTOR SCORE ANALYSIS

Hale, Fiorello, Kavanagh, Hoeppner, and Gaither (2001) contended that Glutting et al.'s (1997) methodology produced misleading results. They argued that the FSIQ on the WISC-III was not an accurate measure of intellectual functioning for all children, especially those demonstrating a significant discrepancy between factor scores. The researchers used an archival data set covering a six-year-period to compare WISC-III factor and FSIQ scores in predicting academic achievement in a sample of children who met diagnostic criteria for learning disabilities.

WISC-III scores were used to predict scores on various achievement measures, including subtests from the WIAT, the Woodcock-Johnson Tests of Achievement–Revised (Woodcock & Johnson, 1989), and the Wide Range Achievement Test–2 (Wilkinson, 1993). Hale et al. (2001) hypothesized that factor scores would account for more academic achievement variance than the FSIQ. They further predicted that the FSIQ would consist mostly of unique, not shared, variance. Similar to the study conducted by Glutting et al. (1997), Hale et al. used hierarchical regression analysis to test their first hypothesis. However, their methodology differed in that they entered the factor scores into the regression equation first, followed by the FSIQ. This decision regarding the order of variable entry was based on the results of a commonality analysis (see Pedhazur, 1997), which suggested that 71.7% of FSIQ variance was uniquely accounted for by the factor scores. Results from the hierarchical regression analysis revealed that for all achievement measures, factor scores accounted for a large portion of variance (22.5–34%), whereas FSIQ accounted for very little additional variance (.2–1.5%). Based on their findings, Hale et al. concluded that FSIQ is "factorially complex" (p. 42), consisting of various skills and abilities as opposed to a single underlying general intelligence construct. Their findings did not support the interpretation of the global FSIQ over factor scores for the learning disabled population or for any other population demonstrating significant variability/scatter among factor scores.

In a later study, Fiorello et al. (2007) attempted to replicate and extend the findings of the Hale et al. (2001) study to the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler, 2003). The researchers used commonality analysis to determine unique and shared variance in the prediction of FSIQ scores for a clinical sample of students diagnosed with a learning disability (LD), attention-deficit/hyperactivity disorder (ADHD), or traumatic brain injury (TBI). Similar to the Hale et al. study, Fiorello et al. found that across disability groups there existed mostly unique variance and little shared variance among factor scores, thus lending support for interpretation of factor scores.

In response to various authors' support of the interpretation of global intelligence over factor scores (Dana & Dawes, 2007; Faust, 2007; Watkins, Glutting, & Lei, 2007), Hale, Fiorello, Kavanagh, Holdnack, and Aloe (2007) argued that profile analysis is especially useful and relevant for clinical samples of children who demonstrate variable scores on intelligence measures. Accordingly, Hale et al. recommend more research with students with disabilities.

CURRENT STUDY

The current study investigated whether global IQ scores are valid long-term predictors of achievement in the presence of significant factor score variability. Specifically, this study compared the predictions of the FSIQ score on the WISC-III for participants demonstrating a significant discrepancy between any of the four factor scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed) and for participants who did not exhibit such a discrepancy. In contrast to previous studies that have used concurrent achievement data, and in order to address hypotheses that cognitive strengths and weaknesses unfold over the course of development, the current study included a longitudinal analysis. Also, in accordance with the recommendations of Hale et al. (2007) for more research with students with disabilities, this study specifically examined a clinical sample. The research that has been conducted in this area has resulted in contradictory findings. It is important that these contradictions be resolved so that psychologists can make accurate, well-founded decisions and provide effective services to children.

Previous research on the effect of factor score discrepancy on the predictive validity of the FSIQ has revealed that different methodologies lead to different results (Glutting et al., 1997; Hale et al., 2001). The methodology used in the current study is consistent with that which is presently used by practitioners in the field when interpreting the results of intelligence tests: The FSIQ is considered before the factor scores (Sattler, 2001). As a result, it was hypothesized that the presence of significant factor score variability would not undermine the predictive ability of the FSIQ. This hypothesis was tested with students matched based on FSIQ and factor score variability categorized as present or absent (factor profile status).

METHOD

Participants

Participants were a subset of a sample of children who participated in a long-term study on the stability of WISC-III scores (Canivez & Watkins, 1998). Participants were evaluated twice for special education eligibility over a period of three years. The original sample consisted of 667 participants. After listwise deletion of cases that were missing information critical to this study (e.g., WISC-III FSIQ or factor scores, reading or math achievement composites, disability status), 202 participants (138 males, 64 females) remained. Participants included children in kindergarten through eighth grade who ranged in age from 6 to 13 years ($M = 8.67$, $SD = 1.79$). A majority of the participants were Caucasian (78.7%), with a smaller percentage of participants classified as Hispanic/Latino (10.4%), African American (8.4%), Native American (1.5%), Asian/Pacific Islander (.5%), and Other (.5%). Participants were classified by multidisciplinary evaluation teams according to local regulations into one of four disability categories: specific learning disability (74%), serious emotional disturbance (6%), mental retardation (12%), and other (8%).

Instruments

Wechsler Intelligence Scale for Children–Third Edition

The WISC-III is an individually administered instrument for assessing the cognitive ability of children ages 6 years through 16 years, 11 months. The WISC-III consists of 13 subtests, which are grouped together to produce three composite IQ scores (i.e., Verbal IQ, Performance IQ, and Full Scale IQ) and four factor index scores (i.e., Verbal Comprehension Index, Perceptual Organization Index, Freedom from Distractibility Index, and Processing Speed Index). The WISC-III was standardized on a nationally representative sample of 2,200 children in accordance with the 1998 United States Census. The WISC-III manual presents further information on the standardization sample and procedures as well as evidence of reliability and validity of the scores (Wechsler, 1991). In addition, several studies have investigated the psychometric properties of the WISC-III (e.g., Canivez, Neitzel, & Martin, 2005; Canivez & Watkins, 1998; Edelman, 1996; Sapp, Abbott, Hinckley, & Rowell, 1997; Slate, 1995; Zimmerman & Woo-Sam, 1997).

Academic Achievement Measures

Various achievement measures demonstrating adequate score reliability and validity were used. Reading and math composite scores obtained from the WIAT, the Woodcock-Johnson Revised Tests of Achievement (WJ-R; Woodcock & Johnson, 1989), and the Kaufman Test of Educational Achievement (K-TEA; Kaufman & Kaufman, 1985) served as the criterion variables in the analyses. Although several different measures of reading and math achievement were used, these composite scores are considered relatively equivalent or at least comparable, as they serve the same purpose and demonstrate adequate reliability and validity (Sattler, 2001).

Procedure

A random sample of school psychologists ($n = 2,000$) from the membership of the National Association of School Psychologists (NASP) was asked to provide test scores and demographic data from recent special education reevaluations. The only criterion given for submitted reevaluation data was two administrations of the WISC-III for the same case. The mean retest intervals for the Specific Learning Disability (SLD), Serious Emotional Disturbance (SED), and Mental Retardation (MR) groups were 2.87 years ($SD = .39$), 2.81 years ($SD = .45$), and 2.90 years ($SD = .49$), respectively. One hundred forty-five school psychologists participated, constituting a 7.25% response rate. They provided an average of 4.6 cases each, with a range from 1 to 25 cases. A full description of this sample was provided by Canivez and Watkins (1998).

Participants were separated into two groups based on the presence or absence of a significant discrepancy between any of the four factor scores on the WISC-III. The flat profile group included 44 participants with a nonsignificant discrepancy between any of the four WISC-III factor scores. Participants assigned to the discrepant profile group ($n = 158$) demonstrated a statistically significant discrepancy between any of the four factor scores. Statistically significant factor discrepancies were determined based on Table B.1 in the WISC-III manual (Wechsler, 1991, p. 261). Values required for statistical significance were taken from the "All Ages" category ($p < .05$).

After dividing the sample into discrepant and nondiscrepant groups, 42 of the participants from the flat profile group were matched with 42 participants from the discrepant profile group on subsequent characteristics in the following order: disability, FSIQ, age, sex, and ethnicity. Efforts were made to match the groups as closely as possible based on the available data. Two participants in the flat profile group were unable to be matched due to extreme FSIQ scores. Thus, the matching process reduced the sample size from 202 to 84 participants (55 males, 29 females). Because categorical variables cannot be entered directly into a multiple

regression model and meaningfully interpreted, the two matched profile groups were dummy coded. Dummy coding transforms categorical data into quantitative form so that the categorical variables can be used as predictors in multiple regression (Cohen, Cohen, West, & Aiken, 2003). In dummy coding, $N - 1$ dummy variables are computed, with N representing the number of groups. The first group is coded 1 on the dummy variable and the remaining group(s) is coded 0. In this case, there were two groups, the flat profile group and the variable profile group, so one dummy variable was created. For these analyses, 1 was used to represent the presence of a factor score discrepancy and 0 to represent the absence of a factor score discrepancy.

Similar to the larger sample, the participants in the matched sample ranged in age from 6 to 13 ($M = 8.61$, $SD = 1.66$). Similar patterns were also maintained for ethnicity and disability status. A majority of the participants were Caucasian (78.6%), followed by Hispanic/Latino (13.1%), African American (7.1%), and Native American (1.2%). Eighty percent of participants in the matched sample were classified as having a specific learning disability, with the remaining participants classified as having a serious emotional disturbance (5%), mental retardation (8%), or other disability (7%). Hierarchical regression analyses were used to examine the predictive validity of the FSIQ across both groups using IQ scores from first testing and academic achievement scores from second testing approximately three years later.

In addition to statistically significant differences between factor scores, clinically significant differences were also investigated. Clinically significant discrepancies were determined based on Table B.2 in the WISC-III manual (Wechsler, 1991, p. 262), which delineates the prevalence of factor score discrepancies in the general population. Subgroups were developed

from the matched discrepant and flat profile groups based on prevalence rates of 15%, 10%, 5%, and 1%. The first subgroup consisted of cases from the discrepant profile group in which the differences between the factor scores were found in 15% of the population ($n = 31$) as well as their respective matches from the flat profile group ($n = 31$). Thus the total sample size was reduced from 84 to 62 for this particular analysis. Subgroups were developed in the same manner for the 10% prevalence level ($n = 40$; 20 discrepant, 20 flat), 5% prevalence level ($n = 20$; 10 discrepant, 10 flat), and 1% prevalence level ($n = 4$; 2 discrepant, 2 flat). Due to the small sample size and subsequent decreased power at the 5% and 1% prevalence levels, hierarchical regression analyses were conducted for only the 15% and 10% prevalence levels.

RESULTS

Descriptive Statistics

Descriptive statistics for the WISC-III FSIQ, factor scores, and reading and math achievement scores are presented in Table 1. The mean IQ and achievement scores of this sample are slightly below average ($M = 100$), which is expected given that this is a clinical sample of individuals with disabilities (Kavale & Nye, 1985–1986). Correlations between all variables were statistically significant ($p < .01$), with coefficients ranging from .40 to .90 ($Mdn = .65$). Table 2 shows descriptive statistics for FSIQ and age based on profile status. The discrepant (significant factor discrepancy) and flat (no significant factor discrepancy) profile groups did not differ significantly on FSIQ, $t(82) = .06$, $p = .96$, $d = .01$; reading achievement, $t(82) = .65$, $p = .52$, $d = .14$; math achievement, $t(82) = .06$, $p = .95$, $d = .01$;

TABLE 1
Descriptive Statistics for WISC-III, Reading, and Math Scores ($N = 202$)

Variable	M	SD	Correlations[a]						
			1	2	3	4	5	6	7
WISC-III									
1. FSIQ	89.41	15.76	—	.90	.89	.77	.66	.65	.75
2. VC	91.12	15.75		—	.65	.73	.44	.62	.69
3. PO	92.51	16.34			—	.57	.59	.50	.63
4. FD	85.39	15.31				—	.49	.64	.66
5. PS	92.58	16.70					—	.40	.49
Achievement									
6. Reading	85.73	15.40						—	.67
7. Math	88.15	16.32							—

Note. FSIQ = Full-Scale IQ, VC = Comprehension, PO = Perceptual Organization, FD = Freedom from Distractibility, PS = Processing Speed. All correlations were significant at the $p < .01$ level.
[a] All IQ score correlations are concurrent. IQ-Achievement score correlations are longitudinal.

TABLE 2
Means and Standard Deviations for Profile Groups on WISC-III FSIQ and Age

	Flat Profile		Discrepant Profile	
Variable	M	SD	M	SD
	Statistically Significant ($n = 84^a$)			
FSIQ	88.83	12.00	88.98	11.82
Age in years	8.71	1.73	8.50	1.60
	15% Prevalence ($n = 62$)			
FSIQ	88.03	12.58	88.13	12.32
Age in years	8.77	1.59	8.52	1.73
	10% Prevalence ($n = 40$)			
FSIQ	86.15	13.32	86.35	13.11
Age in years	8.50	1.47	8.40	1.60

[a] Statistical significance was calculated based on discrepancies reported in Table B.1 on page 261 in the WISC-III manual (Wechsler, 1991). Factor index difference scores required for significance were taken from the "All Ages" category according to a .05 level of significance.

age, $\chi^2(7, N = 84) = 2.511$, $p = .93$; gender, $\chi^2(1, N = 84) = .05$, $p = .82$; ethnicity, $\chi^2(3, N = 84) = 1.82$, $p = .61$; or disability classification, $\chi^2(3, N = 84) = 4.28$, $p = .23$.

The basic assumptions of multiple regression (normality, linearity, and homoscedasticity of residuals) were tested and met. Independence of errors was assumed because the Durbin-Watson statistic was within the recommended range of one to three (Field, 2000). Multicollinearity, examined through the Variance Inflation Factor (VIF), was not deemed present as the VIF values were within the recommended range (Bowerman & O'Connell, 1990).

Factor Profile Status

It was posited that the presence of a significant factor score discrepancy between any of the four factor scores would not invalidate the prediction of reading and math achievement by the FSIQ. FSIQ scores from first testing were used to predict reading and math achievement scores at second testing (around three years later). A total of six hierarchical regression analyses were conducted to test this hypothesis: two to investigate statistically significant differences (one for reading achievement and one for math achievement) and four to investigate clinically significant differences (two at the 15% prevalence level and two at the 10% prevalence level). For all analyses, FSIQ score was entered in the first step, profile status (flat vs. discrepant) in the second step, and the interaction between FSIQ and profile status in the third step. In accordance with recommendations by Cohen et al. (2003), FSIQ scores were centered to make interpretation more meaningful. Without centering, the constant of the regression equation would be the value of the criterion when the FSIQ is equal to zero. An FSIQ of zero is both

impossible and meaningless. As such, the FSIQ scores were centered so that the constant of the regression equation becomes the expected value of the criterion when the FSIQ is at the mean value of the data, which allows for more meaningful interpretation. Due to the number of tests included in each analysis, the criteria for interpretation of statistical significance were based on a fixed alpha level of .01 to minimize Type I error.

Statistically Significant Factor Scores

Two hierarchical regression analyses were conducted to investigate the effect of statistically significant factor discrepancies in predicting reading and math achievement scores. Reading achievement served as the dependent variable in the first analysis and math achievement served as the dependent variable in the second analysis. Unstandardized and standardized regression coefficients, standard errors, and the proportion of variance accounted for (R^2 and R^2 change) by FSIQ and factor profile status are reported in Table 3 for the prediction of reading achievement and in Table 4 for the prediction of math achievement.

FSIQ was statistically significant and accounted for 44% of the variance in the prediction of reading achievement. However, neither factor discrepancy status nor the interaction between FSIQ and factor discrepancy status were statistically significant predictors, accounting for an additional 0.5% and 0.9% of the variance, respectively. A similar pattern of results was found in the prediction of math achievement scores, in that the FSIQ was the only statistically significant contributor to the prediction of math achievement ($R^2 = .48$), whereas factor profile status ($R^2 = .001$) and the interaction between FSIQ and profile status ($R^2 = .003$) were not statistically significant.

TABLE 3
Summary of Hierarchical Regression Analysis for the Prediction of Reading Achievement by FSIQ and Profile Status

Variable	Statistically Significant (n = 84)			15% Prevalence (n = 62)			10% Prevalence (n = 40)		
	B	SE B	β	B	SE B	β	B	SE B	β
Model 1	$R^2 = .441$			$R^2 = .436$			$R^2 = .351$		
Centered FSIQ	0.79	0.10	0.66*	0.72	0.11	0.66*	0.59	0.13	0.59*
Model 2	$R^2\Delta = .005$			$R^2\Delta = .019$			$R^2\Delta = .004$		
Centered FSIQ	0.79	0.10	0.66*	0.72	0.10	0.66*	0.59	0.13	0.59*
Profile Status	1.88	2.31	0.07	3.71	2.56	0.14	1.63	3.38	0.06
Model 3	$R^2\Delta = .009$			$R^2\Delta = .002$			$R^2\Delta = .003$		
Centered FSIQ	0.90	0.14	0.76*	0.76	0.15	0.70*	0.64	0.19	0.64*
Profile Status	1.85	2.31	0.07	3.62	2.58	0.14	1.34	3.49	0.05
FSIQ × Profile Status	−0.23	0.20	−0.14	−0.09	0.21	−0.06	−0.11	0.27	−0.08

Note. $R^2\Delta = R^2$ change. FSIQ (Full-Scale IQ) was centered by subtracting the sample mean from each FSIQ score. Profile status was dummy coded with the presence of significant factor score discrepancy coded as 1 and the absence of such discrepancy coded as 0.
*$p < .01$.

Clinically Significant Factor Scores

Four hierarchical regression analyses were conducted to investigate clinically significant differences between factor scores in the prediction of reading and math achievement scores. For reading and math achievement at the 15% prevalence level, the FSIQ contributed significantly to the prediction of reading ($R^2 = .44$) and math ($R^2 = .48$) achievement scores, whereas factor profile status and the interaction between FSIQ and factor profile status were not significant incremental predictors, with R^2 values ranging from .001 to .019. A similar pattern was found at the 10% prevalence level with FSIQ as the only statistically significant predictor of reading ($R^2 = .35$) and math ($R^2 = .50$) achievement. Again, factor profile status and the interaction between FSIQ and factor profile status did not add significantly to the prediction of either reading or math achievement, with R^2 values ranging from 0 to .009. Unstandardized and standardized regression coefficients, standard errors, and the proportion of variance accounted for (R^2 and R^2 change) at the 15% and 10% prevalence levels are reported in Table 3 for the prediction of reading achievement and Table 4 for the prediction of math achievement.

DISCUSSION

The purpose of this study was to investigate whether significant discrepancies between factor scores render the WISC-III FSIQ invalid as a long-term predictor of academic achievement scores in reading and math for students enrolled in special education classes. It was hypothesized that the presence of significant factor score variability would not undermine the predictive ability of the FSIQ. The findings support this hypothesis. The FSIQ was found to be the only statistically

TABLE 4
Summary of Hierarchical Regression Analysis for the Prediction of Math Achievement by FSIQ and Profile Status

Variable	Statistically Significant (n = 84)			15% Prevalence (n = 62)			10% Prevalence (n = 40)		
	B	SE B	β	B	SE B	β	B	SE B	β
Model 1	$R^2 = .442$			$R^2 = .483$			$R^2 = .502$		
Centered FSIQ	0.79	0.10	0.67*	0.81	0.11	0.70*	0.83	0.13	0.71*
Model 2	$R^2\Delta = .000$			$R^2\Delta = .001$			$R^2\Delta = .000$		
Centered FSIQ	0.79	0.10	0.67*	0.81	0.11	0.70*	0.83	0.14	0.71*
Profile Status	−0.30	2.32	−0.01	0.71	2.67	0.03	−0.49	3.50	−0.02
Model 3	$R^2\Delta = .003$			$R^2\Delta = .003$			$R^2\Delta = .009$		
Centered FSIQ	0.72	0.14	0.61*	0.75	0.15	0.64*	0.72	0.19	0.61*
Profile Status	−0.29	2.33	−0.01	0.83	2.69	0.03	0.13	3.60	0.00
FSIQ × Profile Status	0.14	0.20	0.08	0.13	0.22	0.08	0.23	0.27	0.14

Note. $R^2\Delta = R^2$ change. FSIQ (Full-Scale IQ) was centered by subtracting the sample mean from each FSIQ score. Profile status was dummy coded with the presence of significant factor score discrepancy coded as 1 and the absence of such discrepancy coded as 0.
*$p < .01$.

significant predictor for both reading and math achievement scores, whereas factor profile status was not found to be a statistically significant contributor to the prediction of reading and math achievement scores. The present study's findings are consistent with and extend research in the field that suggests that the global IQ is a valid predictor of academic achievement scores despite significant factor score discrepancy (Glutting et al., 1997; Kahana et al., 2002). The results of the current study indicate that even with the presence of significant factor score discrepancy, the WISC-III FSIQ demonstrated validity as a predictor of academic achievement scores in reading and math.

Limitations

There are several limitations to the current study. As noted in Canivez and Watkins (1998), two limitations of the data set used in this study include the low response rate (7%) of school psychologists and the ineligibility of some students for participation in the study. Only 145 of the 2,000 school psychologists who were randomly sampled from NASP provided data. Additionally, students who were no longer enrolled in special education, did not require reevaluation, or were unavailable for reevaluation were not included in the study, which limits generalization.

Another limitation is the use of the WISC-III, which has been replaced by the WISC-IV. Wechsler (2003) indicated that although the WISC-IV contains new nomenclature and a slightly altered structure, the abilities assessed and summarized by the WISC-III factor scores and WISC-IV composite indices remain fairly consistent between the two versions of the scale. The corrected correlation coefficients between the two instruments for the factor/composite scores range from .72 to .88 and the FSIQ-FSIQ correlation is .89 (Wechsler). Thus, it is hypothesized that the current results would be replicated by studies with the WISC-IV.

A further limitation is the use of a variety of measures to tap academic achievement. Although composites of achievement measures are comparable in purpose and outcome, and all have been judged to be reliable and valid measures (Sattler, 2001), it is not clear whether subtle differences in academic achievement measures may have affected the results. The matching procedure also produces limitations. Although efforts were made to match cases on all relevant variables, this process did not result in exact matches on all variables. In addition, the matching procedure resulted in a somewhat attenuated range of FSIQ scores, as those individuals with extreme FSIQ scores on either end of the continuum were eliminated from the analysis, due to the absence of an equivalent case in the other group.

The relatively small sample size for some of the regression analyses is another limitation. Due to small sample size, analyses could not be conducted for factor discrepancies occurring in 5% ($n = 20$) and 1% ($n = 4$) of the population. In addition, power was low for the analyses due to small sample sizes and the categorical treatment of continuous data.

Based on the sample sizes used in this study, expected incremental effect sizes were not found in all cases. For all sample sizes, the FSIQ had a large enough effect to be detected based on the criteria set forth by Cohen et al. (2003). However, the sample sizes used in this study may not have been adequate to detect small incremental differences when factor score discrepancy and interaction variables were entered into the model. Based on a power level of .80 and an alpha of .01, the smallest effect sizes that could be detected for significant factor score tests for sample sizes of 84, 62, 40, and 202 were .17, .24, .38, and .07, respectively. Despite sample size differences, analyses indicated that the FSIQ accounted for most of the variance in the prediction of reading and math achievement. The contribution of the factor discrepancy scores was always minimal in comparison to that of the FSIQ.

CONCLUSION

The results of the current study provide support for the FSIQ as a valid and powerful long-term predictor of academic achievement scores, even in the presence of significant and rare factor variability. Due to the small sample size and sole focus on the WISC-III in this study, future research is recommended with larger sample sizes and other cognitive ability measures to test the generalizability of the findings of the current study. Although further research is needed to investigate the replicability of the current results with the WISC-IV and other cognitive ability measures, the results of this study support the use of the FSIQ as a predictor of reading and math achievement regardless of underlying factor score variability. Therefore, despite recommendations to interpret the FSIQ with caution in the presence of significant factor score discrepancy (Kaufman, 1994; Weiss, Saklofske, & Prifitera, 2003), the current study provides further support to the body of evidence that says otherwise (Glutting et al., 1997; Kahana et al., 2002): Factor score discrepancy does not invalidate the FSIQ as the best predictor of academic achievement scores for students diagnosed with learning disabilities, emotional disturbance, or mental retardation.

REFERENCES

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed.). Belmont, CA: Duxbury.

Canivez, G. L., Neitzel, R., & Martin, B. E. (2005). Construct validity of the kaufman brief intelligence test, Wechsler Intelligence Scale for Children–Third Edition, and adjustment scales for children and adolescents. *Journal of Psychoeducational Assessment*, *23*, 15–34.

Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children–Third Edition. *Psychological Assessment*, *10*, 285–291.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Dana, J., & Dawes, R. (2007). Comment on fiorello et al. "Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?" *Applied Neuropsychology*, *14*, 21–25.

Donders, J. (1996). Cluster subtypes in the WISC-III standardization sample: Analysis of factor index scores. *Psychological Assessment*, *8*, 312–318.

Edelman, S. (1996). A review of the Wechsler Intelligence Scale for Children–Third Edition (WISC-III). *Measurement and Evaluation in Counseling and Development*, *28*, 219–224.

Elliott, C. D. (1990). *Differential ability scales*. New York: Harcourt Brace Jovanovich.

Faust, D. (2007). Some global and specific thoughts about some global and specific issues. *Applied Neuropsychology*, *14*, 26–36.

Field, A. (2000). *Discovering statistics using SPSS for windows*. London: Sage Publications.

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology*, *14*, 2–12.

Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of WISC-III factor scores in predicting achievement: What do they tell us? *Psychological Assessment*, *9*, 295–301.

Hale, J. B., & Fiorello, C. A. (2001). Beyond the academic rhetoric of "g": Intelligence testing guidelines for practitioners. *The School Psychologist*, *55*, 113–139.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly*, *16*, 31–55.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Holdnack, J. A., & Aloe, A. M. (2007). Is the demise of IQ interpretation justified? A response to special issue authors. *Applied Neuropsychology*, *14*, 37–51.

Kahana, S. Y., Youngstrom, E. A., & Glutting, J. J. (2002). Factor and subtest discrepancies on the differential ability scales: Examining prevalence and validity in predicting academic achievement. *Assessment*, *9*, 82–93.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.

Kaufman, A. S., & Kaufman, N. (1985). *The Kaufman test of educational achievement*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. New York: Wiley.

Kavale, K. A., & Nye, C. (1985–1986). Parameters of learning disabilities in achievement, linguistic, neuropsychological, and social/behavioral domains. *Journal of Special Education*, *19*, 443–458.

McDermott, P., & Glutting, J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests—or, more illusions of meaning? *School Psychology Review*, *26*, 163–176.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Harcourt Brace.

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, *15*, 376–385.

Ryan, J. J., Kreiner, D. S., & Burton, D. B. (2002). Does high scatter affect the predictive validity of WAIS-III IQs? *Applied Neuropsychology*, *9*, 173–178.

Sapp, G. L., Abbott, G., Hinckley, R., & Rowell, A. (1997). Examination of the validity of the WISC-III with urban exceptional students. *Psychological Reports*, *81*, 1163–1168.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego: Jerome M. Sattler.

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

Watkins, M. W., Glutting, J. J., Lei, P. (2007). Validity of the Full-Scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, *14*, 13–20.

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. *Psychological Assessment*, *16*, 133–138.

Watkins, M. W., & Kush, J. C. (1994). Wechsler subtest analysis: The right way, the wrong way, or no way? *School Psychology Review*, *23*, 640–651.

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

Wechsler, D. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale–Third Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale–Third Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition technical and interpretive manual*. San Antonio, TX: Psychological Corporation.

Weiss, L. G., Saklofske, D. H., & Prifitera, A. (2003). Clinical interpretation of the Wechsler Intelligence Scale for Children–Third Edition (WISC-III) index scores. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed., pp. 115–146). New York: The Guilford Press.

Wilkinson, G. (1993). *Wide Range Achievement Test–Second Edition*. Wilmington, DE: Jastak Associates.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psychoeducational Battery–Revised*. Allen, TX: DLM.

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*, 531–546.