Psychological Assessment
2001, Vol. 13, No. 4, 543–548

1040-3590/01/$5.00 DOI: 10.1037//1040-3590.13.4.543

The Frequency of Reliable Component Difference Scores for the Wechsler Intelligence Scale for Children—Third Edition in Two Samples

John C. Caruso and Katie Witkiewitz
University of Montana

Eric A. Youngstrom
Case Western Reserve University

Joseph J. Glutting
University of Delaware

The cumulative percentage frequencies are presented for differences among reliable component analysis (RCA) scores for the verbal comprehension, perceptual organization, freedom from distractibility, and processing speed constructs assessed by the Wechsler Intelligence Scale for Children—Third edition (WISC–III) for the standardization sample and a learning disabled sample. Using RCA scores to form differences has several advantages over traditional equally weighted scores for the WISC–III. J. C. Caruso and N. Cliff (2000) presented tables to assess the statistical significance of differences among the RCA scores for the WISC–III. It is important, however, to use a dual approach in interpreting difference scores; both the statistical significance of a difference and the frequency with which it occurred in a relevant comparison group should be determined. This article contains the information necessary for practitioners to use the recommended dual approach to interpreting RCA difference scores for the WISC–III.

In the course of evaluating the results of an individual administration of the Wechsler Intelligence Scale for Children—Third edition (WISC–III; Wechsler, 1991), differences among the four index-level scores (Verbal Comprehension [VC], Perceptual Organization [PO], Freedom From Distractibility [FD], and Processing Speed [PS]) are often examined. In the WISC–III manual, Wechsler (1991) stated that "an important consideration in interpreting WISC–III results is the amount of difference between the IQ scores or between the factor-based index scores" (p. 173). The WISC–III manual recommends a dual approach to interpreting index score differences: Both the statistical significance of the difference and the frequency of the difference in the standardization sample should be determined. The statistical significance of the difference between two index scores is tested by hypothesizing no actual difference between a person's scores on the two measures (the null hypothesis) and determining if the observed difference refutes that notion. If the difference between the two observed scores is larger than would be expected (under the null hypothesis) at the selected probability level (e.g., $\alpha = .05$), then the examiner can conclude that a true difference exists in the observed direction.

John C. Caruso and Katie Witkiewitz, Department of Psychology, University of Montana; Eric A. Youngstrom, Department of Psychology, Case Western Reserve University; Joseph J. Glutting, Department of Educational Studies, University of Delaware.

This research was funded, in part, by National Science Foundation Grant EPS 9871922. We wish to thank Patricia Bachelder for comments on a draft of this article.

Correspondence concerning this article should be addressed to John C. Caruso, Department of Psychology, University of Montana, 241 Skaggs Building, Missoula, Montana 59812. Electronic mail may be sent to jcaruso@selway.umt.edu.

The other method for interpreting WISC–III difference scores is to examine the frequency of the observed difference in a relevant comparison group. Knowing the frequency with which a particular difference occurs in a relevant comparison group allows us to assess the rarity, or abnormality, of that difference. Payne and Jones (1957) originally described the utility of using the dual approach for appropriately interpreting difference scores. Silverstein (1981) discussed the abnormality of difference scores for the Wechsler Intelligence Scale for Children—Revised (Wechsler, 1974) and portrayed, through a hypothetical example, the problem that might arise when basing interpretation on only the statistical significance of a difference:

> Discrepancies of at least 9, 12, or 16 points will be obtained by 49, 35, or 20% of all subjects, so that while such discrepancies are *reliable* [note that Silverstein equated "*statistically* significant" with "reliable"], they hardly can be regarded as *abnormal*. (p. 394)

The usefulness of the dual approach was subsequently emphasized by several other authors (e.g., Berk, 1982; Kaufman, 1976; Piotrowski, 1978; Reynolds, 1979). The WISC–III manual includes a table (Table B.1; p. 261) that provides the index score differences required for statistical significance (at the .15 and .05 α levels) and a table (Table B.2; p. 262) that presents the frequency of index score discrepancies found in the standardization sample. Both tables are based on differences between equally weighted index scores, which is the method of score construction recommended in the WISC–III manual.

One problem with interpreting difference scores, however, is the fact that they tend to be unreliable. Several authors have recommended that difference-score interpretations should be disregarded, or at least approached with great caution, because of this tendency (e.g., Cattell, 1982; Lord & Novick, 1968; Stanley, 1967). A primary reason that difference scores tend to be unreliable is because the scores on which they are based are often strongly correlated. The formula for the reliability of a difference (between scores with equal means and standard deviations),

$$r_D = \frac{r_{11} + r_{22} - 2r_{12}}{2 - 2r_{12}},$$

where r_{11} and r_{22} are the reliabilities of the individual scores and r_{12} is the correlation between them, makes clear why this is the case: A strong correlation between the variables included in the difference results in a reduction in the reliability of the difference score, whereas a low (or nonexistent) correlation between the constituent variables results in high difference-score reliability. In fact, when scores on the two measures are perfectly uncorrelated, the reliability of the difference score is simply the average of the reliabilities of the two constituent scores and typically a very respectable value.

Reliable Component Scores

To deal with the problem of unreliability, Caruso and Cliff (2000) provided a new method of computing differences for the WISC–III. They recommended the use of reliable component analysis (RCA) to define scores for the constructs assessed by the traditional index scores and to compute differences based on the RCA scores as opposed to the equally weighted scores. Although the RCA scores themselves were shown to have several valuable properties, the primary advantage that RCA difference scores have is high reliability. The advantage that the RCA difference scores have over equally weighted ones in terms of reliability was as high as .10 in the standardization sample: The reliability of the difference between the RCA scores for VCC and POC was .88, compared with .78 for the corresponding equally weighted difference, and the reliability of the difference between the RCA scores for VCC and FDC was .86, compared with .76 for the corresponding equally weighted difference.[1] The RCA method provided a reliability advantage of between .03 and .07 for the other four WISC–III difference scores. The reasons that the RCA difference scores are more reliable than equally weighted ones are the high reliability of the RCA scores (r_{11} and r_{22} in the equation are high) and the fact that the scores are uncorrelated (r_{12} in the equation equals zero).

To aid in the interpretation of RCA differences, Caruso and Cliff (2000) presented the differences required for statistical significance at the .15 and .05 α levels (Table 7, p. 93) that are analogous to those presented in Table B.1 of the WISC–III manual for equally weighted difference scores. However, no information on the frequency of differences was given, and hence RCA score users are currently unable to use the dual approach to difference-score interpretation. Table B.2 of the WISC–III manual cannot be used for RCA difference scores because the standard deviation of a difference depends on the correlation between the constituent scores and these correlations are not the same for RCA and equally weighted scores.

Purpose

The main purpose of this article is to provide those practitioners who use the RCA method of computing difference scores for the WISC–III with the information necessary to use the dual approach to interpreting score differences. Additionally, we use this opportunity to examine certain properties of the RCA scores in a sample that is learning disabled.

Method

Data

Two samples were examined here to provide researchers and practitioners with a choice of relevant comparison groups. The first sample (described in greater detail in the WISC–III manual) is the standardization sample (Wechsler, 1991), which consisted of 2,200 children with a mean age of 11.00 years (SD = 3.11) distributed across an age range of 6 to 16 years. The second is a sample of 615 children from the eastern region of Virginia. All of the children were referred by their schools for psychoeducational assessments, and each child was diagnosed with a learning disability. The diagnosis of a learning disability was made by multidisciplinary teams on the basis of criteria established by the school division, which are consistent with the state guidelines in Virginia. The mean age of the sample was 10.67 years (SD = 2.47), and the range of ages was from 6 to 19 years (only 7 participants in this sample were older than 16 years of age). The majority of children were male (67.8%). With respect to ethnicity, 64.6% were White, 29.6% were African American, 3.8% were Hispanic, 1.0% were Asian American or Pacific Islanders, and 0.2% were Native American.

Procedure

For each sample, equally weighted index scores were computed as recommended in the WISC–III manual for VC, PO, PS, and FD. Differences among these scores were then derived through the simple difference method (as recommended in the WISC–III manual). As noted above, Caruso and Cliff (2000) analyzed the standardization sample of the WISC–III with RCA, and the weights they presented were used to compute the VCC, POC, FDC, and PSC scores for each sample examined here. Differences among RCA scores were also computed using the simple difference method.

Results

Table 1 provides descriptive statistics for each sample. The correlations among RCA scores are below the diagonal, those among equally weighted scores are above the diagonal, and those between corresponding RCA and equally weighted scores are on the diagonal. In the standardization sample, the RCA scores have zero correlations, but the equally weighted scores are strongly correlated. The correlations are exactly zero in the standardization sample because the original RCA for the WISC–III (Caruso & Cliff, 2000) was performed on these data and the components are orthogonal. For equal weighting, the correlation of .63 between VC and PO is the strongest and indicates that the two scores share nearly 40% of their variance. For the sample that was learning disabled, the correlations between RCA scores range from −.21 (for VCC and POC) to .03 (for VCC and PSC), whereas those for equal weighting range from .26 to .47. Although not exactly zero,

[1] In Caruso and Cliff (2000) and here, we follow the two-letter acronym for each of the index scores with an additional C (for component) to differentiate the RCA scores from their equally weighted counterparts. For example, VC is the equally weighted Verbal Comprehension index, and VCC is the verbal comprehension scores derived through RCA.

Table 1
Descriptive Statistics for WISC–III Scores for RCA and Equal Weighting

	Equal weighting					
RCA	VC	PO	FD	PS	*M*	*SD*
	Standardization sample					
VCC	.92	.63	.59	.39	100.00	15.00
POC	.00	.85	.51	.46	100.00	15.00
FDC	.00	.00	.77	.40	100.00	15.00
PSC	.00	.00	.00	.93	100.00	15.00
M	100.29	100.33	101.14	100.88		
SD	15.00	15.10	14.46	14.84		
	Learning disabled sample					
VCC	.89	.47	.44	.37	93.17	15.39
POC	−.21	.87	.36	.45	93.64	18.69
FDC	−.17	−.11	.68	.26	91.32	14.08
PSC	.03	.01	−.13	.94	95.06	16.36
M	91.40	91.24	88.07	94.20		
SD	13.58	15.89	11.86	14.62		

Note. Reliable component analysis (RCA) score correlations are below the diagonal, and equally weighted score correlations are above the diagonal. WISC–III = Wechsler Intelligence Scale for Children—Third edition; VC = Verbal Comprehension index; PO = Perceptual Organization index; FD = Freedom From Distractibility index; PS = Processing Speed index; VCC = verbal comprehension component; POC = perceptual organization component; FDC = freedom from distractibility component; PSC = processing speed component.

the correlations between the RCA scores indicate that they share much less variance (between 0 and 4%) than the equally weighted scores (between 7% and 22%). The large proportion of variance shared by the equally weighted scores results in low reliability for their differences (see Caruso & Cliff, 2000). The values on the diagonals of the correlation matrices presented in Table 1 indicate that the RCA and equally weighted scores have good convergent validity in both the standardization sample (with correlations between .77 [FD and FDC] and .93 [PS and PSC]) and in the learning disabled sample (with correlations between .68 [FD and FDC] and .94 [PS and PSC]). Note that the means and the standard deviations for the RCA scores in the standardization sample are identical for each WISC–III scale because of the way the scores were constructed (see Caruso & Cliff, 2000). In the learning disabled sample, the means for the RCA scores are generally higher than the means for the equally weighted scores. This can be interpreted as a positive aspect of the RCA scores because the WISC–III is explicitly a measure of ability, not achievement, and a common definition of a learning disability is average ability and below average achievement. Therefore, the RCA scores appear to be less biased against the learning disabled individuals in this sample.

Table 2 contains the means and standard deviations for the RCA and equally weighted score differences. The RCA difference scores always have larger standard deviations than the equally weighted ones because the RCA scores have low (or zero) correlations. The equally weighted difference scores are more strongly correlated and therefore have much smaller standard deviations. The means of the RCA difference scores tend to be smaller than those for equal weighting because the RCA score means are less variable (see Table 1).

Table 3 contains the cumulative percentage frequencies of difference scores derived from RCA scores for the standardization sample and the learning disabled sample. Although the main purpose of this table is to allow researchers and practitioners to compare their observed differences to two relevant comparison groups, the values can be examined for a substantive interpretation of the comparability of the two samples, and several points are worth noting. First, the cumulative frequencies for the VCC-FDC, VCC-PSC, and FDC-PSC differences are highly similar in the two samples. Second, there is a tendency for larger VCC-POC, POC-PSC, and POC-FDC differences to exist in the learning disabled sample. Thus, each of the differences involving POC tends to be larger in the learning disabled sample. The bottom of this table contains the means and standard deviations of the absolute values of the differences, as are provided in the WISC–III manual for equally weighted scores.

Discussion

It is a relatively simple process to use Table 3 as an aid in interpreting the results of an administration of the WISC–III when the RCA scoring method is used, and we provide an example here. Suppose that a child receives RCA scores of 100 on VCC and 80 on POC. The difference between these two scores is 20, and on the basis of Table 7 from Caruso and Cliff (2000), this difference is statistically significant at .05. Our conclusion, if our analysis ended there, would be that a potentially actionable discrepancy had been found. After all, if we can be confident that a discrepancy exists, should we not ask why and determine if some remediation is appropriate? However, examining the frequency with which a discrepancy this large or larger occurred in the standardization sample provides additional valuable information: The percentage of the standardization sample with a difference this large or larger,

Table 2
Means and Standard Deviations for RCA and Equal Weighting Score Differences

	RCA		Equal weight	
Difference	*M*	*SD*	*M*	*SD*
	Standardization sample			
VCC-POC	0.00	21.21	−0.04	12.94
VCC-FDC	0.00	21.21	−0.84	13.28
VCC-PSC	0.00	21.21	−0.59	16.49
POC-FDC	0.00	21.21	−0.81	14.66
POC-PSC	0.00	21.21	−0.55	15.47
FDC-PSC	0.00	21.21	0.26	16.02
	Learning disabled sample			
VCC-POC	−0.47	26.58	0.16	15.27
VCC-FDC	1.85	22.59	3.33	13.53
VCC-PSC	−1.89	22.13	−2.80	15.82
POC-FDC	2.32	24.75	3.17	16.03
POC-PSC	−1.42	24.75	−2.96	16.00
FDC-PSC	−3.74	22.89	−6.13	16.29

Note. RCA = reliable component analysis; VCC = verbal comprehension component; POC = perceptual organization component; FDC = freedom from distractibility component; PSC = processing speed component.

Table 3
Cumulative Percentages of the Standardization Sample and Learning Disabled Sample Obtaining Various RCA Score Discrepancies

	WISC–III RCA difference score											
	VCC-POC		VCC-FDC		VCC-PSC		POC-FDC		POC-PSC		FDC-PSC	
Diff	SDS	LDS	SDS	LDS	SDS	LDS	SDS	LDS	SDS	LDS	SDS	LDS
≥49	2.5	9.4	1.9	3.9	2.0	3.9	2.4	5.7	1.8	3.4	2.4	4.4
48	2.8	10.1	2.2	3.9	2.5	4.1	2.8	6.2	2.0	4.1	2.7	4.7
47	3.1	10.4	2.6	3.9	2.7	4.6	3.1	6.2	2.4	4.9	2.9	5.4
46	3.7	10.7	3.0	4.1	3.1	4.6	3.5	6.7	2.9	5.4	3.5	5.4
45	4.4	11.5	3.3	4.1	3.6	5.5	4.0	7.0	3.4	5.4	3.9	5.4
44	4.8	13.0	3.8	4.7	4.1	6.0	4.5	8.1	3.8	7.2	4.3	5.4
43	5.3	13.2	4.2	4.7	4.5	6.2	5.0	8.1	4.1	7.3	4.7	5.7
42	5.7	13.5	4.7	6.3	5.3	7.3	5.5	8.1	4.6	8.6	5.0	5.9
41	6.0	13.8	5.0	6.5	5.6	7.6	6.0	8.1	5.2	9.6	5.7	6.3
40	6.2	15.3	5.7	7.6	6.1	7.8	6.7	9.1	5.7	10.2	6.3	6.3
39	6.9	16.3	6.6	9.3	6.8	8.1	7.3	9.6	6.7	11.7	6.8	7.0
38	7.4	16.6	7.6	10.4	7.5	8.1	8.3	10.7	7.8	12.8	7.4	7.5
37	8.1	17.1	8.6	10.9	8.1	9.3	9.1	12.4	8.6	14.5	8.4	7.5
36	8.8	18.4	9.2	11.1	9.1	10.7	10.0	13.7	9.6	16.1	9.3	9.6
35	9.7	19.5	10.1	11.7	10.2	11.9	10.5	14.1	10.5	17.1	10.7	10.2
34	10.6	22.1	11.5	12.5	11.5	14.1	11.4	16.4	11.5	18.5	11.6	12.4
33	11.9	22.8	12.2	13.8	12.3	14.6	12.3	16.7	12.5	23.4	12.5	13.3
32	13.0	24.1	14.0	14.1	13.4	14.6	13.7	18.4	14.1	25.0	13.7	15.3
31	14.0	25.4	15.1	16.1	14.7	15.3	14.7	19.5	15.5	26.5	14.8	17.7
30	15.1	26.5	16.6	16.4	16.0	17.4	15.8	23.3	17.6	27.8	16.2	19.7
29	16.8	29.4	17.7	19.2	17.8	18.5	17.3	25.7	19.4	29.4	18.0	20.3
28	18.6	30.6	19.6	19.5	19.5	19.0	18.9	28.3	20.8	31.1	19.6	22.1
27	20.4	31.5	21.9	20.5	21.2	20.2	20.6	29.8	22.3	33.0	21.2	22.4
26	21.9	32.5	23.3	22.8	23.0	22.0	22.1	31.1	24.2	34.0	22.8	23.1
25	23.7	33.2	25.0	25.7	25.2	22.8	23.6	32.0	26.0	35.8	24.9	26.8
24	25.8	34.8	26.7	29.8	27.0	25.9	25.6	35.6	27.7	36.6	26.6	28.9
23	28.3	35.9	29.1	32.4	29.1	27.3	28.2	38.5	30.0	38.0	28.8	29.4
22	31.1	37.1	31.7	35.8	31.7	28.0	30.8	39.5	32.2	40.0	31.2	31.2
21	33.1	40.3	33.6	37.2	34.2	32.0	33.4	41.6	34.3	41.0	33.0	33.0
20	35.5	41.6	35.7	40.5	36.5	33.5	35.9	43.6	36.8	42.9	35.8	36.6
19	38.1	43.4	38.3	41.5	39.0	35.3	38.0	45.5	38.9	44.9	37.8	39.3
18	40.4	45.7	40.3	45.0	41.3	37.9	40.5	48.0	41.5	45.4	40.1	43.1
17	43.5	48.5	43.6	48.6	43.9	41.0	43.7	49.6	43.8	48.8	43.0	44.2
16	46.2	49.6	46.5	51.9	46.7	43.7	46.2	51.9	46.3	53.7	45.5	48.0
15	49.9	51.7	49.6	53.2	50.4	47.2	49.0	55.0	49.1	57.2	48.1	48.9
14	53.5	55.1	53.2	56.7	53.7	53.5	52.1	56.4	52.1	60.0	51.5	53.0
13	56.0	59.0	56.2	58.9	57.2	55.6	55.7	59.5	54.7	62.9	54.2	55.6
12	59.4	61.6	59.9	61.3	60.3	59.3	59.2	62.9	57.4	65.7	57.0	58.4
11	62.3	65.5	62.9	66.0	63.7	64.4	62.2	65.2	61.1	67.6	60.3	60.0
10	65.0	68.9	66.8	68.1	67.1	68.0	65.6	69.9	64.2	70.4	64.0	65.2
9	67.9	70.9	69.2	73.7	70.0	71.1	69.4	73.3	67.9	72.2	66.9	71.9
8	71.4	73.7	72.6	75.1	73.3	74.1	73.1	75.1	71.5	75.6	71.2	73.8
7	75.3	77.2	75.8	78.5	76.8	77.2	76.7	76.6	74.4	77.6	75.0	77.1
6	79.1	82.4	78.9	82.6	80.3	81.6	80.1	81.3	78.6	81.3	79.5	80.3
5	83.1	86.3	82.6	85.2	84.2	84.4	84.3	83.7	82.5	84.7	83.1	85.5
4	87.0	88.3	86.7	89.6	88.2	91.2	87.8	88.6	86.5	87.6	86.8	89.6
3	90.7	91.4	90.4	91.9	91.6	94.6	91.0	92.2	90.0	91.9	90.2	94.3
2	94.2	94.0	94.3	94.6	95.1	96.3	95.2	97.1	94.5	94.5	94.0	95.4
1	98.0	96.4	98.2	99.2	98.3	99.5	98.5	98.9	98.5	98.0	98.6	98.9
0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
M	16.9	20.6	17.1	18.1	16.9	17.5	17.0	19.6	16.9	19.9	16.8	18.0
SD	12.9	16.8	12.7	13.6	12.7	13.8	12.8	15.1	12.8	14.7	14.1	14.6

Note. RCA = reliable component analysis; WISC–III = Wechsler Intelligence Scale for Children—Third edition; VCC = verbal comprehension component; POC = perceptual organization component; FDC = freedom from distractibility component; PSC = processing speed component; SDS = standardization sample; LDS = learning disabled sample.

found in Table 3, is 35.5%. If we were to suggest remediation for this child, we would need to make the same suggestion for over one third of the population, putting this child's discrepancy in perspective.

The identification of a difference of 20 as statistically significant is made possible by the high reliability of the VCC-POC difference score (.88; Table 4 of Caruso & Cliff, 2000), but that says nothing about whether the difference is rare, abnormal, important, or deserving of a remediation plan. Larger differences occur less frequently, and at some point determined by the individual practitioner, the difference will be sufficiently rare to warrant further action. The definition of rare that the individual practitioner selects may be determined by resource availability, other theoretical considerations, or, by default, one might use 5% or 10% as a reasonable definition (cf. Cahan, 1986; Glutting, McGrath, Kamphaus, & McDermott, 1992; Silverstein, 1981). The strength of the dual approach is that we can take measurement reliability into account when determining the statistical significance of the difference and then examine the importance of the difference by associating the examinee's difference score with a relevant comparison group.

Using this article and the RCA weights provided in Caruso and Cliff (2000), practitioners will now be able to use RCA to calculate highly reliable difference scores and to examine those differences for both statistical significance and rarity. Furthermore, the rarity of observed differences can be interpreted relative to a "normal" sample (i.e., the standardization sample) or a large learning disabled sample. In general, the standardization sample would provide the more relevant comparison group, as in the above example. Some practitioners or researchers, however, may be interested in the placement of a child within a sample of children formally diagnosed with a learning disability. For example, special education service providers may use the information to gauge the level of service appropriate for a child. A child with a severe discrepancy, relative to other children with learning disabilities, may have an individual education plan that includes a large amount of individual instruction, whereas a child whose scores indicate a less severe relative placement may have an educational plan that includes more group instruction. Of course, other information will be used to develop the educational plan for a child, but the rarity of a discrepancy in the learning disabled sample may be helpful. As more and more students are diagnosed with a learning disability (U.S. Department of Education, 1999), an issue that is becoming increasingly salient is the availability of resources for special education programs and the effect that various scoring methods and cutoff scores have on the number of children declared eligible (Braden & Weiss, 1988; Clarizio & Phillips, 1989; Evans, 1992; Payette & Clarizio, 1994). Comparing a child diagnosed with a learning disability with others with the same diagnosis should be helpful in determining where resources can be most appropriately devoted.

We recommend the following steps be taken when using the RCA approach. First, compute the RCA scores for VCC, POC, FDC, and PSC using the weights provided in Table 1 of Caruso and Cliff (2000). Then compute the difference scores and examine them for statistical significance using Table 7 from the same article. Next, for those differences that are statistically significant at a chosen α level, examine the frequency with which they occurred in the standardization sample using Table 3 of this article. Those differences that are both statistically significant and sufficiently large to be considered rare in the standardization sample should then be the focus of test interpretation. When it is determined that the differences, combined with supporting information from other tests or other sources, are deserving of some individual education plan, the test interpreter may wish to compare the child's score with those of the learning disabled sample for assistance in determining the level of service to provide, as discussed above.

This study has certain limitations. First and foremost, the sample sizes were not large enough at the group level to provide separate analyses for individuals varying with respect to ethnicity. Future research examining the frequency of RCA differences in specific ethnic groups, as well as other aspects of the RCA scores such as their intercorrelation and reliability, would certainly be helpful for researchers who work with minority populations. Also, some practitioners may be hesitant to use differentially weighted scores because the weights are sample dependent (Cohen, 1990) or because the scores are more complicated to compute than the traditional equally weighted scores (Nunnally, 1970). We have addressed these issues elsewhere (Caruso & Cliff, 2000), but the basic defense of differential weighting is that the differential weights are often (as in the case of the WISC–III) highly replicable across samples, and that the added time necessary to use the differential weighting system is trivial compared with the total administration, scoring, and interpretation time for the test. Furthermore, there is no reason to believe that each of the subtests of the WISC–III will contribute equally to their respective factor score. The constructs for which scores are derived on the WISC–III are merely reifications, and it therefore seems allowable (and logical) to construct the scores in the manner that is most advantageous. The equally weighted index scores have several psychometric weaknesses, and this results in questionable validity (see, e.g., Glutting, Youngstrom, Ward, Ward, & Hale, 1997) and potential bias (Perloff & Persons, 1988). The uncorrelatedness (or near uncorrelatedness in the learning disabled sample) of the RCA scores is a valuable property in and of itself that additionally leads to other advantages (e.g., high reliability for differences), and for these reasons, we recommend that RCA be used to compute scores for the WISC–III. The present results will allow for a more comprehensive analysis of differences computed from RCA scores.

References

Berk, R. A. (1982). Verbal–Performance IQ discrepancy score: A comment on reliability, abnormality, and validity. *Journal of Clinical Psychology, 38,* 636–641.

Braden, J. P., & Weiss, L. (1988). Effects of simple difference versus regression discrepancy methods: An empirical study. *Journal of School Psychology, 26,* 133–142.

Cahan, S. (1986). Significance testing of subtest score differences: The rules of the game. *Journal of Psychoeducational Assessment, 4,* 273–280.

Caruso, J. C., & Cliff, N. (2000). Increasing the reliability of Wechsler Intelligence Scale for Children—Third Edition difference scores with reliable component analysis. *Psychological Assessment, 12,* 89–96.

Cattell, R. B. (1982). The clinical use of difference scores: Some psychometric problems. *Multivariate Experimental Clinical Research, 6,* 87–98.

Clarizio, H. F., & Phillips, S. E. (1989). Defining severe discrepancy in the diagnosis of learning disabilities: A comparison of method. *Journal of School Psychology, 27,* 383–391.

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

Evans, L. D. (1992). A comparison of the impact of regression and simple difference discrepancy. *Journal of School Psychology, 30,* 17–29.

Glutting, J. J., McGrath, E. A., Kamphaus, R. W., & McDermott, P. A. (1992). Taxonomy and validity of subtest profiles on the Kaufman Assessment Battery for Children. *Journal of Special Education, 26,* 85–115.

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.

Kaufman, A. S. (1976). Verbal-Performance IQ discrepancies on the WISC–R. *Journal of Consulting and Clinical Psychology, 44,* 739–744.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Nunnally, J. C. (1970). *Introduction to psychological measurement.* New York: McGraw-Hill.

Payette, K. A., & Clarizio, H. F. (1994). Discrepant team decisions: The effects of race, gender, achievement, and IQ on LD eligibility. *Psychology in the Schools, 31,* 40–48.

Payne, R. W., & Jones, H. G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 18,* 115–121.

Perloff, J. M., & Persons, J. B. (1988). Biases resulting from the use of indexes: An application to attributional style and depression. *Psychological Bulletin, 103,* 95–104.

Piotrowski, R. J. (1978). Abnormality of subtest score differences on the WISC–R. *Journal of Consulting and Clinical Psychology, 46,* 569–570.

Reynolds, C. R. (1979). Interpreting the index of abnormality when the distribution of score differences is known: Comment on Piotrowski. *Journal of Consulting and Clinical Psychology, 47,* 401–402.

Silverstein, A. B. (1981). Reliability and abnormality of test score differences. *Journal of Clinical Psychology, 37,* 392–394.

Stanley, J. C. (1967). General and special formulas for the reliability of differences. *Journal of Educational Measurement, 4,* 249–252.

U.S. Department of Education. (1999). *Twentieth annual report to Congress on the implementation of the Individuals with Disabilities Education Act.* Washington, DC: U.S. Government Printing Office.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised manual.* New York: Psychological Corporation.

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

Received November 2, 2000
Revision received May 23, 2001
Accepted May 29, 2001 ■