Psychology in the Schools, Vol. 37(4), 2000

DIAGNOSTIC UTILITY OF THE NUMBER OF WISC-III SUBTESTS DEVIATING FROM MEAN PERFORMANCE AMONG STUDENTS WITH LEARNING DISABILITIES

MARLEY W. WATKINS AND FRANK C. WORRELL

The Pennsylvania State University

This paper examined the diagnostic utility of subtest variability, as represented by the number of subtests that deviate from examinees' mean IQ scores, for identifying students with a learning disability (LD). Participants consisted of the 2,200 students in the WISC-III normative sample and 684 students (*Mdn* grade = 5; *M* age = 10.8) identified as LD. The number of subtests deviating from examinees' Verbal, Performance, and Full Scale IQ by ±3 points for normative and exceptional samples were contrasted via Receiver Operating Curve (ROC) analyses. Results indicated that LD students did not differ from normative sample children at levels above chance. It was concluded that deviation of individual subtest scores from mean IQ scores has no diagnostic utility for hypothesizing about students with learning disabilities. © 2000 John Wiley & Sons, Inc.

School psychologists often use the variability of an individual's subtest scaled scores on the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) as an indicator of neurological functioning (Arizona Department of Education, 1992) or learning disability (Banas, 1993; Kellerman & Burry, 1997; Mayes, Calhoun, & Crowell, 1998). The perceived clinical importance of subtest variability, or scatter, is reflected by the inclusion of a subtest scatter table in the WISC-III manual accompanied by a statement that subtest scatter is "frequently considered as diagnostically significant" (Wechsler, 1991, p. 177). Additionally, many books on test interpretation provide tables and directions for interpretation of subtest scatter (e.g., Cooper, 1995).

Currently, subtest scatter is quantified in three ways. The first method involves examining the range (i.e., the difference between an examinee's highest and lowest subtest scaled scores). The second method involves examining the variance using the variance formula applied to the subtest scores of an individual examinee. Finally, researchers look at the number of subtests differing from the individual examinee's mean score by ±3 points (Schinka, Vanderploeg, & Curtiss, 1997). Research on range and variance scatter with previous Wechsler tests has been unproductive (Sattler, 1992), and WISC-III range and variance scatter indices have also been unable to exhibit adequate diagnostic utility for students with learning disabilities (Daley & Nagle, 1996; Kline, Snyder, Guilmette, & Castellanos, 1993; Mayes et al., 1998). However, the third metric, number of subtests deviating from an individual examinee's mean scores, has not been adequately tested. Consequently, the present study was conducted to test the diagnostic utility of this WISC-III variability index in a large sample of children with learning disabilities.

Method

Participants

Learning disabilities sample. Potential participants consisted of all students who received comprehensive psychoeducational evaluations in three southwestern U.S. suburban school districts during one school year. Participants were selected from special education records based upon two criteria: (a) their cognitive assessment included the 10 mandatory subtests of the WISC-III; (b) their placement status was in a learning disability (LD) program.

Correspondence to: Marley W. Watkins, Department of Educational and School Psychology and Special Education, The Pennsylvania State University, 227 CEDAR Building, University Park, PA 16802. E-mail: mww10@psu.edu.

Students' special education eligibility and placement decisions were determined by multidisciplinary evaluation teams, following assessment by a state-certified school psychologist. Teams followed state special education regulations in which a learning disability was defined by a significant ability-achievement discrepancy (Arizona Department of Education, 1990).

These two selection criteria identified 684 students who were enrolled in kindergarten through 11th grade. Demographic characteristics of this sample are provided in Table 1. On average, these students were more likely to be male and younger than the WISC-III standardization sample. Reflective of their geographic origin, there were proportionately more students of Native American and Hispanic ancestry than found in the United States as a whole.

Academic achievement levels in reading and math were measured with the Woodcock Johnson Tests of Achievement-Revised (WJ-R; Woodcock & Mather, 1989) for 94% of the participants. The academic achievement of the remaining students was assessed with six other achievement tests (e.g., WIAT, KTEA, WRAT, etc.). Mean reading and math achievement scores did not differ between WJ-R and other achievement measures (reading $t(641) = -.28$, $p = .78$; math $t(641) = -.90$, $p = .37$). Table 2 presents summary intellectual and academic achievement scores for participating students. Although lower than average, the cognitive and academic achievement levels of the sample participants are consistent with other compilations of data from children enrolled in special education programs in terms of age, sex, IQ, and achievement levels (Kavale & Nye, 1985). Additionally, ability–achievement discrepancies calculated by subtracting achievement scores from FSIQ reflect underachievement in all academic areas.

Reading disabled subsample. A subsample of participants was identified to allow specialized analyses for students with unambiguous specific reading disabilities. Selection criteria included (a) identification as learning disabled in reading by a multidisciplinary evaluation team following state special education regulations in which a reading disability was defined by a significant ability-reading achievement discrepancy, (b) WISC-III Full Scale IQ (FSIQ) greater than 84, (c) FSIQ-reading achievement discrepancy greater than 14 points, (d) FSIQ-math achievement discrepancy less than 15 points, and (e) not identified as learning disabled in math by a multidisciplinary evaluation team. Thus, selection criteria combined those embedded in state and federal regulations (Arizona Department of Education, 1990) as well as those used in contemporary research on learning disabilities (Vellutino, Scanlon, & Tanzman, 1991) and proposed by reading experts (Stanovich, 1999).

Table 1
Demographic Characteristics of Learning Disabled (10 Subtests), Reading Disabled (10 Subtests), Learning Disabled (12 Subtests), and WISC-III Normative Samples

	Learning Disabled (10 subtests)	Reading Disabled (10 subtests)	Learning Disabled (12 subtests)	Standardization (12 subtests)
N	684	131	332	2,200
Mdn Grade	5	3	4.5	—
M Age	10.8	9.4	10.3	11.5
Sex				
Male	73%	75%	70%	50%
Female	27%	25%	30%	50%
Ethnicity				
White	60.1%	72.5%	48.8%	70%
Native American	18.3%	12.2%	23.8%	—
Hispanic	13.6%	11.5%	16.3%	11%
Black	7.7%	3.8%	10.8%	15%

Table 2
Summary of WISC-III and Academic Achievement Standard Scores, Discrepancy Between Ability and Achievement, and Average Number of Subtests Deviating From Mean Scores for Learning Disabled and Standardized Samples

	Learning Disabled (10 subtests)[a]		Reading Disabled (10 subtests)[b]		Learning Disabled (12 subtests)[c]		Standardization[d] (10/12 subtests)	
Measure	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
VIQ	90.7	11.8	96.8	8.9	90.4	12.0	100	15
PIQ	94.3	13.7	102.6	10.2	93.4	13.3	100	15
FSIQ	91.6	12.0	99.3	8.4	90.9	11.5	100	15
Reading achievement	81.3	13.8	75.2	8.9	80.7	13.5	—	—
Math achievement	84.9	14.2	96.4	10.4	84.6	14.8	—	—
Reading discrepancy	10.2	15.2	24.1	7.6	10.3	14.9	—	—
Math discrepancy	6.7	11.7	2.9	8.1	6.3	12.5	—	—
No. Verbal	.57	.79	.59	.77	.86	1.0	.54/.85	.77/.95
No. Perform.	.98	1.0	1.1	1.1	1.3	1.1	1.04/1.3	1.01/1.1
No. Total	2.0	1.5	2.0	1.4	2.7	1.6	1.95/2.5	1.4/1.6

[a]$N = 684$. [b]$n = 131$. [c]$n = 332$. [d]$N = 2{,}200$.

These criteria selected 131 students from the larger sample of children classified as learning disabled. Their demographic characteristics are provided in Table 1, and their mean cognitive and achievement scores are presented in Table 2. Like the Learning Disabilities sample, there was an overrepresentation of boys and students of Native American ancestry. Additionally, these students were somewhat younger than the other samples. Whereas the general learning disabilities group was marked by FSIQ-reading and FSIQ-math discrepancies of 10.2 and 6.7 points, respectively, the specific reading disabilities subsample had discrepancies in reading and math for 24.1 and 2.9 points, respectively.

Twelve-subtest subsample. A third subsample of participants was formed based upon completion of the two optional WISC-III subtests. This subsample allowed analysis of the incremental influence of the Digit Span and Symbol Search subtests on scatter. With the addition of this criterion, 332 students were identified. Their demographic characteristics are presented in Table 1 and summary cognitive and academic statistics are provided in Table 2. Boys and students of Native American and Hispanic ancestry were again overrepresented when compared to the WISC-III standardization sample.

Nondisabled sample. The WISC-III standardization sample included a representative sample of 2,200 nonexceptional children aged 6–0 to 16–11 years. See Wechsler (1991) and Table 1 for a complete description of the standardization sample.

Instrument

The WISC-III is an individually administered test of intellectual ability for children aged 6–0 to 16–11 years. It consists of 10 mandatory and three optional subtests ($M = 10$; $SD = 3$) that combine to yield Verbal (VIQ), Performance (PIQ), and Full Scale (FSIQ) IQs ($M = 100$; $SD = 15$). Full details of the instrument are available in Wechsler (1991).

Procedure

WISC-III subtest scores for each sample of participants were used to compute three indices of intersubtest variability: (a) number of subtests deviating by ± 3 points from the verbal subtest mean,

(b) number of subtests deviating by ±3 points from the performance subtest mean, and (c) number of subtests deviating by ±3 points from the full scale subtest mean. Verbal, performance, and full-scale means for the learning disabled sample and reading disabled subsample were based upon the 10 mandatory WISC-III subtests. Subtest means for the twelve-subtest subsample were based upon 12 WISC-III subtests with Digit Span being included in the Verbal scale and Symbol Search in the Performance scale.

Analysis

These indices of subtest variability were used to estimate diagnostic utility via Receiver Operating Curve (ROC) analysis (MacMillan & Creelman, 1991; Zweig & Campbell, 1993). ROC statistics are recommended as highly appropriate for assessing the diagnostic accuracy of psychological tests (McFall & Treat, 1999; Rey, Morris-Yates, & Stanislaw, 1992; Kraemer et al., 1999) because they do not depend on the prevalence of disabilities in the population and, consequently, provide a description of diagnostic accuracy that is independent of both base rate and decision threshold effects (Metz, 1978).

A ROC analysis graphically represents a test's diagnostic accuracy across its full range of scores. As illustrated by the dashed diagonal line in Figure 1, the ROC curve of a test with zero discriminating power is called the "random ROC." The more accurately a test is able to discriminate between individuals with and without the target disorder, the more its ROC curve will deviate toward the upper left corner of the graph. The area under the ROC curve (AUC) quantifies its diagnostic utility (Hanley & McNeil, 1982). A perfect test would produce an AUC of 1.0 whereas the random ROC always accounts for .50 of the area under the curve.

RESULTS

The average number of subtests deviating from each examinee's mean by ±3 points for each sample of participants are presented in Table 1. Inspection of the ROC curve of Figure 1, which is

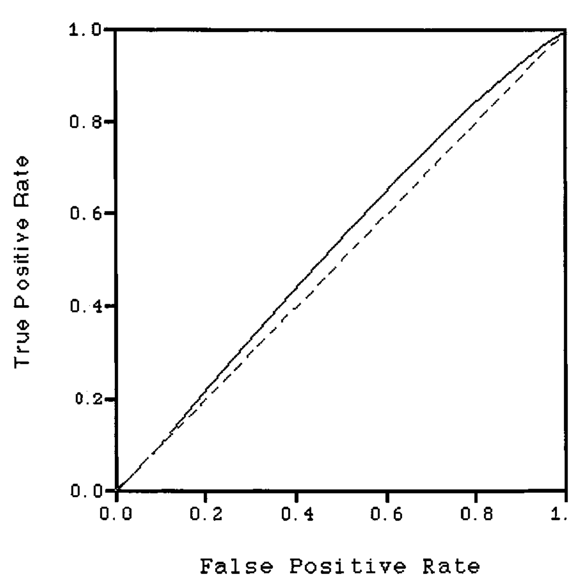

FIGURE 1. ROC analysis of total number of deviating verbal subtests for students with reading disabilities and children from the WISC-III standardization sample.

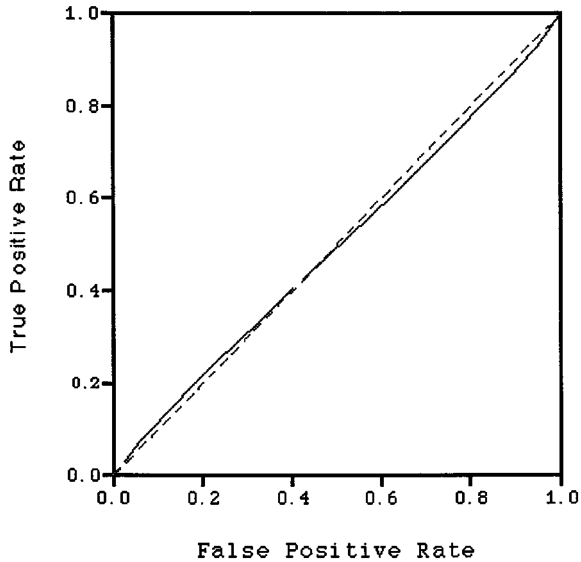

FIGURE 2. ROC analysis of total number of deviating performance subtests for students with reading disabilities and children from the WISC-III standardization sample.

based on the five verbal WISC-III subtests from the reading disabled sample compared to the standardization sample, reveals that it does not diverge from the random ROC. Likewise, Figure 2 indicates that the ROC curve calculated from the five WISC-III performance subtests of these participants is congruous with the random ROC. ROC curves for the other subtest variability comparisons were almost identical to those pictured in Figures 1 and 2.

Based upon nonparametric formulae presented by Hsiao, Bartko, and Potter (1989) and algorithms provided by Metz (1998), the AUC of Figure 1 was .52 and the AUC of Figure 2 was .50. Areas under the curve for each group of children with disabilities are presented in Table 3. AUCs of the magnitude found in Table 3 represent "low" diagnostic accuracy (Swets, 1988). That is, the probability that a randomly chosen child with learning disabilities will have greater subtest variability than a randomly selected child from the WISC-III standardization sample is only .50 to .53. This represents discrimination at a chance level.[1]

DISCUSSION

WISC-III subtest variability as quantified by the number of subtests deviating from examinees' mean subtest scores exhibited no diagnostic utility in distinguishing between children with learning disabilities and children from the WISC-III standardization sample. These negative results are consistent with previous research on Wechsler range and variance scatter indices (Anderson, Kaufman, & Kaufman, 1976; Daley & Nagle, 1996; Gutkin, 1979; Kline et al., 1993; Mayes et al., 1998). When considered within the broader, and generally negative, context of subtest profile research (Kavale & Forness, 1984; Kramer, Henning-Stout, Ullman, & Schnellenberg, 1987; McDremott, Fantuzzo, & Glutting, 1990; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; Mueller, Dennis, & Short, 1986; Watkins & Kush, 1994), subtest variability is not supported as a tool to gen-

[1]ROC analyses were also conducted across White, Hispanic, Native American, and Black subsamples with similar levels of discrimination (i.e., AUC range of .50 to .58).

Table 3
Areas Under the ROC Curve (AUC) for Number of Subtests Deviating From Mean Scores for Learning Disabled and Standardization Samples

	LD (10 subtests)[a]	LD (12 subtests)[b]	RD (10 subtests)[c]
Mean Number Subtests Deviating			
Verbal	.51	.50	.52
Performance	.50	.50	.50
Total	.51	.53	.51

[a]$N = 684$. [b]$n = 131$. [c]$n = 332$.

erate hypotheses regarding learning disabilities. Within the interpretative framework presented by Kamphaus (1998), using subtest variability as an indicator of learning disabilities would constitute a case of acting in opposition to scientific evidence.

References

Anderson, M., Kaufman, A.S., & Kaufman, N.L. (1976). Use of the WISC-R with a learning disabled population: Some diagnostic implications. Psychology in the Schools, 13, 381–386.

Arizona Department of Education. (1990). Working paper: Arizona guidelines for the identification of students with specific learning disabilities. Phoenix, AZ: Author.

Arizona Department of Education. (1992). Standard reporting format for psychoeducational evaluation. Phoenix, AZ: Author.

Banas, N. (1993). WISC-III prescriptions: How to work creatively with individual learning styles. Novato, CA: Academic Therapy Publications.

Cooper, S. (1995). The clinical use and interpretation of the Wechsler Intelligence Scale for Children®—Third Edition. Springfield, IL: Charles C. Thomas.

Daley, C.E., & Nagle, R.J. (1996). Relevance of WISC-III indicators for assessment of learning disabilities. Journal of Psychoeducational Assessment, 14, 320–333.

Gutkin, T.B. (1979). WISC-R scatter indices: Useful information for differential diagnosis? Journal of School Psychology, 17, 368–371.

Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristics (ROC) curve. Radiology, 143, 29–36.

Hsiao, J.K., & Bartko, J.J., & Potter, W.Z. (1989). Diagnosing diagnoses. Archives of General Psychiatry, 46, 664–667.

Kamphaus, R.W. (1998). Intelligence test interpretation: Acting in the absence of evidence. In A. Prifitera & D.H. Saklofske (Eds.), WISC-II clinical use and interpretation (pp. 39–57). San Diego, CA: Academic Press.

Kavale, K.A., & Forness, S.R. (1984). A meta-analysis of the validity of Wechsler scale profiles and recategorizations: Patterns or parodies? Learning Disabilities Quarterly, 7, 136–156.

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

Kellerman, H., & Burry, A. (1997). Handbook of psychodiagnostic testing: Analysis of personality in the psychological report. Boston: Allyn & Bacon.

Kline, R.B., Snyder, J., Guilmette, S., & Castellanos, M. (1993). External validity of the profile variability index for the K-ABC, Stanford-Binet, and the WISC-R: Another cul-de-sac. Journal of Learning Disabilities, 26, 557–567.

Kraemer, H.C., Kazdin, A.E., Offord, D.R., Kessler, R.C., Jensen, P.S., & Kupfer, D.J. (1999). Measuring the potency of risk factors for clinical or policy significance. Psychological Methods, 4, 257–271.

Kramer, J.J., Jenning-Stout, M., Ullman, D.P., & Schnellenberg, R.P. (1987). The viability of scatter analysis on the WISC-R and the SBIS: Examining a vestige. Journal of Psychoeducational Assessment, 5, 37–47.

MacMillan, N.A., & Creelman, C.D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.

Mayes, S.D., Calhoun, S.L., & Crowell, E.W. (1998). WISC-III profiles for children with and without learning disabilities. Psychology in the Schools, 35, 309–316.

McDermott, P.A., Fantuzzo, J.W., & Glutting, J.J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. Journal of Psychoeducational Assessment, 8, 290–302.

McDermott, P.A., Fantuzzo, J.W., & Glutting, J.J., Watkins, M.W., & Baggaley, A.R. (1992). Illusions of meaning in the ipsative assessment of children's ability. Journal of Special Education, 25, 504–526.

McFall, R.M., & Treat, T.A. (1999). Quantifying the information value of clinical assessments with signal detection theory. Annual Review of Psychology, 50, 214–241.

Metz, C.E. (1978). Basic principles of ROC analysis. Seminars in Nuclear Medicine, 8, 283–298.

Metz, C.E. (1998). Rockit [Computer software]. Chicago: Department of Radiology, University of Chicago.

Mueller, H.H., Dennis, S.S., & Short, R.H. (1986). A meta-exploration of WISC-R factor score profiles as a function of diagnosis and intellectual level. Canadian Journal of School Psychology, 2, 21–43.

Rey, J.M., Morris-Yates, A., & Stanislaw, H. (1992). Measuring the accuracy of diagnostic tests using receiver operating characteristics (ROC) analysis. International Journal of Methods in Psychiatric Research, 2, 39–50.

Sattler, J.M. (1992). Assessment of children: Revised and updated third edition. San Diego: Author.

Schinka, J.A., Vanderploeg, R.D., & Curtiss, G. (1997). WISC-III subtest scatter as a function of highest subtest scaled score. Psychological Assessment, 9, 83–88.

Stanovich, K.E. (1999). The sociopsychometrics of learning disabilities. Journal of Learning Disabilities, 32, 350–361.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. Science, 240, 1285–1293.

Vellutino, F.R., Scanlon, D.M., & Tanzman, M.S. (1991). Bridging the gap between cognitive and neuropsychological conceptualizations of reading disability. Learning and Individual Differences, 3, 181–203.

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

Wechsler, D. (1991). Manual for the Wechsler Intelligence Scale for Children—Third Edition. San Antonio, TX: Psychological Corporation.

Woodcock, R.W., & Mather, N. (1989). WJ-R tests of achievement: Examiner's manual. In R.W. Woodcock & M.B. Johnson, Woodcock–Johnson Psycho-Educational Battery-Revised. Allen, TX: DLM.

Zweig, M.H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. Clinical Chemistry, 39, 561–577.