Diagnostic Utility of the Learning Disability Index

Marley W. Watkins, Joseph C. Kush, and Barbara A. Schaefer

Abstract

The Learning Disability Index (LDI) is one of many diagnostic indicators proposed for the identification of students with learning disabilities that relies on patterns of performance on cognitive tests. The LDI is hypothesized to relate to students' specific neuropsychological deficits. The present study investigated the diagnostic utility of the LDI with the third edition of the Wechsler Intelligence Scale for Children by comparing students previously diagnosed with learning disabilities (n = 2,053) to students without learning disabilities (n = 2,200). Subsamples of youth with specific reading (n = 445) and math (n = 168) disabilities permitted further assessment of the efficacy of the LDI. Receiver operating characteristic (ROC) curves revealed that the LDI resulted in a correct diagnostic decision only 55% to 64% of the time. These results demonstrate that the LDI is not a valid diagnostic indicator of learning disabilities.

The identification of students with learning disabilities (LD) in need of special education services is beset with complexity (Chalfant, 1989; Kavale & Forness, 2000). Although some experts have perceived an emerging consensus on diagnostic definitions and procedures (Hammill, 1990), others have continued to see a host of problems (Stanovich, 1999). Given that students with learning disabilities account for more than 50% of the 5 million students enrolled in special education programs in the United States (U. S. Department of Education, 1999), valid diagnostic criteria are crucial to ensure that students receive appropriate educational services (Reschly, 1997).

To this end, state departments of education have promulgated criteria for identifying students with learning disabilities. Although these criteria are not uniform in their requirements, an ability-achievement discrepancy standard is included by most states (Mercer, Jordan, Allsopp, & Mercer, 1996). However, some researchers have proposed that academic achievement alone be used to identify students with learning disabilities (Siegel, 1989), whereas other diagnosticians have suggested procedures that focus on students' performance patterns on cognitive tests (Bannatyne, 1974; Kaufman, 1994; Mayes, Calhoun, & Crowell, 1998; Strauss, Spreen, & Hunter, 2000).

Of the many cognitive test patterns that have been advanced as diagnostic of learning disabilities, the Learning Disability Index (LDI; Lawson & Inglis, 1984) is of particular interest because it has been hypothesized to relate to specific neuropsychological deficits of students with learning disabilities (Lawson & Inglis, 1985). Lawson and Inglis (1984, 1985) conjectured that Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) subtests are sensitive to the presence of learning disabilities in direct proportion to their verbal saturation, which is in turn related to left-hemisphere dysfunction. This theoretical link between test scores and brain functioning is important because contemporary definitions of learning disabilities specify an endogenous etiology related to "central nervous system dysfunction" (Hammill, 1990, p. 82).

Comparisons of groups of students with and without learning disabilities have found significantly higher mean LDI scores among students with learning disabilities than among students in general education (Bellemare, Inglis, & Lawson, 1986; Clampt & Silver, 1990; Lawson & Inglis, 1985; Tittemore, Lawson, & Inglis, 1985). Statistically significant LDI differences between groups have been subsequently interpreted as evidence that the LDI is diagnostically effective. For example, Kaufman (1990) concluded that the LDI taps a sequential–simultaneous processing dimension and is "quite valuable for distinguishing learning-disabled children from normal children" (p. 354).

However, Meehl and Rosen (1955) warned psychologists that they would be misled if they used "validity" or "discrimination" between groups to justify diagnostic decision making. More recently, Elwood (1993) cautioned 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). Thus, the accuracy of the LDI in diagnosing students with learning disabilities awaits determination through the application of appropriate diagnostic utility statistics (Kessel & Zimmerman, 1993; Zarin & Earls, 1993).
Contemporary use of the LDI is also constrained because it was developed with the WISC-R. That scale has been replaced by the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991). Like its predecessor, the WISC-III is the most popular intellectual measure used to determine eligibility for special education services (Wilson & Reschly, 1996). Although the WISC-III is a direct descendant of the WISC-R, only about 73% of the WISC-R items were retained in the WISC-III (Edwards & Edwards, 1993). Moreover, materials and administration procedures were revised for the WISC-III. These changes make it difficult to know whether the results of previous LDI research can be applied to the WISC-III (Strauss et al., 2000).

Given the profound influence that diagnostic decisions have in children’s lives (Dahlstrom, 1993), it is important to fully delineate the diagnostic utility of any indicator used to classify or program for children. Consequently, the present study investigated the diagnostic utility of the WISC-III LDI among a large group of children previously diagnosed with learning disabilities.

**Procedure**

Based on Department of Education records, all 212 special education directors of Arizona school districts were contacted and asked to provide anonymous WISC-III data on students currently enrolled in their special education programs. Personnel from 40 school districts responded with anonymous data on 2,979 students in special education with current psychoeducational evaluations on file (i.e., WISC-III administered within the past 3 years). Of this number, 2,274 students were categorized as having learning disabilities. All participants were diagnosed independently by school district multidisciplinary teams (MDT) based on federal and Arizona special education rules and regulations that required the demonstration of a significant ability-achievement discrepancy exclusive of sensory impairment, mental retardation, emotional disturbance, and environmental, cultural, or economic disadvantage.

**Participants**

Students with Learning Disabilities. Congruent with previous surveys of practitioners (Canivez & Watkins, 1998), optional WISC-III subtests were found to be infrequently administered. However, to maintain consistency with WISC-R LDI research, 10 mandatory subtests and 1 optional subtest (Digit Span) were necessary for the computation of LDI scores. Based on this requirement, 2,053 students with learning disabilities from 37 school districts participated in the current study.

Students were determined by local MDT to exhibit learning disabilities in reading alone (n = 160), math alone (n = 137), written expression alone (n = 412), reading and written expression (n = 580), reading and math (n = 63), math and written expression (n = 203), reading, math, and written expression (n = 493), and not specified (n = 5). Boys constituted 71.9% of the sample and girls 28.1%. Mean age was 10.7 years (SD = 2.6) and ranged from 6 to 16 years. Median grade placement was 5.0, with a range of kindergarten through 11. Ethnic background, as reported on school records, was 67.7% White, 17.3% Hispanic, 5.1% Black, 9.3% Native American, and 0.6% Asian/Pacific. Because data were anonymously retrieved from archival special education records, socioeconomic status could not be determined. However, the participants were distributed across rural, urban, and suburban school districts and were widely dispersed across the state.

Specific Reading Disability. A subsample of participants was identified to allow specialized analyses for students with specific reading disabilities. Selection criteria included:

1. identification of a learning disability in reading by a MDT;
2. discrepancy of 15 or more points between predicted (via regression on FSIQ) and actual reading achievement;
3. no identification as having a learning disability in math by a MDT; and
4. discrepancy of less than 15 points between predicted (via regression on FSIQ) and actual math achievement.

These criteria selected 445 students from the larger sample of children with learning disabilities. Whereas the general learning disabilities group was marked by average FSIQ—reading and FSIQ—math discrepancies of 9.4 and 5.6 points, respectively, the subsample with specific reading disabilities had average discrepancies in reading and math of 22.1 and 1.9 points, respectively. Their mean cognitive and achievement scores are provided in Table 1.

Specific Math Disability. A second subsample of participants was identified to allow specialized analyses for students with specific math disabilities. Selection criteria included,
Students Without Disabilities. The United States WISC-III standardization sample of 2,200 children ages 6 years 0 months through 16 years 11 months served as controls. See Wechsler (1991) for a complete description of the WISC-III standardization sample.

Analyses

LDI. Following the method provided by Lawson and Inglis (1984, 1985), the average intercorrelation matrix from the WISC-III standardization sample (Wechsler, 1991) was subjected to an unrotated principal components analysis. Table 2 provides the results of the two-factor solution in terms of factor loadings and their associated factor score coefficients. As with the WISC-R, the first component reflects a general factor, whereas the second component reveals a verbal–nonverbal dimension. LDI scores were calculated according to the following formula:

\[ \text{LDI} = \sum [W_i(X_i - \bar{M})] \]

where \( W_i \) is the Factor II score coefficient of the \( i \)th subtest multiplied by 100 to remove decimal points, \( X_i \) is the participant’s scaled score on the \( i \)th subtest, and \( \bar{M} \) is the participant’s average scaled score across all eleven subtests.

Diagnostic Utility. There are four possible outcomes when using a LDI score to diagnose learning disabilities: true positive, true negative, false positive, and false negative. Two outcomes are correct (true positive and true negative) and two are incorrect (false positive and false negative). True positives are children with learning disabilities who are correctly identified as such by the LDI. False positives are children identified by the LDI as having a learning disability who do not actually have a learning disability. In contrast, false negatives are children with learning disabilities who are not identified by the LDI as having learning disabilities. A test with a low false negative rate has high sensitivity and a test with a low false positive rate has high specificity.

Although sensitivity and specificity are both desirable attributes of a diagnostic test, they are dependent on cutoff score and prevalence rate. Thus, neither provides a unique measure of diagnostic accuracy (McFall & Treat, 1999). In contrast, by systematically using all possible cutoff scores of a diagnostic test and graphing true positive against false positive decision rates, the full range of that test’s diagnostic utility can be determined. Designated the receiver operating characteristic (ROC), this procedure was originally applied more than 50 years ago to determine how well an electronic receiver was able to distinguish signal from noise (Dawson-Saunders & Trapp, 1990). Because they are not confounded by cutoff scores and prevalence rates, ROC methods have

---

TABLE 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>LD(^a)</th>
<th>Reading LD(^b)</th>
<th>Math LD(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Information</td>
<td>7.73</td>
<td>2.75</td>
<td>8.33</td>
</tr>
<tr>
<td>Similarities</td>
<td>8.36</td>
<td>3.11</td>
<td>9.11</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>7.33</td>
<td>2.58</td>
<td>8.59</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>7.79</td>
<td>2.89</td>
<td>8.47</td>
</tr>
<tr>
<td>Comprehension</td>
<td>8.74</td>
<td>3.25</td>
<td>9.63</td>
</tr>
<tr>
<td>Digit Span</td>
<td>7.32</td>
<td>2.50</td>
<td>7.82</td>
</tr>
<tr>
<td>Picture Completion</td>
<td>9.47</td>
<td>2.92</td>
<td>10.10</td>
</tr>
<tr>
<td>Picture Arrangement</td>
<td>8.97</td>
<td>3.27</td>
<td>10.02</td>
</tr>
<tr>
<td>Block Design</td>
<td>8.89</td>
<td>3.20</td>
<td>10.13</td>
</tr>
<tr>
<td>Object Assembly</td>
<td>9.31</td>
<td>3.12</td>
<td>10.28</td>
</tr>
<tr>
<td>Coding</td>
<td>8.43</td>
<td>3.21</td>
<td>9.47</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>88.8</td>
<td>13.3</td>
<td>93.5</td>
</tr>
<tr>
<td>Performance IQ</td>
<td>94.2</td>
<td>13.9</td>
<td>100.4</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>90.5</td>
<td>12.8</td>
<td>96.3</td>
</tr>
<tr>
<td>Reading</td>
<td>81.1</td>
<td>13.2</td>
<td>74.2</td>
</tr>
<tr>
<td>Math</td>
<td>84.9</td>
<td>14.4</td>
<td>94.4</td>
</tr>
<tr>
<td>Writing</td>
<td>76.8</td>
<td>11.1</td>
<td>77.7</td>
</tr>
<tr>
<td>LDI</td>
<td>140.0</td>
<td>356.0</td>
<td>177.7</td>
</tr>
</tbody>
</table>

\(^a\) \( n = 2,053. \(^b\) \( n = 445. \(^c\) \( n = 168. \)
subsequently been widely adopted in the physical (Swets, 1988), medical (Dawson-Saunders & Trapp, 1990; Hsiao, Bartko, & Potter, 1989), and psychological (Swets, 1996) sciences. They have also found occasional application in special education (Harber, 1981). More recently, ROC methods were strongly endorsed for judging the accuracy of psychological assessments (McFall & Treat, 1999; Swets, Dawes, & Monahan, 2000).

As illustrated in Figure 1, the diagonal dashed line is the random ROC or chance line that indicates zero discriminating power. 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. The accuracy of a ROC can be quantified by calculating the area under its curve. Although both parametric and nonparametric calculation methods are available (Swets, 1988), nonparametric methods produce accurate area estimates without assuming that distributions are normal and of equal variance (Center, 1985; Swets, 1996). Consequently, a nonparametric method was used to calculate the area under the curve (Hanley & McNeil, 1982). Chance diagnostic performance corresponds to an area under the curve of .50, whereas perfect diagnostic performance equates to 1.00. The area under the curve is independent of the cutoff score and the base rate and does not assume that the underlying score distributions are normal. It can be interpreted in terms of two children, one drawn randomly from the distribution of children with the target disorder and one selected randomly from the population of children without the disorder. The area under the curve is the probability of the test correctly rank ordering the children into their appropriate diagnostic groups (Hanley & McNeil, 1982). According to Swets (1996), areas under the curve between .50 and .70 are characterized as showing low accuracy, .70 to .90 represent medium accuracy, and .90 to 1.00 denote high accuracy.

### Results

LDI scores for students from the WISC-III standardization sample averaged 328.9 (see Note). These results are similar to LDI scores from the WISC-R standardization sample (viz., M = 3.2, SD = 306.4; Lawson & Inglis, 1984). In contrast, LDI scores for students with learning disabilities are presented in Table 1. As with the WISC-R (Bellemare et al., 1986; Clampit & Silver, 1990; Lawson & Inglis, 1985), LDI scores of the WISC-III standardization sample were statistically significantly different from LDI scores of students with learning disabilities, t(4,251) = 13.93, p < .001, students with specific reading disabilities, t(2,643) = 10.70, p < .001, and students with specific math disabilities, t(2,366) = 2.40, p = .017. Effect sizes ranged from .19 to .56.

ROC analyses indicated that LDI scores exhibited low diagnostic utility (Swets, 1996). As illustrated in Figure 1, an area under the curve of .61 resulted when students with learning disabilities were compared to students from the WISC-III normative sample. That is, if one student was randomly selected from the students with learning disabilities and one from the WISC-III normative sample, the probability of the LDI correctly rank ordering them into their appropriate diagnostic groups was .61. Results for students with specific reading disabilities (area under the curve = .64) and specific math disabilities (area under the curve = .55) were also of low diagnostic accuracy. Notably, equivalent diagnostic accuracy was achieved by simply comparing all students with learning disabilities with reading or math achievement scores less than 85 to the WISC-III standardization sample (areas under the curve = .66 and .64, respectively).

### Discussion

The use of cognitive subtest profiles or patterns to aid in the diagnosis of learning disabilities is common in train-
subtests have also proven to have little or no diagnostic utility in identifying children with learning disabilities (Watkins, 1996, 1999; Watkins, Kush, & Glutting, 1997a, 1997b; Watkins & Worrell, 2000). When considered within the broader negative context of subtest profile research (Glutting, McDermott, Konold, Snellbaker, & Watkins, 1998; Kramer, Henning-Stout, Ullman, & Schellenberg, 1987; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992; McDermott & Glutting, 1997; Teeter & Korducki, 1998), the LDI is unsupported as a tool in the diagnosis of learning disabilities. Within the interpretative framework presented by Kamphaus (1998), using the LDI as an indicator of learning disabilities constitutes a case of acting in opposition to the scientific evidence.

ABOUT THE AUTHORS

Marley W. Watkins, PhD, is professor in charge of Graduate Programs in School Psychology at The Pennsylvania State University. His current research interests include cognitive and intellectual theory and assessment. Barbara A. Schaefer, PhD, is an associate professor of school psychology at Duquesne University. Her current research interests are learning behaviors, educational and behavioral assessment, gender differences, and instrument development and scaling. Address: Marley W. Watkins, Pennsylvania State University, Department of Educational and School Psychology and Special Education, 227 CEDAR Bldg., University Park, PA 16802 (e-mail: mww10@psu.edu).

NOTE

Standardization data of the Wechsler Intelligence Scale for Children—Third Edition. Copyright © 1990 by The Psychological Corporation. Used by permission. All rights reserved.

REFERENCES


(continued on p. 136)


Torgesen, J. K., & Mathes, P. G. (2000). A basic guide to understanding, assessing, and teaching phonological awareness. Austin, TX: PRO-ED.


(continued from p. 103)


