Archives of Scientific Psychology 2019, 7, 119–128
DOI: http://dx.doi.org/10.1037/arc0000064

2169-3269

American Psychological Association

Archives of Scientific Psychology

www.apa.org/pubs/journals/arc

Profile Reliability of Cognitive Ability Subscores in a Referred Sample

Kara M. Styck
Northern Illinois University

A. Alexander Beaujean and Marley W. Watkins
Baylor University

ABSTRACT

The multidimensionality of intelligence has become commonly accepted among psychologists. As a result, the question "How intelligent is an individual?" has been replaced by the question "In what ways is an individual intelligent?" The construction of modern intelligence tests has followed suit and most intelligence tests today provide scores for some general intellectual attribute as well as multiple specific types of intellectual attributes. This has led to the common practice of interpreting profiles of intellectual strengths and weaknesses, with the subsequent conclusion that these profiles represent real differences in individuals' underlying intellectual attributes. These conclusions are premature, however, because they assume intelligence tests measure these specific intellectual attributes well. A necessary condition for interpreting score profiles is consistency—an individual's profile should be relatively similar across time. The purpose of our study was to evaluate the consistency of intelligence test score profiles on a sample of children who were given a widely used intelligence test two times. We found that strengths and weaknesses of specific types of intelligence were not measured consistently. Thus, although "In what ways is an individual intelligent?" may be the question psychologists want to answer, results of this study suggest that we are currently able to answer only the question, "How intelligent is an individual?"

SCIENTIFIC ABSTRACT

Clinical profile analysis of intelligence test subscores remains a popular practice among psychologists who work in applied settings, despite decades of accumulating evidence indicating that IQ subscores have poor psychometric properties. Bulut, Davison, and Rodriguez (2017) recently developed a method to estimate the within-person (profile pattern) and between-person (profile level) reliability of subscores. Given that reliability is a necessary, albeit insufficient, condition for score interpretation, the purpose of the present investigation was to estimate the within-person and between-person profile reliability for intelligence test subscores using a contemporary version of the Wechsler intelligence scales using a sample of children ($N = 296$) twice assessed for special education eligibility. Results indicated that between-person reliability estimates were higher than within-person reliability estimates at both the subtest (.79 vs. .37) and index score (.78 vs. .53) levels of interpretation, indicating that the profiles were not very reliable. Moreover, this pattern of results remained consistent even when evaluating a subsample of students diagnosed with specific learning disabilities. These findings contribute to the empirical literature base that indicates the interpretation of intelligence test subscore profiles is not psychometrically defensible.

Keywords: clinical profile analysis, cognitive ability, subscores, specific learning disability
Data repository: http://dx.doi.org/10.3886/ICPSR37285.v1

This article was published December 23, 2019.

Kara M. Styck, Department of Psychology, Northern Illinois University; A. Alexander Beaujean, Department of Psychology & Neuroscience, Baylor University; Marley W. Watkins, Department of Educational Psychology, Baylor University.

The authors have made available the data that underlie the analyses presented in this article (see Styck, Beaujean, & Watkins, 2019), thus allowing replication and potential extensions of this work by qualified researchers. Next users are obligated to involve the data originators in their publication plans, if the originators so desire.

Correspondence concerning this article should be addressed to Kara M. Styck, Department of Psychology, Northern Illinois University, 1425 West Lincoln Highway, DeKalb, IL 60115. E-mail: kstyck@niu.edu

Clinical profile analysis (CPA) has been around almost as long as psychological tests (Beaujean & Benson, in press). The idea behind CPA is that score patterns are more useful to interpret than the scores themselves. According to this logic, psychological tests exhibit clinical utility by estimating some "nonlinear joint functions" of the scores comprising a given profile, such as the mean (i.e., elevation), variability of scores about the mean (i.e., scatter), or the location of high and low scores (i.e., shape; Cronbach & Gleser, 1953; Lykken, 1956).

CPA has been applied to many types of psychological tests, including vocational (e.g., Gottfredson & Jones, 1993; Jones, 1989; Maurer & Tarulli, 1997), personality (e.g., Meehl, 1946; Voglmaier et al., 2005; Voglmaier, Seidman, Salisbury, & McCarley, 1997), and intelligence (e.g., Beeldman et al., 2016; Flanagan, Ortiz, & Alfonso, 2013; Letteri, 1980; Raaphorst, de Visser, Linssen, de Haan, & Schmand, 2010; Rizza, McIntosh, & McCunn, 2001).

The use of CPA with intelligence tests became popular after the publication of the Wechsler–Bellevue (Wechsler, 1946). Clinicians, usually with a psychodynamic orientation, believed they could assess noncognitive attributes based on the pattern of scores (Kamphaus, Winsor, Rowe, & Kim, 2012; Sugarman & Kanner, 2000). Since then, score patterns from intelligence tests have been used for a range of purposes, such as diagnosing psychopathology (e.g., specific learning disabilities, autism, attention-deficit/hyperactivity disorder, Canivez, 2013), determining cognitive strengths and weaknesses (Hale et al., 2010; Ortiz, 2015), and developing interventions (Braden & Kratochwill, 1997).

The fervor with which some advocates have promoted CPA has had a profound impact on applied practice. Surveys of psychologists who work in a variety of settings have indicated the widespread use of CPA for individual decision-making (Maki & Adams, 2019; Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000; Sotelo-Dynega & Dixon, 2014). Moreover, the technical and interpretive manuals of all three contemporary iterations of the Wechsler intelligence scales (Wechsler, 2008, 2012, 2014a), the Woodcock-Johnston Tests of Cognitive Ability–Fourth Edition (Schrank, McGrew, & Mather, 2014), and the Stanford-Binet–Fifth Edition (Roid, 2003) contain instructions on how to conduct CPA and recommend it as a means of test score interpretation, without mention of any contradictory peer-reviewed research.

Problems With Interpreting Clinical Profiles

Despite its popularity, methodological research has consistently cautioned against CPA of intelligence test scores (McGill, Dombrowski, & Canivez, 2018; Watkins, 2000). First, for clinical profiles to have meaning, subscores (e.g., subtests, index scores) need to be distinct from aggregate scores (Bulut et al., 2017). Yet, most modern intelligence tests are constructed such that a single aggregate score explains the majority of the variance in test scores (e.g., Canivez, 2014; Canivez, Watkins, & Dombrowski, 2017; Dombrowski, McGill, & Canivez, 2018; Watkins & Beaujean, 2014) and information from inherently unidimensional tests cannot be decomposed to produce useful multidimensional profiles of subscores (Luecht, Gierl, Tan, & Huff, 2006).

Second, subscores should have sufficient validity and reliability evidence for interpretation (Bulut et al., 2017). If subscores assess their target attributes poorly, then the information they yield may not be trustworthy (Sinharay, Puhan, & Haberman, 2011). Because CPA requires interpretating scatter, it shifts the interpretation unit from the original scores to ipsatized versions of the scores (Cattell, 1944). That is, interpretation moves from how individuals perform in comparison to their same age peers to how do individuals perform in comparison to themselves. Thus, it cannot be assumed that any properties of the original subscores apply to the profiles (McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992).

CPA relies heavily upon the existence of statistically significant within-person subscore differences to identify profiles worthy of interpretation and minimizes the importance of base rates or the abnormality of observed score differences. However, the difference between two subscores can be both "real" (i.e., not due to chance) and common (Silverstein, 1981). For example, the authors of the *Wechsler Intelligence Scales for Children–Fifth Edition Technical and Interpretive Manual Supplement* (WISC-V; Wechsler, 2014b) wrote that special population studies indicated "children identified as [having a specific learning disability in the area of mathematics] demonstrate cognitive weaknesses on the VSI, FRI, and QRI" (p. 13). A 12-year-old examinee may demonstrate a significant relative strength on the WISC-V Verbal Comprehension Index (VCI) of 12 points when compared with his or her Visual Spatial Index (VSI) score ($p < .05$), but this observed score difference occurs in 32.4% of the standardization sample. This subscore difference is likely not due to chance but could hardly be considered abnormal.

Subscore Reliability

There are a variety of methods available to assess subscore reliability (Brennan, 2005). Usually, this is examined by assessing the variation within each subscore across all examinees (i.e., between-person) via separate internal consistency estimates for each subscore. When examining profiles in order to determine patterns of strengths and weaknesses, however, the variation among the subscores for each individual (i.e., within-person) is more important. Consequently, reliability of subscore profiles should be assessed using a method that includes within-person variability (Conger & Lipshitz, 1973). If the within-person reliability is not sufficiently high, then high-stakes decisions should not be made using clinical profiles.

Because there is a finite amount of variance in a given set of scores, the between-person and within-person variance are not independent. Increasing the between-person variance (which is often a goal with intelligence tests) comes at the price of decreasing the within-person variance (Huang, 2015). This could be why previous research has typically found subscore profiles tend to be unstable across time (Borsuk, Watkins, & Canivez, 2006; Watkins & Canivez, 2004). Despite this instability, psychologists continue to rely on subscore profile analysis for high-stakes clinical decisions (Maki & Adams, 2019; Sotelo-Dynega & Dixon, 2014; Toffalini, Giofrè, & Cornoldi, 2017). Consequently, the purpose of the present study was to examine the reliability of cognitive subscore profiles. Specifically, we sought to address the following research questions: (a) What is the within-person reliability for cognitive subscore profiles using subtest scores, and (b) What is the within-person reliability for cognitive subscore profiles using index scores?

Method

Participants

Participant data were extracted from archival special education records from two large public-school districts located in the Southwestern United States. Participants were included in the present study if their school records contained scores for all core subtests, index scores, and the full-scale IQ (FSIQ) on the Wechsler Intelligence Scales for Children–Fourth Edition (WISC-IV; Wechsler, 2003a) at two time points. Before collecting the data, the study was approved by an university institutional review board and by school district administrators.

Participants were children ($N = 296$) aged 6.1 to 13.9 years old who were administered the WISC-IV two times, an average of 2.84 ($SD = 0.60$) years apart. Demographic information is presented in Table 1. The majority of participants (66.6%) in the study sample met criteria for a specific learning disability (SLD), which is unsurprising given that it is the most common disability observed in school-based settings across the United States (McFarland et al., 2017), and diagnostic decisions for SLD often involve standardized individually administered intelligence tests (Braden & Althanasiou, 2013). Of participants who were diagnosed with SLD, approximately 21.3% ($n = 42$) displayed deficits in reading, 10.7% ($n = 21$) displayed deficits in mathematics, 7.1% ($n = 14$) displayed deficits in writing, and 60.9% ($n = 120$) displayed deficits in multiple academic areas.

Instrument

The WISC-IV is an individually administered intelligence test for children aged 6 to 16 years old. It contains 10 core subtests that comprise four index scores and the FSIQ score. The VCI is derived from the Vocabulary, Similarities, and Comprehension subtests. The Perceptual Reasoning Index (PRI) is derived from the Matrix Reasoning, Block Design, and Picture Concepts subtests. The Processing Speed Index (PSI) is derived from the Coding and Symbol Search subtests. The Working Memory Index (WMI) is derived from the Digit Span and Letter–Number Sequencing subtests.

Internal consistency reliability coefficients for WISC-IV subtest and index scores range between .79 and .90 and .88 and .94, respectively, while test–retest stability coefficients for WISC-IV subtest and index scores range between .68 and .85 and .79 and .89, respectively (Wechsler, 2003b). Long-term stability (i.e., approximately 11 months) for the subtest and index scores range between .22 and .81 and .49 and .75, respectively, for typically developing children (Ryan, Glass, & Bartels, 2010) and between .28 and .70 and .52. and 76 (i.e., approximately 3 years) for children from referred and clinical samples (Lander, 2010; Watkins & Smith, 2013).

Research investigating the structural validity of the WISC-IV has consistently revealed four factors that match the scoring structure of the test in the standardization sample (Watkins, 2006; Wechsler, 2003b) and in referred and clinical samples (Bodin, Pardini, Burns, & Stevens, 2009; Canivez, 2014; Devena, Gay, & Watkins, 2013; Nakano & Watkins, 2013; Styck & Watkins, 2016, 2017; Watkins, 2010). This four-factor structure has also been demonstrated to be invariant across gender (Chen & Zhu, 2008), age (Keith, Fine, Taub, Reynolds, & Kranzler, 2006), clinical and nonclinical samples (Chen & Zhu, 2012), and testing occasions (Richerson, Watkins, & Beaujean, 2014). Nonetheless, the proportion of reliable variance in WISC-IV subtest scores due to variance in the four index scores tends to be low: ranging between .26 and .48 for the VCI, ranging between .02 and .17 for the PRI, ranging between .33 and .53 for the PSI, and ranging between .10 and .23 for the WMI (Canivez, 2014; Gomez, Vance, & Watson, 2016, 2017; Styck & Watkins, 2016, 2017).

Table 1
Demographic Information for Study Participants (N = 296)

Variable	%	n
Characteristic		
Female	32.1	95
White	78.7	233
Hispanic	11.1	33
Black	7.1	21
Asian/Pacific Islander	1.7	5
Missing	.3	3
Primary diagnosis		
Specific learning disability	66.6	197
Other health impairment	11.1	33
Emotional disturbance	7.4	22
None	6.4	19
Autism	3.4	10
Speech/language impairment	2.4	7
Intellectual disability	2.0	6
Multiple disabilities	.3	1
Hearing impairment	.3	1
Secondary diagnosis		
None	76.7	227
Speech/language impairment	10.8	32
Missing	4.4	13
Specific learning disability	3.7	11
Other health impairment	2.4	7
Emotional disturbance	1.7	5
Hearing impairment	.3	1

Analyses

Bulut and colleagues (2013; Bulut et al., 2017) proposed a method of estimating reliability of clinical profiles that divides total subscore variability obtained from parallel test forms into within-person and between-person variability. It produces between-person (ρ_B) and within-person (ρ_W) reliability estimates.

ρ_B is the proportion of variance in the observed profile levels that can be attributed to variance in true profile levels across test score profiles. It is calculated as

$$\rho_B = \frac{\tau_B}{\sigma_B^2}. \quad (1)$$

The true between-person variability (τ_B) is calculated as

$$\tau_B = D\sigma_{\bar{T}_j}^2, \quad (2)$$

where D is the total number of subscores and $\sigma_{\bar{T}_j}^2$ is the variance in true profile levels (i.e., variance of average true subscore values between persons). The between-person observed score variance (σ_B^2) is calculated as

$$\sigma_B^2 = D\sigma_{\bar{X}_j}^2, \quad (3)$$

where $\sigma_{\bar{X}_j}^2$ is the variance in observed profile levels (i.e., variance of average observed subscore values between persons).

ρ_W is the proportion of variance in the observed profile patterns due to variance in true profile patterns across test score profiles. It is calculated as

$$\rho_W = \frac{\tau_W}{\sigma_W^2}. \quad (4)$$

The true within-person variability (τ_W) is calculated as

$$\tau_W = \sum_{d=1}^{D}\left(\frac{\sum_{j=1}^{J}[T_{jd} - \bar{T}_j]^2}{J}\right), \quad (5)$$

where T_{jd} is the true subscore value for person j on subscore d, $\bar{T}_j$ is the average true score for person j across all d subscores, and J is the total number of people in the sample. The observed within-person variance (σ_W^2) is calculated as

$$\sigma_W^2 = \sum_{d=1}^{D}\left(\frac{\sum_{j=1}^{J}[X_{jd} - \bar{X}_j]^2}{J}\right), \quad (6)$$

where X is an observed score and the rest of the terms are defined the same as in Equation 5.

Total profile reliability (ρ_T) is the weighted average of ρ_B and ρ_W. It is calculated as

$$\rho_T = \frac{1}{\sigma_B^2 + \sigma_W^2}\left(\sigma_B^2\rho_B + \sigma_W^2\rho_W\right), \quad (7)$$

where terms are the same those defined in Equation 1 through Equation 6. The value of ρ_T is constrained to be between ρ_B and ρ_W.

As ρ_W gets closer to one, the distinct information provided by subscores becomes more precise. Thus, if ρ_W is large, then test interpretation needs to account for subscore patterns. ρ_B is a measure of reliability for the total test score, so values closer to one indicate there will be more precision with the actual test scores. If ρ_B is large, then test interpretation should focus on the individual scores. If both ρ_W and ρ_B are large, then score interpretation may be able to include both subscore values and their patterns.

Bulut et al. (2017) found that for a fixed subtest length, as the correlations among subscores increased, ρ_B increased at the expense of ρ_W decreasing. This led them to conclude that "the test conditions that lead to subscores with high between-person reliability . . . significantly reduce the distinctiveness of the subscores," which then result "in subscores with unacceptably low within-person reliability" (p. 102). In other words, there is a trade-off between maximizing ρ_B and ρ_W (Huang, 2015).

Evidence to support the reliability of subscore profiles for the WISC-IV would be indicated if ρ_W is close to one. If ρ_W is close to zero, this would indicate subscore profiles do not have sufficient reliability for individual interpretation. Of course, reliability is necessary evidence for interpretation, but not sufficient. So, high reliability values should be supplemented with validity and utility evidence before using subscore profiles to make high stakes decisions.

The sample included in the present study is theoretically equivalent to administering two essentially τ-equivalent forms of a test, which is a requirement for Bulut's (2013; Bulut et al., 2017) profile reliability method. Essentially τ-equivalent forms exist when the true score variance is equal across forms (Graham, 2006). Same-time and cross-time correlations between pairs of subscores should, therefore, be highly similar. To examine this assumption, we calculated the same-time and cross-time correlations for all the WISC-IV scores in the study sample.

To examine consistency across administrations, we transformed mean differences between scores at time one and time two using Hedges' (1981) standardized effect size measure (g) as well as its confidence interval. Subsequently, we estimated the between-person and within-person profile reliability from participants' WISC-IV subtests and the four index scores. Some researchers have asserted that students' diagnosed with SLD possesses specific patterns of cognitive and academic patterns of strengths and weaknesses which can be used for diagnostic and intervention decisions (Flanagan et al., 2013; Hale et al., 2010). Consequently, we examined profile reliability separately for the subsample of respondents with a primary classification of SLD. All analyses were conducted in R (Version 3.4.1; R Core Team, 2017). Profile reliability was estimated using the *profileR* package (Version 0.3–4; Bulut & Desjardins, 2017).

Results

Summary statistics, Time 1 to Time 2 score correlations (i.e., test–retest reliabilities), and between-administration effects sizes are presented in Table 2. The Block Design and Coding subtests were both somewhat lower at Time 2 than Time 1, but overall there were minimal mean differences between scores across time.

Tables 3 and 4 contain same-time (see Table 3) and cross-time correlations (see Table 4) for all the WISC-IV scores for the study

Table 2
Descriptive Statistics for Wechsler Intelligence Scales for Children–Fourth Edition Scores at Time 1 and Time 2 (N = 296)

	Time 1		Time 2			Effect size		
Variable	*M*	*SD*	*M*	*SD*	$r_{T1,T2}$	*g*	95% CI	
Subtest scores								
Block design	9.25	2.74	8.81	2.88	.69	.20	.04	.36
Similarities	8.81	2.58	9.15	2.69	.56	−.13	−.30	.03
Digit span	8.07	2.51	7.83	2.53	.58	.10	−.06	.26
Picture concepts	9.63	3.25	10.15	2.89	.43	−.16	−.32	.00
Coding	8.61	3.13	7.66	2.83	.50	.32	.16	.48
Vocabulary	8.65	2.47	8.43	2.65	.65	.11	−.06	.27
Letter–number sequencing	8.12	2.74	8.18	3.01	.45	−.02	−.18	.14
Matrix reasoning	9.11	2.84	9.27	2.94	.61	−.06	−.22	.10
Comprehension	9.00	2.54	8.97	2.48	.44	−.00	−.16	.16
Symbol search	8.53	3.19	8.75	3.03	.51	−.07	−.23	.09
Index scores								
Verbal Comprehension Index	92.94	11.74	93.09	12.46	.70	−.02	−.18	.14
Perceptual Reasoning Index	95.91	14.26	96.44	14.68	.73	−.05	−.21	.11
Working Memory Index	88.78	12.41	88.08	13.67	.63	.06	−.10	.22
Processing Speed Index	92.08	14.83	89.94	14.64	.62	.17	.00	.33
Full-scale score								
Full-scale IQ score	90.87	12.67	90.48	13.25	.79	.05	−.12	.21

Note. Subtest scores are standardized to have a mean of 10 and a standard deviation of 3; and, index scores and the full-scale IQ score are standardized to have a mean of 100 and a standard deviation of 15.g = Hedge's g effect size; $r_{T1,T2}$ = correlations between scores at Times 1 and 2; CI = confidence interval.

Table 3
Same-Time Correlations of WISC-IV Scores at Time 1 and Time 2

Subtest	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. BD	**.69**	.22	.44	.37	.22	.26	.37	.51	.17	.34	.26	.77	.47	.33	.63
2. SI	.39	**.56**	.30	.36	.02	.60	.30	.32	.44	.33	.81	.39	.35	.21	.60
3. DS	.40	.38	**.58**	.33	.14	.36	.44	.36	.18	.32	.34	.48	.83	.28	.61
4. PCn	.40	.41	.28	**.43**	.18	.38	.28	.42	.36	.35	.44	.78	.36	.32	.67
5. CD	.23	.04	.24	.18	**.50**	.14	.29	.13	.12	.40	.12	.23	.26	.84	.45
6. VC	.32	.70	.38	.38	.09	**.65**	.38	.33	.60	.28	.87	.42	.44	.26	.68
7. LN	.40	.35	.51	.29	.24	.46	**.45**	.41	.30	.30	.39	.44	.86	.35	.65
8. MR	.62	.45	.44	.50	.21	.40	.39	**.61**	.28	.34	.37	.80	.46	.28	.66
9. CO	.32	.48	.37	.36	.31	.60	.43	.34	**.44**	.25	.82	.35	.29	.23	.57
10. SS	.35	.22	.27	.25	.60	.22	.36	.36	.32	**.51**	.34	.44	.37	.83	.65
11. VCI	.41	.85	.45	.45	.17	.90	.49	.47	.80	.30	**.70**	.46	.43	.28	.74
12. PRI	.82	.51	.45	.77	.25	.45	.44	.87	.41	.39	.54	**.73**	.54	.40	.83
13. WMI	.46	.42	.84	.33	.28	.49	.89	.48	.46	.37	.54	.52	**.63**	.38	.75
14. PSI	.33	.15	.28	.24	.89	.18	.34	.31	.35	.90	.27	.36	.36	**.62**	.66
15. FSIQ	.69	.67	.64	.62	.50	.69	.68	.73	.67	.61	.80	.83	.76	.62	**.79**

Note. Time 1 correlations are depicted in the upper triangle, Time 2 correlations are depicted in the lower triangle, and test–retest correlations are depicted on the diagonal in boldface type. WISC-IV = Wechsler Intelligence Scales for Children–Fourth Edition; BD = block design; SI = similarities; DS = digit span; PCn = picture concepts; CD = coding; VC = vocabulary; LN = letter–number sequencing; MR = matrix reasoning; CO = comprehension; SS = symbol search; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index; FSIQ = full-scale IQ score.

sample. Absolute deviations between same-time and cross-time correlations for all WISC-IV subtest score pairs ranged between .00 and .41 (*Mdn* = .06); for the index scores, the absolute deviations ranged between .00 to .21 (*Mdn* = .08). Consequently, differences between same-time and cross-time correlations appear to be negligible, which lends support to the assumption of essential τ-equivalence.

Profile reliability estimates for the subtest and index scores for the entire sample are provided in the top part of Table 5, while estimates for the SLD subgroup are provided in the bottom part of Table 5. The results were relatively similar across subscores and groups as is evident by the mean subtest and index scores depicted in Figures 1 and 2. Specifically, the between-person reliability estimates were higher than the within-person estimates, with the difference in values being larger for the subtests than the index scores. Using the Shrout and Lane's (2012) reliability classifications, the between-person reliability estimates are all moderate, whereas the within-person reliability estimates are slight-to-fair. Within-person reliability coefficients of this magnitude do not provide sufficient evidence for interpreting within-person cognitive ability strengths and weaknesses. Of note, the between-person reliability estimates are close to the correlation between FSIQ scores at time one and time two. Because the FSIQ scores have the highest reliability of any score on the WISC-IV (Wechsler, 2003b), likely the between-score reliability estimates are maximized for this particular sample.

Discussion

The purpose of the present study was to examine profile reliability of intelligence test subscores. Using Bulut et al.'s (2017)

Table 4
Cross-Time Correlations of WISC-IV Scores at Time 1 and Time 2

	Time 2 subtest														
Time 1 subtest	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. BD	**.69**	.35	.34	.39	.20	.32	.35	.56	.29	.35	.38	.67	.40	.31	.60
2. SI	.21	**.56**	.29	.28	.02	.50	.30	.28	.34	.17	.54	.32	.34	.12	.46
3. DS	.34	.32	**.58**	.25	.11	.33	.45	.36	.21	.22	.35	.39	.59	.18	.48
4. PCn	.34	.40	.31	**.43**	.10	.34	.34	.40	.35	.24	.43	.47	.38	.19	.50
5. CD	.21	.03	.17	.25	**.50**	.06	.14	.22	.17	.45	.11	.28	.18	.52	.34
6. VC	.25	.58	.37	.31	.08	**.65**	.36	.30	.50	.20	.68	.35	.41	.16	.55
7. LN	.32	.29	.38	.20	.27	.34	**.45**	.37	.23	.33	.34	.36	.48	.34	.50
8. MR	.51	.35	.32	.39	.17	.33	.34	**.61**	.29	.29	.38	.61	.38	.26	.56
9. CO	.20	.45	.29	.27	.16	.47	.35	.24	**.44**	.22	.53	.29	.37	.21	.46
10. SS	.35	.22	.28	.23	.40	.19	.32	.32	.24	**.51**	.25	.37	.35	.51	.48
11. VCI	.27	.63	.38	.34	.11	.64	.41	.33	.51	.24	**.70**	.38	.45	.20	.59
12. PRI	.64	.47	.41	.51	.20	.42	.44	.65	.40	.37	.50	**.73**	.49	.32	.70
13. WMI	.38	.35	.56	.27	.23	.40	.53	.43	.26	.33	.40	.44	**.63**	.31	.58
14. PSI	.33	.15	.27	.29	.54	.15	.28	.31	.25	.57	.22	.38	.31	**.62**	.49
15. FSIQ	.56	.55	.52	.49	.34	.55	.54	.59	.48	.49	.62	.66	.61	.47	**.79**

Note. Cross-time correlations between Time 1 (rows) and Time 2 (columns) are depicted in the upper and lower triangles. Test–retest correlations are depicted on the diagonal in boldface type. WISC-IV = Wechsler Intelligence Scales for Children–Fourth Edition; BD = block design; SI = similarities; DS = digit span; PCn = picture concepts; CD = coding; VC = vocabulary; LN = letter–number sequencing; MR = matrix reasoning; CO = comprehension; SS = symbol search; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index; FSIQ = full-scale IQ score.

Table 5
WISC-IV Subtest and Index Subscore Profile Reliability Estimates

Subscore	Profile reliability estimate		
	ρ_B	ρ_W	ρ_T
Entire sample (N = 296)			
Subtest	.79	.37	.54
Index	.78	.53	.67
SLD subsample (n = 197)			
Subtest	.77	.36	.51
Index	.76	.52	.64

Note. WISC-IV = Wechsler Intelligence Scales for Children–Fourth Edition; ρ_B = between-person reliability; ρ_W = within-person reliability; ρ_T = total reliability; SLD = specific learning disability.

method for examining profile stability, we estimated the between-person (ρ_B) and within-person (ρ_W) reliability for the WISC-IV subtests and index scores using a sample of students twice referred for special education services. Results indicated that ρ_B was substantially higher than ρ_W for both subtest and index score profiles. Moreover, this pattern of results remained when examining profiles for a subsample of students diagnosed with SLD. Bulut et al. stated that "subscores with high between-person reliability and low within-person reliability would indicate that subscores may not provide any valuable information about examinees beyond what the total test score already provides" (p. 92). Consequently, subscore profiles using subtest and index scores from the WISC-IV do not appear to provide reliable information.

These results contribute to a growing empirical literature base that strongly suggests subscore profiles from intelligence tests do not yield clinically meaningful information (McGill et al., 2018; McGill, Styck, Palomares, & Hass, 2016; Watkins & Glutting, 2000; Watkins, Glutting, & Youngstrom, 2005). This makes one wonder why test publishers and clinicians continue to recommend that these patterns of subscores contain clinically useful information and should be interpreted (e.g., Flanagan et al., 2013; Hale et al., 2010; Wechsler, 2003b, 2014a). As per Grice et al. (2017), this may be an example of inappropriately attempting to explain individual-level phenomenon with group-level data.

It is unknown exactly why empirical studies that examine subscore patterns from intelligence tests tend to demonstrate poor

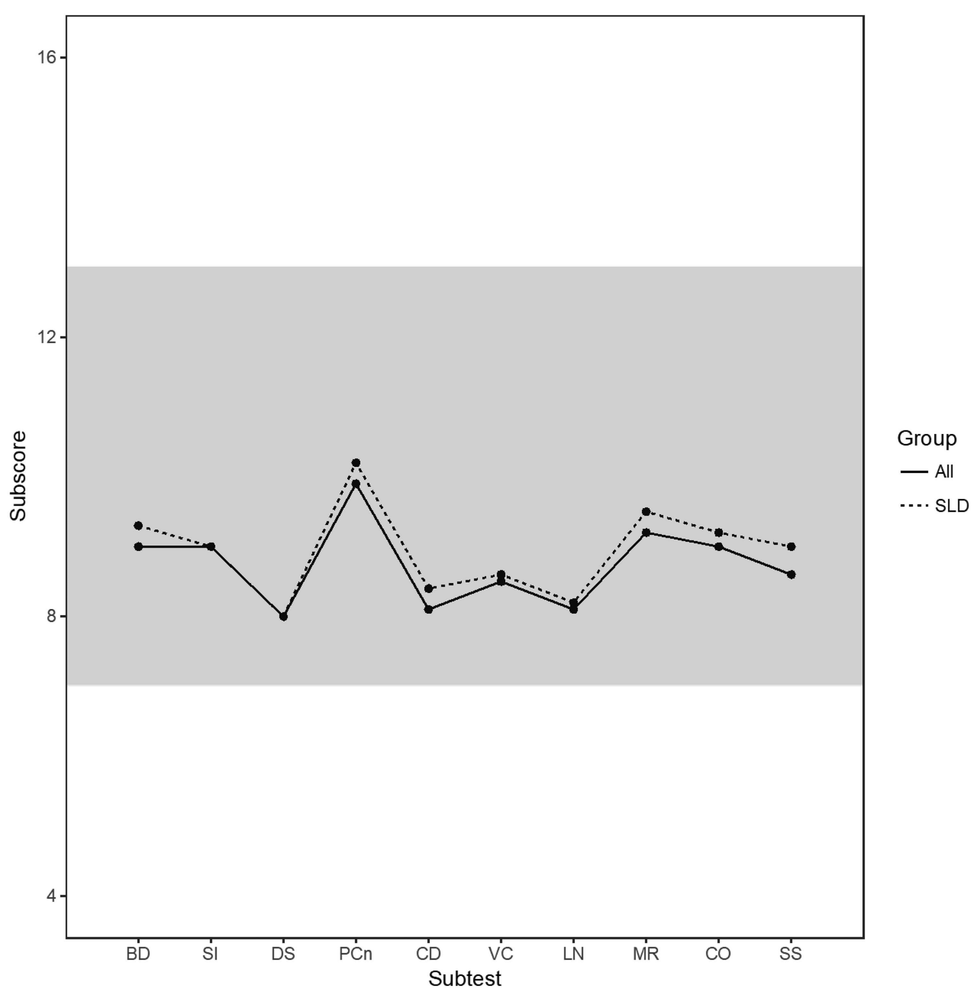

Figure 1. Mean Wechsler Intelligence Scales for Children–Fourth subtest scores for a sample of 296 children referred for special education evaluations and a subsample of 197 children identified with specific learning disabilities by multidisciplinary evaluation teams. BD = block design; SI = similarities; DS = digit span; PCn = picture concepts; CD = coding; VC = vocabulary; LN = letter-number sequencing; MR = matrix reasoning; CP = comprehension; SS = symbol search; SLD = specific learning disabilities.

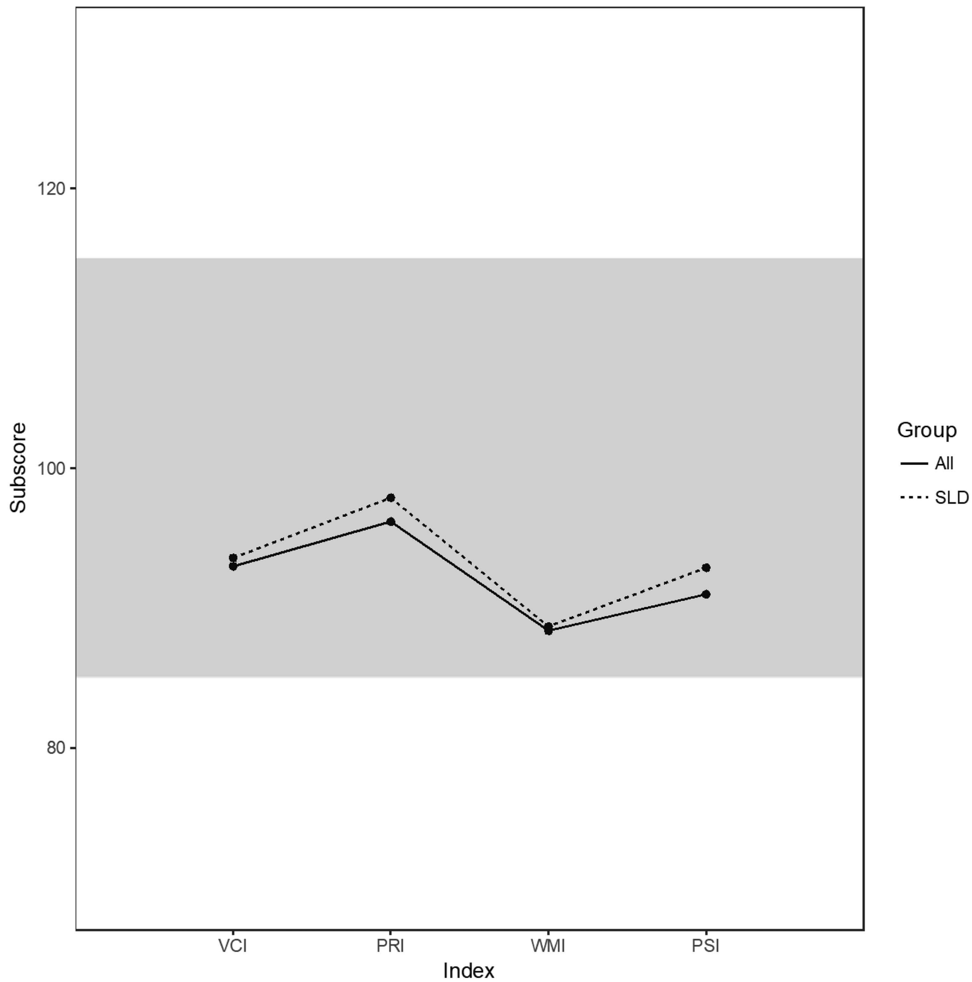

Figure 2. Mean Wechsler Intelligence Scales for Children–Fourth Edition index scores for a sample of 296 children referred for special education evaluations and a subsample of 197 children identified with specific learning disabilities by multidisciplinary evaluation teams. VCI = verbal comprehension index; PRI = perceptual reasoning index; WMI = working memory index; PSI = processing speed index; SLD = specific learning disabilities.

psychometric properties—both in the present study as well as in many other studies (e.g., McGill & Busse, 2015; Smith & Watkins, 2004; Watkins, Kush, & Schaefer, 2002). Likely, there are at least two contributing factors. First, cognitive ability tests are implicitly designed to maximize between-person reliability, which functionally precludes having high within-person reliability.

Subscores have added value beyond a total test score when they have high reliability, are distinct from other subscores (i.e., low subscore correlations), and the total test score has low reliability (Haberman, 2008; Sinharay, 2010; Sinharay, Haberman, & Puhan, 2007). Most intelligence tests are designed to maximize efficiency, avoid examinee fatigue, and produce a total test score (e.g., FSIQ in the WISC-IV) with very high reliability. Moreover, the between-person correlations among subscores on intelligence tests tends to be relatively high (Bodin et al., 2009; Canivez, 2014; Devena et al., 2013; Nakano & Watkins, 2013; Styck & Watkins, 2016, 2017; Watkins, 2010). Consequently, some of the very properties cognitive test developers seek to maintain in order to support the structural validity of their between-person score interpretations are likely contributing to the poor properties of within-person subtest and index subscore profiles.

Another likely contributing factor is that the scores from these cognitive ability tests, at best, have an ordinal measurement structure (Michell, 2012). Thus, from a measurement perspective, interpretations that require direct score comparisons (e.g., $a-b > c-d$) are questionable because they assume homogeneity of magnitude differences that have not been empirically demonstrated. Moreover, in their desire to make intelligence tests as commercially appealing as possible, test publishers and authors provide a bevy of scores appealing to a wide range of clinicians. A result of this is that many of these scores are neither theoretically nor psychometrically defensible (Beaujean & Benson, 2018,).

The present study does contain limitations. Children in our subsample of participants diagnosed with SLD exhibited unexpected underachievement in a variety of academic areas. Sample sizes precluded the estimation of SLD within specific academic areas (i.e., reading, writing, mathematics) and there is some evidence to suggest that WISC-IV subtest and index scores may not measure intelligence in the same way for children with and without SLD (Giofrè & Cornoldi, 2015). Future studies may wish to apply the procedures outlined in Bulut et al. (2017) to estimate the subscore profile reliability in more homogeneous samples of students with SLD in specific academic areas or in other homogenous clinical groups. In addition, the sample contained in the present study is a referred sample. Referred samples have been demonstrated to have lower mean scores that are less variable when compared with test standardization samples (e.g., Bodin et al., 2009; Canivez, 2014; Styck & Watkins, 2016, 2017) and results may only generalize to other referred samples.

Despite these limitations, the present study adds to the body of research questioning the practice of cognitive ability clinical profile interpretation. Clinicians are advised to eschew interpretation of subscore profiles until evidence indicates that they are reliable and contain information that is not available from the total test score (Bulut et al., 2017; Sinharay et al., 2011; Wainer & Feinberg, 2015).

References

Beaujean, A. A., & Benson, N. F. (in press). The one and the many: Enduring legacies of Spearman and Thurstone on intelligence test score interpretation. *Applied Measurement in Education*.

Beaujean, A. A., & Benson, N. F. (2018). Theoretically-consistent cognitive ability test development and score interpretation. *Contemporary School Psychology*. Advance online publication. http://dx.doi.org/10.1007/s40688-018-0182-1

Beeldman, E., Raaphorst, J., Klein Twennaar, M., de Visser, M., Schmand, B. A., & de Haan, R. J. (2016). The cognitive profile of ALS: A systematic review and meta-analysis update. *Journal of Neurology, Neurosurgery, and Psychiatry, 87,* 611–619. http://dx.doi.org/10.1136/jnnp-2015-310734

Bodin, D., Pardini, D. A., Burns, T. G., & Stevens, A. B. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15,* 417–424. http://dx.doi.org/10.1080/09297040802603661

Borsuk, E. R., Watkins, M. W., & Canivez, G. L. (2006). Long-term stability of membership in a Wechsler Intelligence Scale for Children-Third Edition (WISC-III) subtest core profile taxonomy. *Journal of Psychoeducational Assessment, 24,* 52–68. http://dx.doi.org/10.1177/0734282905285225

Braden, J. P., & Althanasiou, M. S. (2013). Psychological assessment in school settings. In J. R. Graham & J. A. Naglieri (Eds.), *Assessment psychology: Handbook of psychology* (2nd ed., pp. 291–314). Hoboken, NJ: Wiley.

Braden, J. P., & Kratochwill, T. R. (1997). Treatment utility of assessment: Myths and realities. *School Psychology Review, 26,* 475–485.

Brennan, R. L. (2005). *Some test theory for the reliability of individual profiles* (Research Report 12). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment.

Bulut, O. (2013). *Between-person and within-person subscore reliability: Comparison of unidimensional and multidimensional IRT models* (Unpublished doctoral dissertation). Department of Educational Psychology, University of Minnesota, Twin Cities, MN.

Bulut, O., Davison, M. L., & Rodriguez, M. C. (2017). Estimating between-person and within-person subscore reliability with profile analysis. *Multivariate Behavioral Research, 52,* 86–104. http://dx.doi.org/10.1080/00273171.2016.1253452

Bulut, O., & Desjardins, C. D. (2017). profileR: Profile analysis of multivariate data in R (R package Version 0.3–4) [Computer software]. Retrieved from https://rdrr.io/cran/profileR/

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 84–112). New York, NY: Oxford University Press.

Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly, 29,* 38–51. http://dx.doi.org/10.1037/spq0000032

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children-Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29,* 458–472. http://dx.doi.org/10.1037/pas0000358

Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, interactive. *Psychological Review, 51,* 292–303. http://dx.doi.org/10.1037/h0057299

Chen, H., & Zhu, J. (2008). Factor invariance between genders of the Wechsler Intelligence Scale for Children-Fourth Edition. *Personality and Individual Differences, 45,* 260–266. http://dx.doi.org/10.1016/j.paid.2008.04.008

Chen, H., & Zhu, J. (2012). Measurement invariance of WISC-IV across normative and clinical samples. *Personality and Individual Differences, 52,* 161–166. http://dx.doi.org/10.1016/j.paid.2011.10.006

Conger, A. J., & Lipshitz, R. (1973). Measures of reliability for profiles and test batteries. *Psychometrika, 38,* 411–427. http://dx.doi.org/10.1007/BF02291663

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50,* 456–473. http://dx.doi.org/10.1037/h0057173

Devena, S. E., Gay, C. E., & Watkins, M. W. (2013). Confirmatory factor analysis of the WISC-IV in a hospital referral sample. *Journal of Psychoeducational Assessment, 31,* 591–599. http://dx.doi.org/10.1177/0734282913483981

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). Hierarchical exploratory factor analyses of the Woodcock-Johnson IV full test battery: Implications for CHC application in school psychology. *School Psychology Quarterly, 33,* 235–250. http://dx.doi.org/10.1037/spq0000221

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley.

Giofrè, D., & Cornoldi, C. (2015). The structure of intelligence in children with specific learning disabilities is different as compared to typically development children. *Intelligence, 52,* 36–43. http://dx.doi.org/10.1016/j.intell.2015.07.002

Gomez, R., Vance, A., & Watson, S. D. (2016). Structure of the Wechsler Intelligence Scale for Children-Fourth Edition in a Group of Children with ADHD. *Frontiers in Psychology, 737*. http://dx.doi.org/10.3389/fpsyg.2016.00737

Gomez, R., Vance, A., & Watson, S. (2017). Bifactor model of WISC-IV: Applicability and measurement invariance in low and normal IQ groups. *Psychological Assessment, 29,* 902–912. http://dx.doi.org/10.1037/pas0000369

Gottfredson, G. D., & Jones, E. M. (1993). Psychological meaning of profile elevation in the Vocational Preference Inventory. *Journal of Career Assessment, 1,* 35–49. http://dx.doi.org/10.1177/106907279300100105

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement, 66,* 930–944. http://dx.doi.org/10.1177/0013164406288165

Grice, J., Barrett, P., Cota, L., Felix, C., Taylor, Z., Garner, S., . . . Vest, A. (2017). Four bad habits of modern psychologists. *Behavioral Sciences, 7,* 1–21.

Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics, 33,* 204–229. http://dx.doi.org/10.3102/1076998607302636

Hale, J. B., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., & Yalof, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disabilities Quarterly, 33,* 223–236. http://dx.doi.org/10.1177/073194871003300310

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128. http://dx.doi.org/10.3102/10769986006002107

Huang, L. (2015). *Improving the use of subscores on a test battery: Some reliability and validity evidence from the Wechsler Intelligence Scale for Children-Fourth Edition* (Unpublished doctoral dissertation). Department of Educational Psychology, University of Minnesota, Twin Cities, MN.

Jones, L. K. (1989). Measuring a three-dimensional construct of career indecision among college students: A revision of the Vocational Decision Scale—The Career Decision Profile. *Journal of Counseling Psychology, 36,* 477–486.

Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A history of intelligence test interpretation. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (3rd ed., pp. 56–70). New York, NY: Guilford Press.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review, 35,* 108–127.

Lander, J. (2010). *Long-term stability of scores on the Wechsler Intelligence Scale for Children-Fourth Edition in children with learning disabilities* (Doctoral dissertation). Retrieved from ProQuest. (Accession No. 3407030)

Letteri, C. A. (1980). Cognitive profile: Basic determinant of academic achievement. *The Journal of Educational Research, 73,* 195–199. http://dx.doi.org/10.1080/00220671.1980.10885234

Luecht, R. M., Gierl, M. J., Tan, X., & Huff, K. (2006, April 8–10). *Scalability and the development of useful diagnostic scales*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Lykken, D. T. (1956). A method of actuarial pattern analysis. *Psychological Bulletin, 53,* 102–107. http://dx.doi.org/10.1037/h0043709

Maki, K. E., & Adams, S. R. (2019). A current landscape of specific learning disability identification: Training, practices, and implications. *Psychology in the Schools, 56,* 18–31. http://dx.doi.org/10.1002/pits.22179

Maurer, T. J., & Tarulli, B. A. (1997). Managerial work, job analysis, and Holland's RIASEC vocational environment dimensions. *Journal of Vocational Behavior, 50,* 365–381. http://dx.doi.org/10.1006/jvbe.1996.1549

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. http://dx.doi.org/10.1177/002246699202500407

McFarland, J., Hussar, B., de Brey, C., Synder, T., Want, X., Wilkinson-Flicker, S., . . . Hinz, S. (2017). *The condition of education 2017 (NCES 2017–144)*. Washington, DC: National Center for Education Statistics.

McGill, R. J., & Busse, R. T. (2015). Incremental validity of the WJ-III COG: Limited predictive effects beyond the GIA-E. *School Psychology Quarterly, 30,* 353–365. http://dx.doi.org/10.1037/spq0000094

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology, 71,* 108–121. http://dx.doi.org/10.1016/j.jsp.2018.10.007

McGill, R. J., Styck, K. M., Palomares, R. S., & Hass, M. R. (2016). Critical issues in specific learning disability identification: What we need to know about the PSW model. *Learning Disability Quarterly, 39,* 159–170. http://dx.doi.org/10.1177/0731948715618504

Meehl, P. E. (1946). Profile analysis of the Minnesota Multiphasic Personality Inventory in differential diagnosis. *Journal of Applied Psychology, 30,* 517–524. http://dx.doi.org/10.1037/h0062318

Michell, J. (2012). "The constantly recurring argument:" Inferring quantity from order. *Theory & Psychology, 22,* 255–271. http://dx.doi.org/10.1177/0959354311434656

Nakano, S., & Watkins, M. W. (2013). Factor structure of the Wechsler Intelligence Scales for Children-Fourth Edition among referred Native American students. *Psychology in the Schools, 50,* 957–968. http://dx.doi.org/10.1002/pits.21724

Ortiz, S. O. (2015). CHC theory of intelligence. In S. Goldstein, D. Princiotta, & J. Naglieri (Eds.), *Handbook of intelligence* (pp. 209–227). New York, NY: Springer.

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly, 15,* 376–385. http://dx.doi.org/10.1037/h0088795

Raaphorst, J., de Visser, M., Linssen, W. H. J. P., de Haan, R. J., & Schmand, B. (2010). The cognitive profile of amyotrophic lateral sclerosis: A meta-analysis. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases, 11,* 27–37. http://dx.doi.org/10.3109/17482960802645008

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Richerson, L. P., Watkins, M. W., & Beaujean, A. A. (2014). Longitudinal invariance of the Wechsler Intelligence Scale for Children-Fourth Edition in a referred sample. *Journal of Psychoeducational Assessment, 32,* 597–609. http://dx.doi.org/10.1177/0734282914538802

Rizza, M. G., McIntosh, D. E., & McCunn, A. (2001). Profile analysis of the Woodcock-Johnson III Tests of Cognitive Abilities with gifted students. *Psychology in the Schools, 38,* 447–455. http://dx.doi.org/10.1002/pits.1033

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Torrance, CA: WPS.

Ryan, J. J., Glass, L. A., & Bartels, J. M. (2010). Stability of the WISC-IV in a sample of elementary and middle school children. *Applied Neuropsychology, 17,* 68–72. http://dx.doi.org/10.1080/09084280903297933

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Shrout, P. E., & Lane, S. P. (2012). Reliability. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics* (pp. 643–660). Washington, DC: American Psychological Association.

Silverstein, A. B. (1981). Reliability and abnormality of test score differences. *Journal of Clinical Psychology, 37,* 392–394. http://dx.doi.org/10.1002/1097-4679(198104)37:2<392::AID-JCLP2270370230>3.0.CO;2-8

Sinharay, S. (2010). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement, 47,* 150–174. http://dx.doi.org/10.1111/j.1745-3984.2010.00106.x

Sinharay, S., Haberman, S. J., & Puhan, G. (2007). Subscores based on classical test theory: To report or not to report. *Educational Measurement: Issues and Practice, 26,* 21–28. http://dx.doi.org/10.1111/j.1745-3992.2007.00105.x

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice, 30,* 29–40. http://dx.doi.org/10.1111/j.1745-3992.2011.00208.x

Smith, C. B., & Watkins, M. W. (2004). Diagnostic utility of the Bannatyne WISC-III pattern. *Learning Disabilities Research & Practice, 19,* 49–56. http://dx.doi.org/10.1111/j.1540-5826.2004.00089.x

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 51,* 1031–1045. http://dx.doi.org/10.1002/pits.21802

Styck, K. M., Beaujean, A. A., & Watkins, M. W. (2019). *Profile reliability of cognitive ability subscores in a referred sample* [Data set]. Ann Arbor, MI: InterUniversity Consortium for Political and Social Research. http://dx.doi.org/10.3886/ICPSR37285.v1

Styck, K. M., & Watkins, M. W. (2016). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities, 49,* 216–224. http://dx.doi.org/10.1177/0022219414539565

Styck, K. M., & Watkins, M. W. (2017). Structure of the WISC-IV for students with ADHD. *Journal of Attention Disorders, 21,* 921–928. http://dx.doi.org/10.1177/1087054714553052

Sugarman, A., & Kanner, K. (2000). The contribution of psychoanalytic theory to psychological testing. *Psychoanalytic Psychology, 17,* 3–23. http://dx.doi.org/10.1037/0736-9735.17.1.3

Toffalini, E., Giofrè, D., & Cornoldi, C. (2017). Strengths and weaknesses in the intellectual profile of different subtypes of specific learning disorder: A study on 1,049 diagnosed children. *Clinical Psychological Science, 5,* 402–409. http://dx.doi.org/10.1177/2167702616672038

Voglmaier, M. M., Seidman, L. J., Niznikiewicz, M. A., Dickey, C. C., Shenton, M. E., & McCarley, R. W. (2005). A comparative profile analysis of neuropsychological function in men and women with schizotypal personality disorder. *Schizophrenia Research, 74,* 43–49. http://dx.doi.org/10.1016/j.schres.2004.09.013

Voglmaier, M. M., Seidman, L. J., Salisbury, D., & McCarley, R. W. (1997). Neuropsychological dysfunction in schizotypal personality disorder: A profile analysis. *Society of Biological Psychiatry, 41,* 530–540. http://dx.doi.org/10.1016/S0006-3223(96)00056-X

Wainer, H., & Feinberg, R. (2015). For want of a nail: Why unnecessarily long tests may be impeding the progress of Western civilization. *Significance, 12,* 16–21. http://dx.doi.org/10.1111/j.1740-9713.2015.00797.x

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15,* 465–479. http://dx.doi.org/10.1037/h0088802

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children-Fourth Edition. *Psychological Assessment, 18,* 123–125.

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children-Fourth Edition among a national sample of referred students. *Psychological Assessment, 22,* 782–787. http://dx.doi.org/10.1037/a0020043

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition. *School Psychology Quarterly, 29,* 52–63. http://dx.doi.org/10.1037/spq0000038

Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. *Psychological Assessment, 16,* 133–138. http://dx.doi.org/10.1037/1040-3590.16.2.133

Watkins, M. W., & Glutting, J. J. (2000). Incremental validity of WISC-III profile elevation, scatter, and shape information for predicting reading and math achievement. *Psychological Assessment, 12,* 402–408. http://dx.doi.org/10.1037/1040-3590.12.4.402

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues*— (2nd ed., pp. 251–268). New York, NY: Guilford Press.

Watkins, M. W., Kush, J. C., & Schaefer, B. A. (2002). Diagnostic utility of the Learning Disability Index. *Journal of Learning Disabilities, 35,* 98–103, 136. http://dx.doi.org/10.1177/002221940203500201

Watkins, M. W., & Smith, L. G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children-Fourth Edition. *Psychological Assessment, 25,* 477–483. http://dx.doi.org/10.1037/a0031653

Wechsler, D. (1946). *Wechsler-Bellevue Intelligence Scale*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Psychological Corporation.

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

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition technical and interpretive manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition technical and interpretive manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2014a). *Wechsler Intelligence Scale for Children-Fifth Edition technical and interpretive manual*. San Antonio, TX: Pearson.

Wechsler, D. (2014b). *Wechsler Intelligence Scale for Children-Fifth Edition technical and interpretive manual supplement: Special group validity studies with other measures and additional tables*. San Antonio, TX: Pearson.

Received December 27, 2017
Revision received January 2, 2019
Accepted January 7, 2019 ■