# Diagnostic Utility of the WISC-III Developmental Index as a Predictor of Learning Disabilities

Marley W. Watkins


## Abstract

Wechsler's Deterioration Index (WDI) was developed as an indicator of cognitive impairment in adults but has been applied to children, because neuropsychological deficits have often been hypothesized to account for learning difficulties during the development period. Renamed the Wechsler *Developmental* Index, this measure has been used to discriminate among groups of children with and without learning disabilities. The present study replicated those findings with the Wechsler Intelligence Scale for Children–Third Edition, but also applied more appropriate diagnostic efficiency statistics to analyze the actual diagnostic utility of the WDI. These analyses revealed that the WDI performed at chance levels when distinguishing 611 students diagnosed with learning disabilities from those diagnosed with emotional disabled ($n$ = 80) or mental retardation ($n$ = 33), as well as from 2,200 simulated random nondisabled cases. It was concluded that mean group differences were not adequate and that ipsative indicators must be definitively validated in experimental environments before they can be applied in practice.


Wechsler's (1958) Deterioration Index (WDI) was originally developed as an indicator of cognitive impairment that was hypothesized to be sensitive to brain injury in adults. Conceptually, the WDI was composed of two groups of Wechsler subtest scores: *hold* subtests, which were considered to be insensitive to deterioration in brain injury (Vocabulary, Information, Object Assembly, and Picture Completion), and *don't hold* subtests, which were judged vulnerable to intellectual decline (Digit Span, Similarities, Coding, and Block Design).

Use of the WDI with adults has received considerable theoretical attention (Lezak, 1983; Livesay, 1986), but its diagnostic utility with that population should be cautiously interpreted within the broader context of neuropsychological assessment (Faust, 1991; Prigatano & Redner, 1993; Wedding & Faust, 1989). Application of the WDI with children was recently suggested by Bowers et al. (1992), given that neuropsychological deficits have often been hypothesized to account for learning and attentional difficulties in children (Accardo & Whitman, 1991; Goodyear & Hynd, 1992). For example, contemporary definitions of learning disabilities employ such neurology-based terms as "brain injury," "presumed to be due to central nervous system dysfunction," and "of presumed neurological origin" (Hammill, 1990, p. 82) to conjecture an endogenous etiology.

The WDI is conceptually consistent with a theoretical explanation of learning problems that attributes difficulty to a general neurocognitive deficit (Reschly & Gresham, 1989). It is inconsistent with competing theories that discount global conceptions of neurological dysfunction in favor of localized cognitive deficits (Mather & Roberts, 1994) and with those that focus on specific subtypes of learning disabilities based upon discrete patterns of neuropsychological assets and deficits (Rourke, 1994). It is also inconsistent with theories that reject neuropsychological constructs entirely (Coles, 1987; Klatt, 1991). Thus, the WDI as a reflection of a homogeneous pattern of Wechsler performance has most relevance to the theoretical position that global neuropsychological impairments account for learning disabilities.

The general-deficit concept was investigated by Bowers, Washburn, and Livesay (1986), who found a low, but significant, correlation between the WDI and the Halstead Impairment Index in children. In a follow-up study, Bowers et al. (1992) analyzed the Wechsler Intelligence Scale for Children–Revised (WISC-R; Wechsler, 1974) performance of two samples of children with learning disabilities and two samples with attention-deficit/hyperactivity disorder (ADHD). The WDI did not predict severity of academic delay for the two groups of children diagnosed with learning disabilities ($n$ = 26 and $n$ = 35). However, the groups of children with ADHD ($n$ = 10 and $n$ = 17) displayed significantly higher WDI scores than the nondisabled children. Bowers et al.

JOURNAL OF LEARNING DISABILITIES
VOLUME 29, NUMBER 3, MAY 1996
PAGES 305–312

recommended that the WDI be renamed the Wechsler *Developmental* Index, as children's cognitive skills are not deteriorating but, rather, assumed to be developing unevenly. Methodological problems led Bowers et al. to suggest that their research be replicated to provide greater assurance of generalizability.

A replication was subsequently conducted by Klein and Fisher (1994). In addition to a larger number of children with learning disabilities ($n$ = 104), that study included 72 nondisabled children and analyzed scores from the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991). Results indicated that children in learning disability programs scored significantly higher on the WDI than did children not placed in special education programs. Based on statistically significant group differences, Klein and Fisher concluded that the WDI was useful in predicting which students would be found eligible for learning disability services. Thus, their detection of WDI group differences for children with learning disabilities corresponded to the findings of Bowers et al. (1992) for children with ADHD. The methodology used in the replication does not, however, rule out alternative hypotheses that might account for the results.

The current emphasis on correlational methods and statistical significance testing of the null hypothesis in educational and psychological research has been widely criticized as an overused and misunderstood practice that readily lends itself to misinterpretation and often detracts from more important considerations (Cohen, 1990; Meehl, 1978; Pollard, 1993; Rosnow & Rosenthal, 1989; Shaver, 1993; Thompson, 1989, 1994a, 1994b).

When these methodological standards are applied to the Bowers et al. (1992) experiment, it becomes apparent that a group of 17 children diagnosed as ADHD is insufficient for ensuring that a representative sample of the population was selected. Statistical conclusions cannot be drawn from those data, because the power to detect a valid research hypothesis was inadequate, and alternative hypotheses, such as chance, are equally plausible (Tversky & Kahneman, 1993; Wedding & Faust, 1989).

More than 40 years ago, Meehl and Rosen (1955) demonstrated that efficient diagnosis depends on the psychometric instruments employed *and* on a consideration of base rates, or prevalence, of the criterion condition in both the nondisabled and clinical populations. That insight was emphatically rearticulated by Elwood (1993), who asserted that "significance alone does *not* reflect the size of the group differences nor does it imply the test can discriminate subjects with sufficient accuracy for clinical use" (p. 409).

Viable solutions to the daunting methodological problems encountered in diagnostic decision making have been developed and employed in such diverse fields as medicine, materials testing, and weather forecasting (Colliver, Vu, & Barrows, 1992; Swets, 1988; Swets et al., 1979; Wedding & Faust, 1989). Early attempts to introduce them into the study of decision making in special education foundered (Harber, 1981), but they are now applied with regularity in research on the diagnosis of psychopathology (Fagot, 1992; McDermott et al., 1995).

Kessel and Zimmerman (1993) recently contributed a succinct, comprehensive presentation of these diagnostic-efficiency statistics. As outlined in Table 1, their list of eight common statistical indices allows analysis of a test's accuracy in relation to two pervasive alternative interpretations: base rate and chance (Cohen, 1990; Rosnow & Rosenthal, 1989).

Although positive WDI results were reported by both Bowers et al. (1992) and Klein and Fisher (1994), some questions remain unanswered regarding the diagnostic utility of the WDI. Two are the most pressing. First, are these results replicable across diverse samples of children with and without learning problems? Thompson (1994a) presented a detailed discussion regarding the pivotal role that replication plays in validating educational and psychological research. Second, did the methodology used in previous studies accurately differentiate between groups for diagnostic purposes, or are other hypotheses more credible? Werthamer-Larsson (1994) identified critical issues surrounding the diagnosis of childhood disorders and stressed the necessity of applying rigorous methodological designs so as to exclude alternative plausible interpretations. The present research was designed to address these two basic questions with the WDI.

## Method

### *Participants*

Students enrolled in three southwestern suburban school district special education programs who received comprehensive psychological evaluations during a 1-year period served as participants. Students were selected from special education records based upon two criteria: (a) that their cognitive assessment included the WISC-III, and (b) a diagnosis of learning disability (LD), emotional disability (ED), or mental retardation (MR).

The state's special education rules and regulations, which governed diagnostic placements, were similar to IDEA standards. That is, *learning disability* was defined as a significant ability–achievement discrepancy (using a regression approach with 1.5 standard errors of estimate suggested as a severe discrepancy); *emotional disability* entailed one of five emotional characteristics adversely affecting educational progress; and the *mental retardation* label required deficits in both intellectual functioning and adaptive behavior. Learning disability diagnoses were further delineated by local multidisciplinary teams into eligibility areas: reading, math, and/or written expression.

These selection criteria identified 724 students. Of this number, 91% were

**TABLE 1**
Nomenclature and Description of Recommended Diagnostic Efficiency Statistics

| Nomenclature | Description |
|---|---|
| Sensitivity | True positive rate. Proportion of participants with a target disorder who are identified by a positive test finding. |
| Specificity | True negative rate. Proportion of participants free of the disorder who are correctly identified by a negative test result. |
| Positive predictive power | Proportion of participants identified by a positive test score who truly have the target disorder. |
| Negative predictive power | Proportion of participants identified by a negative test score who truly do *not* have the target disorder. |
| False positive rate | Proportion of participants identified by a positive test score who truly do *not* have the target disorder. |
| False negative rate | Proportion of participants identified by a negative test score who truly have the target disorder. |
| Overall correct classification | Hit rate. Proportion of participants with *and* without the target disorder who were correctly classified by the test. |
| Kappa | Proportion of agreement between the test and actual condition of the participants (disordered vs. nondisordered) beyond that accounted for by chance alone. |

placed in Grades 1 through 8. Special education enrollment was 84% in LD, 11% in ED, and 5% in MR programs. Gender distribution was 73% male and 27% female. Acknowledged ethnic membership was 63% White, 7% Black, 13% Hispanic, and 17% Native American.

Total combined enrollment in these three participating districts summed to approximately 45,000 students across Grades K through 12. Student socioeconomic levels were estimated from the zip codes of each home address. Students lived in 18 separate zip code zones; poverty rates ranged from 2.0% to 28.6% and per capita income ranged from $9,735 to $37,288 in 1989 dollars (*Sourcebook*, 1992).

Academic achievement levels in reading, math, and written expression were primarily measured with the Woodcock-Johnson Tests of Achievement–Revised (Woodcock & Mather, 1989). Eight other achievement measures accounted for only 7% of the students' academic achievement scores. Table 2 presents intellectual and academic achievement scores for participating students by special education classification and gender. Results are congruent with previous compilations of data from children with learning disabilities (Kavale & Nye, 1985).

As Keith and Reynolds (1990) noted, a hypothesis-testing approach is required if assessment research is to lead to empirically validated test applications. Given the large number of children with disabilities included in this sample (724), a fundamental research issue was the relative ability of the WDI to distinguish between disabled and nondisabled children. The WISC-III normative sample was deemed an ideal comparison group for answering this question. However, the Psychological Corporation, publisher of the WISC-III, refused access to the WISC-III normative sample (C. Doebbler, personal communication, March 3, 1994; A. Prifitera, personal communication, May 16, 1994). Consequently, a 2,200-case multivariate nondisabled comparison sample was randomly generated (Bentler, 1993; Norusis, 1990). Results closely approximated the published WISC-III norm group subtest distributions and intercorrelations. Random subtest means ranged from 9.9 to 10.0, and standard deviations, from 3.0 to 3.4; and the maximum subtest intercorrelation deviation from those reported in Table 3.12 (p. 281) of the WISC-III manual was .03. Although less satisfactory than the 2,200 normative cases, this sample permits a test of the WDI's ability to distinguish between children with various disabilities and multivariate random cases. Thus, the experimental sample contains four discrete groups: (a) LD—students with learning disabilities, (b) ED—students with emotional disabilities, (c) MR—students diagnosed with mental retardation, and (d) RN—a group of random normal scores generated by computer.

### *Procedure*

WDI scores were calculated for all subjects with complete WISC-III subtest data by applying the usual formula: WDI = (*hold* – *don't hold*) ÷ *hold*. As standardized in previous research (Bowers et al., 1992; Klein & Fisher, 1994), subjects were categorized as impaired if their WDI exceeded .20 and nonimpaired if their WDI was .20 or less. WDI scores for subjects by group and classification are provided in Table 3.

## Results

A one-way ANOVA indicated that mean WDI scores of the LD, ED, MR, and NR groups were different at a statistically significant level, $F(3, 2920) = 19.69$, $p < .001$. A Tukey HSD post hoc analysis revealed that the average WDI scores of the LD and ED groups were significantly higher than the MR and NR group means ($p < .05$).

Achievement lags were calculated by subtracting reading, math, and written expression scores from WISC-III Full Scale IQ (FSIQ) scores to create three separate areas of achievement delay for each student. The relationship between WDI and achievement delay was estimated via the correlation of WDI scores with those three simple ability–achievement discrepancy measures. Results were weak to nonexistent (reading $r = .03$, math $r = .03$, and written expression $r = .10$).

As expected, the impaired group demonstrated a statistically significant higher mean WDI score than the nonimpaired group, t (2,922) = 29.53, $p < .0001$. Intellectual and academic skills were similar for the impaired and nonimpaired groups, with the exception of reading, $t(715) = 2.62$, $p < .01$, where the nonimpaired group exceeded the impaired group (85.2 vs. 81.3, respectively).

Diagnostic-efficiency statistics for the WDI are applied to the current groups of students with and without learning disabilities in Table 4. Although the results displayed therein demonstrate that the WDI is unable to efficiently diagnose learning disabilities, it is possible that the WDI cutting score (.20 by convention) was not properly calibrated, and the index may be effective at some other cutting point. This notion was tested with an extension of the diagnostic-efficiency statistics previously discussed. Originally developed in engineering as a way to tell how well an electronics receiver was able to distinguish signal from noise, this procedure has been adapted and reformulated for biostatistic applications (Hanley & McNeil, 1982, 1983; Kraemer, 1988; Murphy et al., 1987; Swets, 1988).

Designated the receiver operating characteristic (ROC), this procedure entails plotting the balance between the sensitivity and specificity of a diagnostic test while systematically moving the cutting score across its full range of values. As illustrated in Figure 1, the ROC curve of a test with zero discriminating power is a diagonal line dubbed the "line of no information" or the "random ROC." The more clearly a test is able to discriminate between individuals with and without the target disorder, the farther its ROC curve will deviate toward the upper left corner of the graph. Visual inspection of the ROC curve of Figure 1, which is based on the current WDI data, reveals that it does not substantially diverge from the random ROC. This illustration can be quantified by calculating the area under the ROC curve (AUC; Colliver et al., 1992; Hanley & McNeil, 1982) to produce an overall index of the accuracy of the discrimination provided by the WDI. A perfect test would produce an AUC of 1.0, whereas the random ROC always accounts for .50 of the area under the curve. Based upon the for-

**TABLE 2**
Intellectual and Achievement Standard Scores by Gender of Students Classified as Learning Disabled (LD), Emotionally Disabled (ED), and Mentally Retarded (MR)

| | LD | | ED | | MR | |
|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female |
| WISC-III FSIQ | | | | | | |
| *M* | 92.5 | 88.5 | 93.1 | 86.0 | 56.8 | 60.3 |
| *SD* | 11.7 | 12.4 | 12.5 | 14.4 | 8.0 | 8.1 |
| WISC-III VIQ | | | | | | |
| *M* | 91.5 | 88.2 | 92.8 | 86.0 | 60.6 | 62.8 |
| *SD* | 11.7 | 11.5 | 11.6 | 13.1 | 7.5 | 7.3 |
| WISC-III PIQ | | | | | | |
| *M* | 95.2 | 90.9 | 95.0 | 88.4 | 59.5 | 63.5 |
| *SD* | 13.2 | 14.6 | 14.2 | 15.0 | 10.1 | 10.0 |
| Reading | | | | | | |
| *M* | 81.5 | 81.4 | 91.9 | 95.0 | 65.3 | 68.4 |
| *SD* | 14.1 | 12.6 | 19.3 | 9.6 | 16.4 | 11.8 |
| Math | | | | | | |
| *M* | 85.5 | 82.7 | 92.5 | 91.1 | 58.7 | 60.2 |
| *SD* | 13.6 | 15.5 | 14.3 | 12.6 | 14.6 | 11.4 |
| Written Expression | | | | | | |
| *M* | 75.9 | 78.0 | 82.7 | 88.3 | 58.0 | 66.9 |
| *SD* | 11.2 | 11.6 | 15.3 | 9.2 | 13.6 | 9.0 |

*Note.* WISC-III = Wechsler Intelligence Scale for Children–Third Edition; FSIQ = Full-Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ.

mulas provided by Hsiao, Bartko, and Potter (1989), the AUC of Figure 1 summed to .57. As explained by Hanley and McNeil, "the area under the curve can be thought of simply as measuring the probability of a correct ranking of a (normal, abnormal) pair" (p. 31). Conceptually, the probability is .57 that one of the participants with a learning disability in this study would obtain a higher WDI score than a randomly selected participant who did not have a learning disability. In contrast, tossing a coin would result in a correct classification rate of 50%. Thus, there is no WDI cutting score more efficacious than the original .20 standard.

**TABLE 3**
WDI Scores for Subjects by Experimental Group and Impaired Versus Nonimpaired Classification

| | WDI | | |
|---|---|---|---|
| Group | *n* | *M* | *SD* |
| Learning disabilities | 611 | .03 | .19 |
| Emotional disabilities | 80 | .06 | .16 |
| Mental retardation | 33 | −.10 | .47 |
| Random nondisabled | 2200 | −.03 | .21 |
| Impaired | 306 | .28 | .07 |
| Nonimpaired | 2618 | −.05 | .19 |

*Note.* WDI = Wechsler Developmental Index.

**TABLE 4**
Diagnostic Efficiency of the WDI When Used to Predict Membership in Groups With and Without Disabilities in Three Studies

| | Study | | |
|---|---|---|---|
| Diagnostic-efficiency statistic | 1 | 2 | 3 |
| Sensitivity | .15 | .21 | .15 |
| Specificity | .91 | .89 | .80 |
| Postive predictive power | .30 | .73 | .80 |
| Negative predictive power | .80 | .44 | .15 |
| False positive rate | .09 | .11 | .20 |
| False negative rate | .85 | .79 | .85 |
| Overall correct classification | .75 | .49 | .25 |
| Kappa | .07 | .09 | −.02 |

*Note.* Study 1 identifies the current investigation with 2,924 subjects (LD compared to ED, MR, and NR cases). Study 2 refers to 176 subjects classified by Klein and Fisher (1994). Study 3 refers to the current study with 611 subjects in the LD group and 113 subjects in the nonLD group (80 ED and 33 MR cases). WDI = Wechsler Developmental Index; LD = children with learning disabilities; ED = children with emotional disabilities; MR = children with mental retardation; NR = simulated random nondisabled cases.

## Discussion

The ability of the WDI to serve as a distinctive measure of neurocognitive impairment in children with learning disabilities was investigated in this study. It has been hypothesized that the WDI is an index of children's intellectual functioning that reflects developmental lags in important intellectual skills (Bowers et al., 1992). The present study found that a large group of children with learning disabilities exhibited average WDI scores that were significantly higher than those of children with diagnoses of mental retardation and than those calculated from cases constructed from a random multivariate normal distribution of scores.

The relationship of WDI scores to ability–achievement discrepancies was weak, and overall intellectual skills were similar across all three groups of students with disabilities. These results are consistent with those reported by Klein and Fisher (1994) for children with learning disabilities, and with those presented by Bowers et al. (1992) for children with ADHD. The present group of students with elevated WDI scores did, however, exhibit significantly higher reading scores than did students with nonelevated WDI scores. Thus, mean WDI score differences between groups of students with and without specific disabilities were replicated in a diverse group of students.

Clearly, the WDI classification of these students, when compared to their true condition, agreed at near chance levels (kappa = .07). Only 36% of the students identified by an elevated WDI were actually students previously diagnosed with learning disabilities. Similarly, a WDI of less than .20 (nonimpaired) frequently identified as nondisabled students who were actually enrolled in learning disability programs. Inaccurate classifications were replicated across all WDI values, as illustrated in the ROC curve of Figure 1, so these imprecise results were not caused by an improperly calibrated WDI cutting score. These statistics indicate that the WDI is incapable of assisting in the diagnostic decision-making process when students with learning disabilities are to be distinguished from students with other disabilities or from students without disabilities. The unavoidable conclusion is that the WDI is ineffectual in accurately discriminating among students and is, therefore, an invalid addition to the psychoeducational diagnostic process. Elwood's (1993) assertion that statistically significant group differences do not allow discrimination of subjects

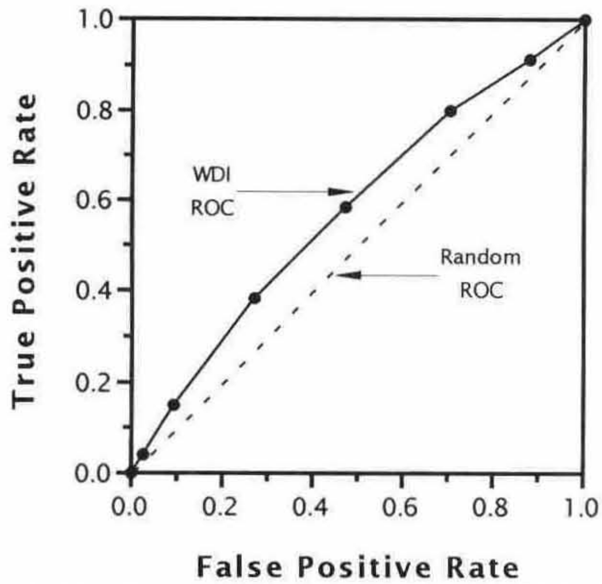


**FIGURE 1.** Receiver operating characteristic (ROC) of Wechsler Developmental Index used to distinguish between subjects with and without learning disabilities.

with sufficient accuracy for clinical use was substantiated.

Table 4 also clearly displays comparable results when the Klein and Fisher (1994) WDI classifications were reanalyzed with diagnostic-efficiency statistics. That is, near chance levels of agreement between actual diagnosis and WDI level were observed (kappa = .09). Klein and Fisher acknowledged that the WDI *alone* might not be adequate but hypothesized that when it was used in conjunction with other techniques, it would improve diagnostic accuracy. Ironically, Meehl and Rosen (1955) identified just this situation as one that could result in an "increase of erroneous clinical decisions" (p. 215). That is, adding a random component to the decision-making process cannot increase, and may decrease, diagnostic accuracy when trying to detect a low-incidence condition.

Although the current results were negative, they were calculated from a large sample of students with disabilities; students without disabilities were only modeled by random scores. The lack of nondisabled comparison group is a pervasive flaw in educational and psychological research (McDermott, Fantuzzo, & Glutting, 1990) and always constitutes a serious threat to the external validity of the results. There is no way to adjust for or accommodate the lack of a nondisabled comparison group, but it is instructive to recalculate all diagnostic utility statistics from this study after excluding the random cases. The resulting figures (see Table 4) are not supportive of the WDI hypothesis, even though the base rate is heavily skewed in favor of a diagnosis of learning disabilities. This suggests that the failure of the WDI as a diagnostic indicator is a robust outcome that must be accepted as more plausible than results based on mean differences between groups.

These negative results may be so uniform because they reflect underlying measurement limitations inherent in any ipsative measure (Dunlap & Cornwell, 1994; Macmann & Barnett, 1994; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992). Thus, any diagnosis of childhood disorders achieved through recategorization or comparison of Wechsler subtest scores must be considered unproven until definitively validated through appropriate diagnostic-efficiency statistics.

### ABOUT THE AUTHOR

*Marley W. Watkins, PhD, received his doctorate in school psychology from the University of Nebraska–Lincoln. He is a diplomate in school psychology of the American Board of Professional Psychology and an associate professor of education (school psychology) at Pennsylvania State University. Address: Marley W. Watkins, Department of Educational and School Psychology and Special Education, Pennsylvania State University, 227 CEDAR Building, University Park, PA 16802.*

### AUTHOR'S NOTE

*Preliminary versions of this paper were presented at annual meetings of the Arizona Association of School Psychologists and the National Association of School Psychologists.*

### REFERENCES

Accardo, P. J., & Whitman, B. Y. (1991). The misdiagnosis of the hyperactive child. In P. J. Accardo, T. A. Blondis, & B. Y. Whitman (Eds.), *Attention deficit disorders and hyperactivity in children* (pp. 1–21). New York: Marcel Dekker.

Bentler, P. (1993). *EQS structural equations program manual.* Los Angeles: BMDP Statistical Software.

Bowers, T. G., Risser, M. G., Suchanec, J. F., Tinker, D. E., Ramer, J. C., & Domoto, M. (1992). A developmental index using the Wechsler Intelligence Scale for Children: Implications for the diagnosis and nature of ADHD. *Journal of Learning Disabilities, 25,* 179–185.

Bowers, T., Washburn, S., & Livesay, J. (1986). Predicting neuropsychological impairment by screening instruments and intellectual evaluation indices: Implications for the measuring of Kaufman's Factor III. *Psychological Reports, 59,* 487–493.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304–1312.

Coles, G. (1987). *The learning mystique: A critical look at "learning disabilities."* New York: Pantheon Books.

Colliver, J. A., Vu, N. V., & Barrows, H. S. (1992). Screening test length for sequential testing with a standardized-patient examination: A receiver operating characteristic (ROC) analysis. *Academic Medicine, 67,* 592–595.

Dunlap, W. P., & Cornwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research, 29,* 115–126.

Elwood, R. W. (1993). Psychological tests and clinical discriminations: Beginning to address the base rate problem. *Clinical Psychology Review, 13,* 409–419.

Fagot, B. I. (1992). Assessment of coercive parent discipline. *Behavioral Assessment, 14,* 387–406.

Faust, D. (1991). Forensic neuropsychology: The art of practicing a science that does not yet exist. *Neuropsychology Review, 2,* 205–231.

Goodyear, P., & Hynd, G. W. (1992). Attention-deficit disorder with (ADD/H) and without (ADD/WO) hyperactivity: Behavioral and neuropsychological differentiation. *Journal of Clinical Child Psychology, 21,* 273–305.

Hammill, D. D. (1990). On defining learning disabilities: An emerging consensus. *Journal of Learning Disabilities, 23,* 74–84.

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

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148,* 839–843.

Harber, J. R. (1981). Assessing the quality of decision making in special education. *The Journal of Special Education, 15,* 77–90.

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

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

Keith, T. Z., & Reynolds, C. R. (1990). Measurement and design issues in child assessment research. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 29–61). New York: Guilford.

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment, 5,* 395–399.

Klatt, H. (1991). Learning disabilities: A questionable construct. *Educational Theory, 41,* 47–60.

Klein, E. S., & Fisher, G. S. (1994, March). *The usefulness of the Wechsler Deterioration Index as a predictor of learning disabilities in children.* Paper presented at the annual meeting of the National Association of School Psychologists, Seattle, WA.

Kraemer, H. C. (1988). Assessment of 2 × 2 associations: Generalization of signal-detection methodology. *American Statistician, 42,* 37–49.

Lezak, M. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Livesay, J. (1986). Clinical utility of Wechsler's Deterioration Index in screening for behavioral impairment. *Perceptual and Motor Skills, 25,* 1191–1194.

Macmann, G. M., & Barnett, D. W. (1994). Structural analysis of correlated factors: Lessons from the verbal–performance dichotomy of the Wechsler scales. *School Psychology Quarterly, 9,* 161–197.

Mather, N., & Roberts, R. (1994). Learning disabilities: A field in danger of extinction? *Learning Disabilities Research & Practice, 9,* 49–58.

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. *The Journal of Special Education, 25,* 504–526.

McDermott, P. A., Watkins, M. W., Sichel, A. F., Weber, E. M., Keenan, J. T., Holland, A. M., & Leigh, N. M. (1995). The accuracy of new national scales for detecting emotional disturbance of children and adolescents. *The Journal of Special Education, 29,* 337–354.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806–834.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194–216.

Murphy, J. M., Berwick, D. M., Weinstein, M. C., Borus, J. F., Budman, S. H., & Klerman, G. L. (1987). Performance on screening and diagnostic tests. *Archives of General Psychiatry, 44,* 550–555.

Norusis, M. J. (1990). *SPSS base system user's guide.* Chicago: SPSS, Inc.

Pollard, P. (1993). How significant is "significance?" In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 449–460). Hillsdale, NJ: Erlbaum.

Prigatano, G. P., & Redner, J. E. (1993). Uses and abuses of neuropsychological testing in behavioral neurology. *Behavioral Neurology, 11,* 219–231.

Reschly, D. J., & Gresham, F. M. (1989). Current neuropsychological diagnosis of learning problems: A leap of faith. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Handbook of clinical child neuropsychology* (pp. 503–519). New York: Plenum Press.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276–1284.

Rourke, B. P. (1994). Neuropsychological assessment of children with learning disabilities. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 475-514). Baltimore: Brookes.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education, 61,* 293–316.

*Sourcebook of ZIP code demographics.* (1992). Arlington, VA: CACI Marketing Systems.

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

Swets, J. A., Pickett, R. M., Whitehead, S. F., Getty, D. J., Schnur, J. A., Swets, J. B., & Freeman, B. A. (1979). Assessment of diagnostic technologies. *Science, 205,* 753–759.

Thompson, B. (1989). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. *Measurement and Evaluation in Counseling and Development, 22,* 2–6.

Thompson, B. (1994a). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality, 62,* 157–176.

Thompson, B. (1994b). *The concept of statistical significance testing* (Report No.

EDO-TM-94-1). Washington, DC: ERIC Clearinghouse on Assessment and Evaluation.

Tversky, A., & Kahneman, D. (1993). Belief in the law of small numbers. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 341–349). Hillsdale, NJ: Erlbaum.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised.* San Antonio, TX: Psychological Corp.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corp.

Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4,* 233–265.

Werthamer-Larsson, L. (1994). Methodological issues in school-based services research. *Journal of Clinical Child Psychology, 23,* 121–132.

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.