



## Are There Cognitive Profiles Unique to Students With Learning Disabilities? A Latent Profile Analysis of Wechsler Intelligence Scale for Children–Fourth Edition Scores

Marley W. Watkins<sup>a</sup>  and Gary L. Canivez<sup>b</sup> 

<sup>a</sup>Baylor University; <sup>b</sup>Eastern Illinois University

### ABSTRACT

It is often assumed that children with learning disabilities (LD) exhibit unique profiles of ability scores that reflect idiosyncratic cognitive strengths and weaknesses. Interpretation of cognitive ability profiles initially focused on visual inspection of subtest scores followed by statistical comparisons to identify significant cognitive strengths and weaknesses. However, subsequent research demonstrated that these subtest profiles lacked sufficient reliability, validity, and diagnostic utility. Profile research typically utilized variable-centered methods, but person-centered methods might be more appropriate. The present study utilized latent profile analysis (LPA), a person-centered method that is model-based and flexible, with 1,830 school-identified students with LD and 2,200 simulated normative participants. Four broad ability score profiles distinguished by level rather than shape emerged. Thus, this latent mixture model analysis found no mixture of subpopulations, suggesting that WISC-IV score variation was due to underlying continuous latent factors rather than a typology unique to LD.

### IMPACT STATEMENT

It is often assumed that children with learning disabilities (LD) exhibit unique profiles of ability test scores that reflect idiosyncratic cognitive strengths and weaknesses. However, this person-centered analysis found no profile of ability scores distinctive of children with LD. These results suggest that cognitive test scores may vary due to underlying continuous latent factors rather than a typology unique to LD.

### ARTICLE HISTORY

Received October 22, 2020  
Accepted April 16, 2021

### KEYWORDS

Learning disabilities, latent profile analysis, WISC-IV

### ASSOCIATE EDITOR

Francis Huang

Although formally recognized more than 50 years ago, there is still considerable debate about the diagnosis and treatment of specific learning disabilities (LD; Grigorenko et al., 2020). Although it is generally accepted that learning disabilities are marked by unexpected academic underachievement, their manifestation is otherwise heterogeneous (Grigorenko et al., 2020). Given this heterogeneity, three methods for identifying LD are sanctioned in federal law (U.S. Department of Education, 2004): (a) severe ability–achievement discrepancy; (b) failure to respond to intervention (RTI); and (c) a pattern of cognitive strengths and weaknesses (PSW; Maki et al., 2015; Zirkel, 2017). Further, individual states can permit, prohibit, or require one or more of these methods and local education agencies have wide latitude in implementing federal and state legal mandates (Cottrell & Barrett, 2016; Zumeta et al., 2014).

Surveys have consistently found that a PSW approach for the identification of LD students is favored by school

psychologists (Benson et al., 2020; Kamphaus et al., 2018; Kranzler et al., 2020; Pfeiffer et al., 2000; Sotelo-Dynerga & Dixon, 2014). In fact, PSW approaches have a long history in psychology, first appearing with the emergence of multidimensional intelligence tests and the visual inspection of peaks and valleys in the subtest profile (Rapaport et al., 1945). Statistical methods were subsequently incorporated to “impose some empirical order on profile interpretation; to make sensible inferences from the data with full awareness of errors of measurement and to steer the field away from the psychiatric couch” (Kaufman et al., 2016, p. 7). PSW methods that utilize statistical comparisons to determine cognitive processing strengths and weaknesses are currently widespread (Flanagan et al., 2013; Flanagan & Alfonso, 2017; Groth-Marnat & Wright, 2016; Sattler, 2018). Although these modern approaches are improvements over visual inspections of peaks and valleys, numerous problems and inadequacies have been identified in the extant literature when assessing their

reliability, validity, and diagnostic utility (Canivez, 2013; McGill et al., 2018; Watkins, 2000; Watkins et al., 2005).

The Wechsler scales are often used for assessing children's intellectual ability (Benson et al., 2019) and as a result there have been numerous studies examining the reliability, validity, and diagnostic utility of scores from various Wechsler Intelligence Scale for Children versions (e.g., WISC-R, WISC-III, WISC-IV). Because PSWs historically involved subtest scores, research regarding the reliability, validity, and diagnostic utility of subtest scores, various combinations of subtest scores (e.g., the ACID profile comprised of lower scores on the Arithmetic, Coding, Information, and Digit Span subtests or the SCAD profile comprised of lower scores on the Symbol Search, Coding, Arithmetic, and Digit Span subtests), and ipsative comparisons (the individual's relative strengths/weaknesses) have been of particular interest. Reviews of research indicated inadequate internal consistency for most subtests for individual decision making; longitudinal instability for subtest, pseudocomposite, and ipsative scores; and chance level diagnostic utility (Canivez, 2013; McGill et al., 2018; Watkins et al., 2005). Consequently, the diagnostic use of cognitive strengths and weaknesses has been deemphasized by some authors in favor of a hypothesis generating function that must be confirmed with clinical or testing information (Kaufman & Lichtenberger, 2006; Sattler, 2018). However, as Watkins and Canivez (2004) pointed out, searching for confirming information to support unreliable subtest profiles would likely increase errors rather than reduce them.

Attempts to identify normative subtest profiles using cluster analyses of intelligence tests, including Wechsler scales, were reported and summarized by Canivez (2013) as comparing an individual's subtest profile to a normative profile typology might be considered an actuarial method in locating deviant profiles. Although normative profiles have been identified for some Wechsler scales (WISC-R, WISC-III, WAIS-R, WPPSI) and other intelligence tests (DAS, KABC, UNIT), results have revealed that variation was mostly driven by level/elevation (an indication of general intelligence) rather than shape/pattern; and long-term stability of profile classification appears inadequate (Canivez, 2013; Watkins et al., 2021). Consequently, contemporary approaches have deemphasized PSWs based on subtest scores in favor of PSWs based on factor scores (Flanagan & Alfonso, 2017).

## CONTEMPORARY MODELS FOR IDENTIFYING COGNITIVE PROCESSING PROBLEMS

The view that PSWs are useful in distinguishing LD from rival disorders or low academic achievement has been

embodied in several models that use the broad ability scores (factor scores) provided by modern intelligence tests because it is assumed that those broad ability scores reflect a variety of cognitive processes (Flanagan et al., 2013; Flanagan & Alfonso, 2017; Hale & Fiorello, 2004; Naglieri, 2011). These models conjecture that "individuals with SLD typically present an uneven profile of abilities demonstrating difficulty with some types of learning, but ease with others" (Mather, 2009, p. 41). In essence, it is assumed that children with LD will exhibit unique *profiles* of broad ability scores that will reflect idiosyncratic cognitive strengths and weaknesses that differentially impact academic achievement (Hale et al., 2010; Miller et al., 2016; Saklofske et al., 2016). These models are accepted by school psychologists (Benson et al., 2020; Lockwood & Farmer, 2020). For example, one recent survey of school psychologists found that almost 65% interpreted profiles of broad ability test scores for the identification of LD (Kranzler et al., 2020). Further, a pattern of cognitive strengths and weaknesses for the identification of LD has been endorsed by professional organizations (Christo & Jones, 2014; Learning Disabilities Association of America, 2010) and was declared to be best professional practice (Hale et al., 2010).

## Methods to Assess Profiles

### Variable Centered Methods

Most of the research on the cognitive patterns of children with LD has utilized variable-centered methods (e.g., Compton et al., 2012; Johnson et al., 2010; Niileksela & Reynolds, 2014; Saklofske et al., 2016). That is, methods such as regression and factor analysis that assume all individuals in a sample are drawn from a single population such that estimated parameters apply equally to all members of the group. For example, Lecerf et al. (2016) found significant differences between unexceptional and exceptional groups of children on two cognitive ability scores from the French version of the WISC-IV. Likewise, Saklofske et al. (2016) compared a group of children with LD to a matched comparison sample and found significant cognitive score differences between the two groups. However, even when statistically significant group differences are obtained, the effect sizes may not be large and the distributional overlap may be too great to allow accurate individual diagnostic classification, rendering such differences unhelpful in clinical application (Watkins, 2009). In practice, homogeneity of parameters is an unrealistic assumption because individuals within groups always differ so that inferences based on group-level data may not generalize to individuals (Fisher et al., 2018; Kagan, 2018).

## Person-Centered Methods

In contrast, person-centered approaches relax the assumption of population homogeneity, which allows parameters to vary across unobserved subpopulations of participants (Morin et al., 2020; Wang & Wang, 2020). Person-centered methods “generate a typology in which participants are classified into qualitatively and quantitatively distinct profiles based on their specific combinations of strengths and weaknesses on the same array of competencies” (Morin & Marsh, 2015, p. 40). Until recently, cluster analysis was the only easily accessible person-centered method available to researchers. Although a great number of cluster analysis studies identifying LD subtypes have been conducted, the variety of clustering methods, samples, and measures have made results difficult to generalize (McKinney, 1984; Saklofske et al., 2016; Speece, 2003). For example, cluster analyses that employed correlation as a similarity measure produced different results than analyses that relied on distance measures (Loehlin et al., 2018; McDermott et al., 1989).

Loehlin (2019) reviewed studies that identified cognitive clusters and concluded that “perhaps stable and consistent cognitive clusters simply do not exist. Perhaps they exist, but only for particular populations, or across particular sets of cognitive measures, or are detectable by only certain clustering methods” (p. 22). Importantly, many cluster analysis studies failed to include samples of unexceptional learners making it difficult to determine whether their results were characteristic of LD or simply represented individual differences among unexceptional learners (McKinney, 1984). Additionally, “cluster analysis techniques will always produce clusters, even in random data. The investigator must be concerned about whether the resulting clusters are discovered or forced by the technique” (Speece, 1994, p. 36).

Latent mixture models have recently eclipsed cluster analysis because latent mixture models are model-based, more flexible, less subjective, better able to accommodate complex multivariate interaction effects and measurement error, and provide a probabilistic classification of each individual (DiStefano & Kamphaus, 2006; Hickendorff et al., 2018; Morin et al., 2020; Oberski, 2016; Peugh & Fan, 2013; Woo et al., 2018). Statistical software to conduct such analyses are also now more readily available. Mixture models assume that an observed sample includes several homogeneous subpopulations, each with its own distribution that results in a ‘mixture’ of parameters (Morin et al., 2020). Mixture models are characterized by categorical latent variables where each category represents an inferred subpopulation. Memorably, “mixture modeling is the art of unscrambling eggs: It recovers hidden groups from observed data” (Oberski, 2016, p. 275).

Called latent profile analysis (LPA) when the observed indicator variables are continuous (Hickendorff et al., 2018; Wang & Wang, 2020), the goal of LPA “is to uncover

latent profiles or groups of individuals who share a meaningful and interpretable pattern of responses on the measures of interest” (Ferguson et al., 2020, p. 459). The use of LPA with LD populations has been recommended by measurement specialists (Bray & Dziak, 2018; Willson & Rupley, 2013) and several studies have successfully applied mixture models with standardized and unstandardized academic and cognitive indicators among participants with LD (e.g., Geary et al., 2009; Niileksela & Templin, 2019; Swanson et al., 2018). Although variable-centered methods have suggested that unique broad ability score profiles characterize children with LD (Giofrè et al., 2017; Miller et al., 2016; Toffalini et al., 2017), LPA using the broad cognitive ability scores from modern intelligence tests has not been conducted among children with LD.

## Present Study

The present study was designed to fill that evidential lacuna by applying LPA to the broad ability (factor index) scores from the WISC-IV (Wechsler, 2003a) of a large number of exceptional and unexceptional learners to ascertain the number and kind of cognitive subtypes that emerge and to determine if a cognitive typology of children with LD manifests. If unique LD cognitive profiles exist, it is hypothesized that one or more profiles exclusively or predominantly populated by students with LD will be revealed.

## METHOD

### Participants

The first sample was composed of 1,830 (59.6% male) children with school-identified LD with an average age of 10.4 years ( $SD=2.4$ ). Ethnic background of these participants was 67% White, 14% Hispanic, 10% Black, 6% Native American, 1% Asia/Pacific, and 2% Other or Unspecified. No additional information about the participants with LD was available to ensure anonymity. The second sample contained 2,200 simulated normative participants (available as [supplementary material](#)). Although data from the WISC-IV normative sample was preferable, access to those data was twice refused by the publisher (Pearson, personal communication, February 12, 2010 and September 8, 2020).

### Instruments

The WISC-IV (Wechsler, 2003a) was standardized on a nationally representative sample of 2,200 children aged 6–16 years closely approximating the 2000 United States

Census based on sex, race, parent education level, and geographic region. The WISC-IV contains 10 core subtests ( $M = 10$ ,  $SD = 3$ ) that form four factor index scores: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI). The Full Scale IQ (FSIQ) and factor indices have means of 100 and standard deviations of 15. Substantial reliability and validity evidence was provided by Wechsler (2003b).

Examiners used a variety of academic achievement measures, but the majority of scores were from a version of the Wechsler Individual Achievement Test (47.3%) and a version of the Woodcock-Johnson Tests of Achievement (42.3%). Both measures are well-developed scales with nationally representative normative samples and strong psychometric characteristics (Thorndike & Thorndike-Christ, 2010).

## Procedure

The sample of children with learning disabilities was drawn from the data sets of four published studies that examined the structural validity evidence of the WISC-IV among referred samples (Canivez, 2014; Devena et al., 2013; Styck & Watkins, 2016; Watkins, 2010). Data from three of the studies were obtained through file review and data from the fourth study were from anonymous school psychologists electronically submitting test scores and relevant information for anonymous children. In total, there were 2,669 students in those validity studies but only 1,830 were identified as LD and had WISC-IV index scores. Those 1,830 participants were evaluated by 440 school psychologists in 23 states. As reflected in Table 2, participants with LD were heterogeneous with respect to academic area of disability but ability-achievement discrepancies generally reflected school-identified area of specific academic disability.

**Table 1.** Scores on the Wechsler Intelligence Scale for Children-Fourth Edition for 1,830 Participants With Learning Disabilities and 2,200 Simulated Normative Participants

| Score                      | LD   |      | Simulated Norm |      |
|----------------------------|------|------|----------------|------|
|                            | Mean | SD   | Mean           | SD   |
| Block Design               | 9.1  | 2.7  | 10.0           | 3.0  |
| Similarities               | 8.8  | 2.6  | 10.1           | 3.0  |
| Digit Span                 | 8.0  | 2.5  | 10.0           | 3.0  |
| Picture Concepts           | 9.9  | 2.8  | 10.1           | 3.0  |
| Coding                     | 8.5  | 2.9  | 10.0           | 2.9  |
| Vocabulary                 | 8.5  | 2.5  | 10.0           | 3.0  |
| Letter-Number Sequencing   | 8.4  | 2.7  | 10.1           | 3.0  |
| Matrix Reasoning           | 9.3  | 2.6  | 10.1           | 3.0  |
| Comprehension              | 9.1  | 2.5  | 10.1           | 2.9  |
| Symbol Search              | 8.9  | 2.8  | 10.0           | 3.0  |
| Verbal Comprehension Index | 92.8 | 12.0 | 100.0          | 15.0 |
| Perceptual Reasoning Index | 96.8 | 12.9 | 100.0          | 15.0 |
| Working Memory Index       | 89.6 | 11.9 | 100.0          | 15.0 |
| Processing Speed Index     | 92.9 | 13.6 | 100.0          | 15.0 |
| Full Scale IQ              | 91.6 | 11.1 | 100.0          | 15.0 |

Note. LD = learning disability. Univariate skew and kurtosis < 1.0 for all scores.

The simulated normative sample was generated with EQS 6.4 (Bentler, 2006) using the descriptive statistics and factor structure provided by Wechsler (2003b). As is apparent from Table 1, the scores from this simulated sample exhibited means and standard deviations almost identical to those reported by Wechsler (2003b) for the real normative sample. Additionally, all simulated correlations were within  $\pm .04$  of those reported by Wechsler (2003b) and a confirmatory factor analysis found that the oblique first-order structure favored by Wechsler (2003b) was a good fit to these simulated data ( $RMSEA = .04$ ,  $CFI = .98$ ).

## Analyses

### LPA Model Specification

It has been suggested that broad ability scores (i.e., factor index scores) are the optimal level of interpretation of IQ tests and that analysis of index scores is good clinical practice (Beal et al., 2019; Kovacs & Conway, 2019). Further, Morin et al. (2020) recommended that factor scores be used in LPA because they offer better control of measurement error than observed scale scores and they preserve the underlying measurement structure of the instrument. Accordingly, the four WISC-IV factor index scores from the total sample of 4,030 cases served as indicator variables in LPA conducted with Mplus version 8.4 (Muthén & Muthén, 1998–2019) using robust maximum likelihood estimation. To ensure convergence on the global maximum, 3,000 sets of random start values were initially generated for each model. If the best loglikelihood value was not found by at least three different start values, then the number of random starts was increased until the best loglikelihood was replicated or 10,000 start values were exhausted (Morin et al., 2020).

Two constraints were imposed on LPA solutions to minimize the number of parameters to be estimated: (a) the local independence assumption that latent profile membership explains covariance among indicators; and (b) the homogeneity assumption of equality of profile-specific variances (Peugh & Fan, 2013). Local independence is similar to the assumption of uncorrelated residuals in factor analysis and its relaxation is not generally recommended (Morin et al., 2020; Nylund-Gibson & Choi, 2018; Schweizer, 2012). However, unconstrained models may require fewer profiles to fit the data, so a post-hoc strategy of assessing the effects of relaxing these constraints was also employed (Berlin et al., 2014; Lubke & Luningham, 2017; Masyn, 2013).

### LPA Model Fit

Mixture models are typically exploratory (Hickendorff et al., 2018; Lubke & Luningham, 2017; Morin et al., 2020). Therefore, LPA models were fit in a stepwise fashion,

**Table 2.** Academic Performance of 1,830 Participants With Learning Disabilities by Area of Academic Disability

| LD Area                  | WISC-IV FSIQ |      |      | Reading |      | Math |      | Writing |      |
|--------------------------|--------------|------|------|---------|------|------|------|---------|------|
|                          | %            | Mean | SD   | Mean    | SD   | Mean | SD   | Mean    | SD   |
| Reading                  | 18.6         | 94.6 | 10.4 | 82.4    | 8.6  | 95.3 | 10.0 | 95.7    | 9.9  |
| Math                     | 13.0         | 89.7 | 9.8  | 93.1    | 8.4  | 80.7 | 9.8  | 94.2    | 12.3 |
| Writing                  | 8.0          | 95.0 | 11.5 | 93.0    | 10.1 | 97.6 | 11.0 | 84.2    | 12.0 |
| Reading & Math           | 9.3          | 90.1 | 10.2 | 81.4    | 9.3  | 81.5 | 10.3 | 94.0    | 11.7 |
| Reading & Writing        | 21.9         | 95.1 | 11.6 | 79.9    | 9.4  | 95.8 | 10.6 | 84.1    | 12.7 |
| Math & Writing           | 5.0          | 87.4 | 8.3  | 90.7    | 8.1  | 81.2 | 8.6  | 80.5    | 11.8 |
| Reading, Math, & Writing | 20.6         | 87.9 | 10.9 | 76.7    | 10.8 | 80.2 | 10.6 | 80.2    | 12.4 |
| Total Reading            | 70.4         | 92.1 | 11.3 | 79.8    | 9.9  | 89.2 | 12.8 | 87.5    | 13.5 |
| Total Math               | 48.0         | 88.5 | 10.3 | 83.5    | 12.0 | 80.7 | 10.1 | 87.1    | 14.0 |
| Total Writing            | 55.5         | 91.6 | 11.5 | 81.5    | 11.6 | 88.8 | 13.2 | 82.4    | 12.6 |
| Total                    | 96.4         | 91.7 | 11.2 | 83.3    | 11.2 | 88.3 | 12.9 | 87.9    | 13.6 |

Note. LD = learning disability. 3.6% of the participants with LD were missing FSIQ scores, 3.0% were missing reading scores, 3.2% were missing math scores, and 15.7% were missing writing scores.

starting by estimating a one-profile solution and then successively adding profiles until the model failed to converge or produced a statistically improper solution, indicating that a more parsimonious model might be appropriate (Ferguson et al., 2020; Hickendorff et al., 2018; Masyn, 2013; Morin et al., 2020; Nylund-Gibson & Choi, 2018; Wang & Wang, 2020).

Unfortunately, there is no universally accepted measure of statistical adequacy for LPA models (Ferguson et al., 2020; Masyn, 2013; Nylund-Gibson & Choi, 2018). Consequently, the Bayesian information criterion (BIC; Schwarz, 1978), sample size adjusted BIC (aBIC; Sclove, 1987), adjusted Vuong-Lo-Mendell-Rubin likelihood ratio test (aLMR; Lo et al., 2001), and Bootstrapped likelihood ratio test (BLRT; McLachlan & Peel, 2000) were employed to determine model fit (Ferguson et al., 2020; McLachlan & Peel, 2000; Morin et al., 2020; Nylund-Gibson & Choi, 2018; Wang & Wang, 2020).

**Information Criteria.** The BIC and aBIC are information criteria that balance parsimony and goodness-of-fit. BIC adjusts for the number of free parameters and sample size and aBIC adjusts for the number of free parameters but decreases the sample size penalty of the BIC (McLachlan & Peel, 2000). With both criteria, the smallest value points to the preferred profile solution. The BIC has been favored in several simulation studies, especially for continuous indicator variables (Morgan, 2015; Morin & Wang, 2016; Nylund et al., 2007). In contrast, the aBIC sometimes overestimated the number of profiles, especially with large sample sizes (Dziak et al., 2020).

**Inferential Tests.** The aLMR and BLRT are inferential tests of the difference between the current model and a model with one less profile. Statistical nonsignificance indicates that model fit is not statistically improved by the addition

of the current profile. Both tests are influenced by sample size. The BLRT has outperformed other indices in simulation research (Morgan, 2015; Nylund et al., 2007). Given that multiple tests can inflate the Type I error rate (Lubke & Lunningham, 2017), statistical significance was set at  $.05 \div 7 = .007$  for each aLMR and BLRT test to maintain an overall alpha level of .05.

**Classification Diagnostics.** Although not used for initial model selection, these classification criteria may be used “to judge the utility of the [LPA] directly applied to a particular set of indicators to produce highly-differentiated groups in the sample” (Masyn, 2013, p. 570). Entropy is a measure of the overall precision of classification of individuals into profiles. Entropy values  $\geq .80$  are preferred, values  $\geq .70$  are acceptable, and values  $\leq .60$  are inadequate (Asparouhov & Muthén, 2014; Ferguson et al., 2020; Nylund-Gibson & Choi, 2018; Wang & Wang, 2020). The average posterior probability within each profile is a measure of the precision of classification within each profile. Average posterior probability values  $\geq .80$  indicate a good profile solution and values  $\geq .70$  indicate acceptable profile classification (Nylund-Gibson & Choi, 2018; Wang & Wang, 2020). Additionally, the relative size of emergent profiles was reviewed to ensure sufficient power for generalization to the population and to guard against a small set of outliers ( $< 1\%$ ; Berlin et al., 2014; Ferguson et al., 2020). Finally, each solution was evaluated for its substantive meaning and theoretical coherence (Ferguson et al., 2020; Morin et al., 2020; Morin & Wang, 2016; Nylund-Gibson & Choi, 2018; Wang & Wang, 2020).

### Post Hoc Analyses

Although it may not control measurement error (Wang & Wang, 2020), the classify–analyze approach was used

because it more closely aligns with clinical practice (Benson et al., 2020) and prior cluster analyses (Poletti et al., 2018). That is, each individual's most probable profile membership was treated as a nominal variable based on posterior probabilities (Hickendorff et al., 2018). Those nominal variables were used for subsequent analyses.

## RESULTS

### WISC-IV Scores

Descriptive statistics for the WISC-IV scores of both samples are presented in Table 1. WISC-IV scores were relatively normally distributed but were slightly lower and less variable than average for the LD sample. Similar patterns have been found with other samples of referred students (Canivez & Watkins, 1998; Johnson et al., 2010).

### LPA Models

Honoring the local independence and homogeneity assumptions, constrained models with one through eight profiles were assessed. Those results are provided in Table 3. As is often found (Bray & Dziak, 2018; Nylund-Gibson & Choi, 2018), no single model was unequivocally identified. The aBIC reached a nadir at seven profiles and the BLMR also signaled that no more than seven profiles were needed. However, both aBIC and BLMR may have been biased by the large sample size (Dziak et al., 2020). Additionally, the entropy of model seven was low (.61) and two of its profiles were marked by inadequate average posterior probability values (.47). Models with six and eight profiles also exhibited inadequate average posterior probability values and/or profile membership <1% of the sample. Accordingly, models six, seven, and eight were considered unsuitable.

**Table 3.** Fit Statistics and Classification Accuracy for Latent Profile Analysis of Factor Scores From the Wechsler Intelligence Scale for Children-Fourth Edition for 1,830 Students With Learning Disabilities and 2,200 Simulated Normative Participants (N = 4,030)

| Profile | Loglikelihood | BIC           | aBIC          | $\Delta$ BIC | $p$          |              |         |
|---------|---------------|---------------|---------------|--------------|--------------|--------------|---------|
|         |               |               |               |              | aLMR         | BLMR         | Entropy |
| 1       | -65864        | 131794        | 131768        | 3820         | -            | -            | -       |
| 2       | -64386        | 128880        | 128839        | 906          | .0001        | .0001        | 0.70    |
| 3       | -63974        | 128097        | 128039        | 123          | .0001        | .0001        | 0.70    |
| 4       | -63891        | <b>127974</b> | 127901        | 0            | .0001        | .0001        | 0.67    |
| 5       | -63872        | 127977        | 127888        | 3            | <b>.0136</b> | .0001        | 0.65    |
| 6       | -63857        | 127988        | 127883        | 14           | .0113        | .0001        | 0.66    |
| 7       | -63837        | 127989        | <b>127868</b> | 15           | .1619        | .0001        | 0.61    |
| 8       | -63827        | 128012        | 127875        | 38           | .0224        | <b>.0380</b> | 0.64    |

Note. BIC = Bayesian information criterion; aBIC = sample size adjusted BIC; aLMR = adjusted Vuong-Lo-Mendell-Rubin likelihood ratio test; BLMR = Bootstrapped likelihood ratio test. Lowest information values and first nonsignificant (< .007) inferential tests in bold.

The BIC value was minimal for model four but its so-called 'Occam's window' (i.e.,  $\Delta$ BIC values of 0-3), also included the five profile model (Kass & Raftery, 1995). However, one profile in model five exhibited inadequate average posterior probability (.44), making model five less desirable. Model four was also identified by the aLMR test and exhibited near-acceptable entropy values (.67), acceptable to good average posterior probability values (.73-.84), and sizeable profile membership ( $n \geq 157$ ). Model four seemed to have been derived from the collapse of two profiles from model five (including the one with inadequate precision) into one more adequate profile in model four. Namely, the mean of the four indicator scores for the profiles in model five compared to model four were: 79 and 87 versus 84; 98 versus 96; 108 versus 107; and 121 versus 121.

Constrained models can provide an approximate upper bound for the number of profiles and allow consideration of a smaller number of plausible unconstrained models (Dziak et al., 2020; Lubke & Luningham, 2017). Relaxing the homogeneity assumption for models with three through five profiles did not improve BIC or aBIC values over the constrained models. Relaxing the local independence assumption resulted in a failure to replicate the best loglikelihood value and/or statistically inappropriate solutions for those models. Thus, the more parsimonious constrained models are more probable. Among those models, model four was identified as most appropriate by the BIC and aLMR and its indicator variables appeared to reasonably separated and homogeneous (see Table 4); consequently, the model with four profiles was judged the most adequate in terms of both statistical fit and theoretical interpretability (Bray & Dziak, 2018). However, these profiles were differentiated quantitatively rather than qualitatively. That is, they differed in level but not in shape as illustrated in Figure 1.

### Post Hoc Analyses

As expected, there were statistically significant ( $p < .001$ ) differences in VCI, PRI, WMI, and PSI scores across the four profiles and across the two samples (Table 4). However, the size of the index score differences between the simulated norm and LD samples within each profile tended to be of little practical significance. Ferguson (2009) suggested that  $d > .40$  is the recommended minimum effect size that represents a practically significant effect for social science data. Table 4 shows that 5 of the 16 comparisons met this practical significance level but two of those five were based on an unstable sample ( $n = 9$ ). The WMI score was practically smaller for LD participants in profiles 2 and 3 while the PRI score was practically smaller for the simulated normative sample in profile 3.

**Table 4.** Characteristics of the Four Profiles for 1,830 Students With LD and 2,200 Simulated Normative Participants

|          | Profile 1 |      | Profile 2 |       | Profile 3 |        | Profile 4 |        |
|----------|-----------|------|-----------|-------|-----------|--------|-----------|--------|
|          | Norm      | LD   | Norm      | LD    | Norm      | LD     | Norm      | LD     |
| VCI      | 82.1      | 83.0 | 95.9      | 94.6  | 110.5     | 109.5  | 124.6     | 121.1* |
| PRI      | 81.4      | 84.5 | 97.2      | 99.9  | 108.9     | 112.7* | 126.5     | 127.4  |
| WMI      | 83.8      | 79.7 | 96.1      | 91.9* | 110.1     | 102.7* | 121.0     | 121.9  |
| PSI      | 85.6      | 84.6 | 97.9      | 95.0  | 107.3     | 103.0  | 118.5     | 109.2* |
| <i>n</i> | 415       | 539  | 920       | 1,095 | 717       | 187    | 148       | 9      |
| APP      | 0.80      |      | 0.76      |       | 0.84      |        | 0.73      |        |

Note. Norm = simulated normative sample, LD = children with learning disability, VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index, and APP = average posterior probability.

\* Standardized difference (*d*) between Norm and LD samples > .40.

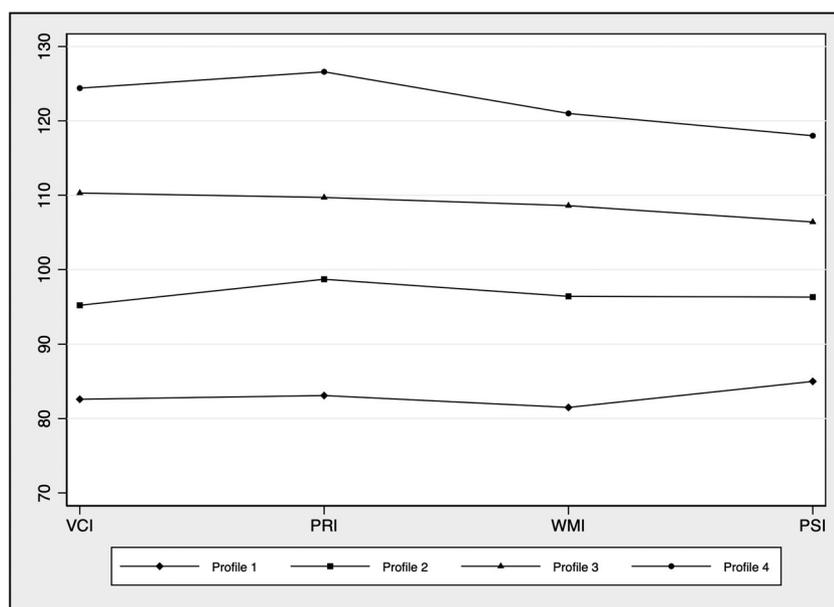
In terms of profile membership, there was little differentiation between the number of simulated normative and LD participants in each cognitive profile with the exception that participants with LD were less likely to be members of higher IQ profiles than lower IQ profiles (comprising 56.5%, 54.3%, 20.7%, and 5.7% of the membership of ascending IQ profiles). Thus, children with low average to average IQ scores were almost equally likely to belong to simulated normative and LD groups and there was no profile exclusively or even predominantly populated by children with LD.

Given that LPA was simultaneously applied to both LD and simulated normative samples, it might be speculated that a unique LD profile did not emerge because it was overpowered by the large number of normative participants within the LPA. To evaluate the verity of this supposition, LPA was applied to the 1,830 children with LD and,

employing the same methodology as above, a model with three profiles was judged the most adequate in terms of both statistical fit and theoretical interpretability. These three profiles were again distinguished by level rather than shape. The 2,200 simulated normative participants were then probabilistically classified into the three profiles defined by the LD sample. No profile was comprised exclusively or predominantly by LD members. For example, profile 1 with a mean WISC-IV index score of 85 contained 88% normative and 12% LD participants, profile 2 with a mean index score of 99 contained 52% normative and 48% LD participants, and profile 3 with a mean index scores of 114 contained 88% normative and 12% LD participants. Likewise, none of these profiles was comprised exclusively or predominantly by participants with a specific academic disability in reading, math, or writing. For instance, children with a specific reading LD comprised 19%, 22%, and 14% of the total LD membership in profiles 1 through 3, respectively, with mean WISC-IV index scores of 87, 98, and 111, respectively.

## DISCUSSION

Loehlin (2019) asked if “individuals tend to fall into a small number of distinct clusters based on their patterns of cognitive skill?” (p. 19). This study applied LPA to the WISC-IV index scores of 1,830 children with LD and 2,200 simulated normal participants to answer that question. LPA, a latent mixture method that assumes a heterogeneous sample includes several homogeneous subpopulations, results favored a model with four profiles that differed in level but

**Figure 1.** WISC-IV Factor Index Means for the Four-Profile Model

Note. VCI = Verbal Comprehension Index, PRI = Perceptual Reasoning Index, WMI = Working Memory Index, PSI = Processing Speed Index.

not in shape. Thus, this latent mixture model analysis found no mixture of subpopulations (Harring & Hodis, 2016), suggesting that WISC-IV score variation was due to underlying continuous latent factors rather than a typology (Berlin et al., 2014; Lubke & Miller, 2015; Morin & Wang, 2016).

These results are not entirely unexpected because cluster analysis studies of cognitive abilities have often found clusters characterized by level rather than shape (Canivez, 2013; Loehlin, 2019; McDermott et al., 1989). Branum-Martin et al. (2013, Nov-Dec) demonstrated that clusters or profiles can artificially emerge when cut-scores are applied to correlated indicator variables. Reading and math skills have also been shown to be dimensional rather than categorical (Child et al., 2019; Snowling & Hulme, 2012).

Characteristics of the WISC-IV may also have contributed to the present results. As illustrated by Watkins (2010) with the WISC-IV standardization sample and Watkins et al. (2006), Watkins (2010), and Canivez (2014) with samples of referred students, WISC-IV factor index scores conflate general and group factor variance and when that variance is decomposed, the VCI, PRI, and WMI scores account for little unique variance. This problem is not limited to the WISC-IV as similar results were observed for the WISC-V (Canivez et al., 2016, 2017), WJ IV (Dombrowski et al., 2017, 2018), and DAS-II (Canivez et al., 2020; Canivez & McGill, 2016). Insufficient unique variability among factor index scores may make it difficult to identify and describe profiles beyond level.

The present results, in combination with past research documenting the psychometric inadequacies of PSW approaches, indicate that such methods should not be used or promoted in assessment and classification for LD. Both National Association of School Psychologists (2010) and APA (2002, 2010) ethical standards require that test scores and procedures be interpreted in light of empirical evidence and a basic tenet of scientifically based psychology is that replicated, peer-reviewed evidence is needed *before* a method is accepted for clinical practice (McFall, 2000). Some might suggest that those reporting negative research findings offer a replacement. However, “the burden of proof in science rests on the individual making a claim, not the critic” (Lilienfeld et al., 2015, p. 8). Therefore, it is incumbent on those promoting PSW methods to provide evidence for their reliability, validity, diagnostic utility, and ultimately, treatment validity and until that supportive evidence is presented, such methods cannot be advocated.

## LIMITATIONS

This study is unique in applying LPA to large samples of children with and without LD but it is not without limitations. First, the sample of children with LD were extracted

from data sets obtained for published structural validity studies with participants that may have been selected based on ability-achievement discrepancies that were prevalent in the past. This diagnostic standard may have biased the sample in unknown ways. Further, there is no way to know exactly what the classification criteria were for individual LD identification or the extent to which multidisciplinary teams might have deviated from the state or local criteria or methods in their classifications of LD. Also, the sample of children with LD was heterogeneous in respect to area of academic eligibility (see Table 2). Results might differ if there were sufficient numbers of children for analyses based on homogeneous academic eligibility (i.e., reading only, math only, etc.). Relatedly, these data and analyses pertain to the WISC-IV which was replaced by the WISC-V in 2014. LPA with the WISC-V might produce different results but to date no such studies have appeared in the peer reviewed literature. Another limitation relates to the normative sample used in the present study which was simulated rather than actual. Although the simulated data were psychometrically similar to the WISC-IV normative sample based on descriptive statistics, correlations, and factor structure, it might differ in unknown ways. It is unfortunate that the publisher denied access to the normative sample, so simulating data to match normative data was the only option available.

## CONCLUSIONS

A pattern of cognitive strengths and weaknesses for the identification of LD has been endorsed by professional organizations (Christo & Jones, 2014; Learning Disabilities Association of America, 2010), was declared to be best professional practice (Hale et al., 2010), and is commonplace among school psychologists (Benson et al., 2020; Kranzler et al., 2020; Lockwood & Farmer, 2020). These practices are based on the assumption that children with LD will exhibit unique *profiles* of broad ability scores that will reflect idiosyncratic cognitive strengths and weaknesses that differentially impact academic achievement (Hale et al., 2010; Miller et al., 2016; Saklofske et al., 2016).

Despite its popularity, interpretation of cognitive test profiles is not supported by research evidence (Floyd & Kranzler, 2019; Grigorenko et al., 2020; Kranzler et al., 2019, 2020; McGill et al., 2018; Miciak et al., 2018; Watkins, 2009) and has been described as a “shared professional myth” (Watkins, 2000, p. 465) and a “misuse of IQ scores” (Beaujean et al., 2018, p. 18). As summarized by Schneider and Kaufman (2017), “the evidence for the utility of using cognitive ability tests to diagnose learning disabilities is brittle and weak, and many assessment practices urgently need reform.” Clinical conjectures, personal anecdotes,

and rhetoric backed by citations of similar subjective claims do not constitute scientific evidence (Schneider & Kaufman, 2017), they offer only an “illusion of validity” (Kahneman & Klein, 2009, p. 517). Appeals to theory absent supporting empirical evidence are equally inadequate in scientific clinical practice. In contrast, clinicians must operate on the best available scientific evidence (Wodarski & Hopson, 2012) within an evidence-based assessment framework (Youngstrom et al., 2017). In agreement with Loehlin’s (2019) review of the cluster analysis literature, we did not find a typology of children with LD identified by cognitive profiles, but rather, these WISC–IV data were best represented by continuous latent factors rather than by categorical latent profiles. Until additional research is conducted which demonstrates the presence of such profiles, the best available evidence suggests that broad ability factor scores do not aggregate into cognitive profiles unique to children with LD. Thus, PSW methods to identify children with LD are not supported by the available evidence and should be eschewed.

## DISCLOSURE

We have no conflicts of interest to disclose.

## ORCID

Marley W. Watkins  <http://orcid.org/0000-0001-6352-7174>  
 Gary L. Canivez  <http://orcid.org/0000-0002-5347-6534>

## REFERENCES

- American Psychological Association. (2002, 2010 Amendments). *Ethical principles of psychologists and code of conduct*. APA.
- Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 329–341. <https://doi.org/10.1080/10705511.2014.915181>
- Beal, A. L., Holdnack, J. A., Saklofske, D. H., & Weiss, L. G. (2019). Practical considerations in WISC–V interpretation and intervention. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC–V: Clinical use and interpretation* (2nd ed., pp. 23–54). Academic Press.
- Beaujean, A. A., Benson, N. F., McGill, R. J., & Dombrowski, S. C. (2018). A misuse of IQ scores: Using the dual discrepancy/consistency model for identifying specific learning disabilities. *Journal of Intelligence*, 6(3), 36–25. <https://doi.org/10.3390/jintelligence6030036>
- Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology*, 72, 29–48. <https://doi.org/10.1016/j.jsp.2018.12.004>
- Benson, N. F., Maki, K. E., Floyd, R. G., Eckert, T. L., Kranzler, J. H., & Fefer, S. A. (2020). A national survey of school psychologists’ practices in identifying specific learning disabilities. *School Psychology*, 35(2), 146–157. <https://doi.org/10.1037/spq0000344>
- Bentler, P. M. (2006). *EQS 6 structural equation program manual*. Multivariate Software.
- Berlin, K. S., Williams, N. A., & Parra, G. R. (2014). An introduction to latent variable mixture modeling (part 1): Overview and cross-sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, 39(2), 174–187. <https://doi.org/10.1093/jpepsy/jst084>
- Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2013, Nov-Dec). Classification and identification of reading and math disabilities: The special case of comorbidity. *Journal of Learning Disabilities*, 46(6), 490–499. <https://doi.org/10.1177/0022219412468767>
- Bray, B. C., & Dziak, J. J. (2018). Commentary on latent class, latent profile, and latent transition analysis for characterizing individual differences in learning. *Learning and Individual Differences*, 66, 105–110. <https://doi.org/10.1016/j.lindif.2018.06.001>
- Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwann (Eds.), *The Oxford handbook of child psychological assessment* (pp. 84–112). 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: The Official Journal of the Division of School Psychology, American Psychological Association*, 29(1), 38–51. <https://doi.org/10.1037/spq0000032>
- Canivez, G. L., & McGill, R. J. (2016). Factor structure of the Differential Ability Scales–Second Edition: Exploratory and hierarchical factor analyses with the core subtests. *Psychological Assessment*, 28(11), 1475–1488. <https://doi.org/10.1037/pas0000279>
- Canivez, G. L., McGill, R. J., & Dombrowski, S. C. (2020). Factor structure of the Differential Ability Scales–Second Edition core subtests: Standardization sample confirmatory factor analyses. *Journal of Psychoeducational Assessment*, 38(7), 791–815. <https://doi.org/10.1177/0734282920914792>
- Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children–Third Edition. *Psychological Assessment*, 10(3), 285–291. <https://doi.org/10.1037/1040-3590.10.3.285>
- Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children–Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, 28(8), 975–986. <https://doi.org/10.1037/pas0000238>
- 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(4), 458–472. <https://doi.org/10.1037/pas0000358>
- Child, A. E., Cirino, P. T., Fletcher, J. M., Willcutt, E. G., & Fuchs, L. S. (2019). A cognitive dimensional approach to understanding shared and unique contributions to reading, math, and attention skills. *Journal of Learning Disabilities*, 52(1), 15–30. <https://doi.org/10.1177/0022219418775115>
- Christo, C., & Jones, J. (2014). *CASP position paper: Specific learning disabilities and patterns of strengths and weaknesses*. <https://www.casonline.org/pdfs/position-papers/SLD-PSW%20position%20paper%20final.pdf>

- Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45*(1), 79–95. <https://doi.org/10.1177/0022219410393012>
- Cottrell, J. M., & Barrett, C. A. (2016). Defining the undefinable: Operationalization of methods to identify specific learning disabilities among practicing school psychologists. *Psychology in the Schools, 53*(2), 143–157. <https://doi.org/10.1002/pits.21892>
- 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*(6), 591–599. <https://doi.org/10.1177/0734282913483981>
- DiStefano, C., & Kamphaus, R. W. (2006). Investigating subtypes of child development: A comparison of cluster analysis and latent class cluster analysis in typology creation. *Educational and Psychological Measurement, 66*(5), 778–794. <https://doi.org/10.1177/0013164405284033>
- Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ-IV Cognitive at school age. *Psychological Assessment, 29*(4), 394–407. <https://doi.org/10.1037/pas0000350>
- Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). An alternative conceptualization of the theoretical structure of the WJ IV Cognitive at school age: A confirmatory factor analytic investigation. *Archives of Scientific Psychology, 6*(1), 1–13. <https://doi.org/10.1037/arc0000039>
- Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jerimiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics, 21*(2), 553–565. <https://doi.org/10.1093/bib/bbz016>
- Farmer, R. L., McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2020, December). Why questionable assessment practices remain popular in school psychology: Instructional materials as pedagogic vehicles. *Canadian Journal of School Psychology*. Advance Online Publication. <https://doi.org/10.1177/0829573520978111>
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538. <https://doi.org/10.1037/a0015808>
- Ferguson, S. L., Moore, E. W. G., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development, 44*(5), 458–468. <https://doi.org/10.1177/0165025419881721>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences, 115*(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>
- Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of the WISC-V assessment*. Wiley.
- Flanagan, D., Ortiz, S., & Alfonso, V. C. (2013). *Essentials of cross battery assessment*. (3rd ed.). Wiley.
- Floyd, R. G., & Kranzler, J. H. (2019). Remediating student learning problems: Aptitude-by-treatment interaction versus skill-by-treatment interaction. In M. K. Burns (Ed.), *Introduction to school psychology: Controversies and current practice* (pp. 413–431). Oxford University Press.
- Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L. (2009). First-grade predictors of mathematical learning disabilities: A latent class trajectory analysis. *Cognitive Development, 24*(4), 411–429. <https://doi.org/10.1016/j.cogdev.2009.10.001>
- Giofrè, D., Toffalini, E., Altoe, G., & Cornoldi, C. (2017). Intelligence measures as diagnostic tools for children with specific learning disabilities. *Intelligence, 61*, 140–145. <https://doi.org/10.1016/j.intell.2017.01.014>
- Grigorenko, E. L., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *The American Psychologist, 75*(1), 37–51. <https://doi.org/10.1037/amp0000452>
- Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Wiley.
- Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., Cohen, M., Davis, A., Decker, S., Denckla, M., Dumont, R., Elliott, C., Feifer, S., Fiorello, C., Flanagan, D., Fletcher-Janzen, E., Geary, D., Gerber, M., Gerner, M., ... Yalof, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33*(3), 223–236. <https://doi.org/10.1177/073194871003300310>
- Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. Guilford.
- Harring, J. R., & Hodis, F. A. (2016). Mixture modeling: Applications in educational psychology. *Educational Psychologist, 51*(3–4), 354–367. <https://doi.org/10.1080/00461520.2016.1207176>
- Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences, 66*, 4–15. <https://doi.org/10.1016/j.lindif.2017.11.001>
- Johnson, E. S., Humphrey, M., Mellard, D. F., Woods, K., & Swanson, H. L. (2010). Cognitive processing deficits and students with specific learning disabilities: A selective meta-analysis of the literature. *Learning Disability Quarterly, 33*(1), 3–18. <https://doi.org/10.1177/073194871003300101>
- Kagan, J. (2018). Kinds of individuals defined by patterns of variables. *Development and Psychopathology, 30*(4), 1197–1209. <https://doi.org/10.1017/S095457941800055X>
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise. *American Psychologist, 64*(6), 515–526. <https://doi.org/10.1037/a0016755>
- Kamphaus, R. W., PierceWinsor, A., Rowe, E. W., & Kim, S. (2018). A history of intelligence test interpretation. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 56–72). Guilford.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). Wiley.
- Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V*. Wiley.
- Kovacs, K., & Conway, A. R. A. (2019). A unified cognitive/differential approach to human intelligence: Implications for IQ testing. *Journal of Applied Research in Memory and Cognition, 8*(3), 255–272. <https://doi.org/10.1016/j.jarmac.2019.05.003>
- Kranzler, J. H., Gilbert, K., Robert, C. R., Floyd, R. G., & Benson, N. F. (2019). Further examination of a critical assumption

- underlying the dual-discrepancy/consistency approach to specific learning disability identification. *School Psychology Review*, 48(3), 207–221. <https://doi.org/10.17105/SPR-2018-0008.V48-3>
- Kranzler, J. H., Maki, K. E., Benson, N. F., Eckert, T. L., Floyd, R. G., & Fefer, S. A. (2020). How do school psychologists interpret intelligence tests for the identification of specific learning disabilities? *Contemporary School Psychology*, 24(4), 445–456. <https://doi.org/10.1007/s40688-020-00274-0>
- Learning Disabilities Association of America. (2010). The Learning Disabilities Association of America's white paper on evaluation, identification, and eligibility criteria for students with specific learning disabilities. <https://ldaamerica.org/wp-content/uploads/2013/10/LDA-White-Paper-on-IDEA-Evaluation-Criteria-for-SLD.pdf>
- Lecerf, T., Bovet-Boone, F., Peiffer, E., Kieng, S., & Geistlich, S. (2016). WISC-IV GAI and CPI profiles in healthy children and children with learning disabilities. *European Review of Applied Psychology*, 66(3), 101–107. <https://doi.org/10.1016/j.erap.2016.04.001>
- Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (2015). *Science and pseudoscience in clinical psychology* (2nd ed., pp. 1–16). Guilford.
- Lo, Y., Mendell, N., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767–778. <https://doi.org/10.1093/biomet/88.3.767>
- Lockwood, A. B., & Farmer, R. L. (2020). The cognitive assessment course: Two decades later. *Psychology in the Schools*, 57(2), 265–283. <https://doi.org/10.1002/pits.22298>
- Loehlin, J. C. (2019). Cognitive clustering—How general? *Intelligence*, 75, 19–22. <https://doi.org/10.1016/j.intell.2019.03.001>
- Loehlin, J. C., Wright, M. J., Hansell, N. K., & Martin, N. G. (2018). Are there distinct cognitive types? *Intelligence*, 70, 7–11. <https://doi.org/10.1016/j.intell.2018.07.002>
- Lubke, G. H., & Luningham, J. (2017). Fitting latent variable mixture models. *Behaviour Research and Therapy*, 98, 91–102. <https://doi.org/10.1016/j.brat.2017.04.003>
- Lubke, G. H., & Miller, P. J. (2015). Does nature have joints worth carving? A discussion of taxometrics, model-based clustering and latent variable mixture modeling. *Psychological Medicine*, 45(4), 705–715. <https://doi.org/10.1017/S003329171400169X>
- Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A comprehensive review. *School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association*, 30(4), 457–469. <https://doi.org/10.1037/spq0000109>
- Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods: Statistical analysis* (pp. 551–611). Oxford University Press.
- Mather, N. (2009). The intelligent testing of children with specific learning disabilities. In J. C. Kaufman (Ed.), *Intelligent testing: Integrating psychological theory and clinical practice* (pp. 30–52). Cambridge University Press.
- McDermott, P. A., Glutting, J. J., Jones, J. N., Watkins, M. W., & Kush, J. (1989). Core profile types in the WISC-R national sample: Structure, membership, and applications. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1(4), 292–299. <https://doi.org/10.1037/1040-3590.1.4.292>
- McFall, R. M. (2000). Elaborate reflections on a simple manifesto. *Applied and Preventive Psychology*, 9(1), 5–21. [https://doi.org/10.1016/S0962-1849\(05\)80035-6](https://doi.org/10.1016/S0962-1849(05)80035-6)
- 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. <https://doi.org/10.1016/j.jsp.2018.10.007>
- McKinney, J. D. (1984). The search for subtypes of specific learning disability. *Journal of Learning Disabilities*, 17(1), 43–50. <https://doi.org/10.1177/002221948401700113>
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley.
- Miciak, J., Taylor, W. P., Stuebing, K. K., & Fletcher, J. M. (2018). Accuracy using a pattern of processing strengths and weaknesses method with multiple measures. *Journal of Psychoeducational Assessment*, 36(1), 21–33. <https://doi.org/10.1177/0734282916683287>
- Miller, J. L., Saklofske, D. H., Weiss, L. G., Drozdick, L., Llorente, A. M., Holdnack, J. A., & Prifitera, A. (2016). Issues related to the WISC-IV assessment of cognitive functioning in clinical and special groups. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (pp. 287–343). Academic Press.
- Morgan, G. B. (2015). Mixed mode latent class analysis: An examination of fit index performance for classification. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 76–86. <https://doi.org/10.1080/10705511.2014.935751>
- Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 39–59. <https://doi.org/10.1080/10705511.2014.919825>
- Morin, A. J. S., McLarnon, M. J. W., & Litalien, D. (2020). Mixture modeling for organizational behavior research. In Y. Griep & S. D. Hansen (Eds.), *Handbook on the temporal dynamics of organizational behavior* (pp. 351–379). Edward Elgar.
- Morin, A. J. S., & Wang, J. C. K. (2016). A gentle introduction to mixture modeling using physical fitness performance data. In N. Ntoumanis & N. D. Myers (Eds.), *An introduction to intermediate and advanced statistical analyses for sport and exercise scientists* (pp. 183–209). Wiley.
- Muthén, L. K., & Muthén, B. O. (1998–2019). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Naglieri, J. A. (2011). The discrepancy/consistency approach to SLD identification using the PASS theory. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of specific learning disability identification* (pp. 145–172). Wiley.
- National Association of School Psychologists. (2010). *Principles for professional ethics*. NASP.
- Niileksela, C. R., & Reynolds, M. R. (2014). Global, broad, or specific cognitive differences? Using a MIMIC model to examine differences in CHC abilities in children with learning disabilities. *Journal of Learning Disabilities*, 47(3), 224–236. <https://doi.org/10.1177/0022219412453172>
- Niileksela, C. R., & Templin, J. (2019). Identifying dyslexia with confirmatory latent profile analysis. *Psychology in the Schools*, 56(3), 335–359. <https://doi.org/10.1002/pits.22183>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary*

- Journal*, 14(4), 535–569. <https://doi.org/10.1080/10705510701575396>
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4), 440–461. <https://doi.org/10.1037/tps0000176>
- Oberski, D. L. (2016). Mixture models: Latent profile and latent class analysis. In J. Robertson & M. Kaptein (Eds.), *Modern statistical methods for HCI* (pp. 275–287). Springer.
- Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 616–639. <https://doi.org/10.1080/10705511.2013.824780>
- 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(4), 376–385. <https://doi.org/10.1037/h0088795>
- Poletti, M., Carretta, E., Bonvicini, L., & Giorgi-Rossi, P. (2018, Jan/Feb). Cognitive clusters in specific learning disorder. *Journal of Learning Disabilities*, 51(1), 32–42. <https://doi.org/10.1177/0022219416678407>
- Rapaport, D., Schafer, R., & Gill, M. M. *Diagnostic psychological testing: The theory, statistical evaluation, and diagnostic application of a battery of tests*. Year Book Publishers.
- Saklofske, D. H., Weiss, L. G., Breaux, K., & Beal, A. L. (2016). WISC-V and the evolving role of intelligence testing in the assessment of learning disabilities. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *WISC-V assessment and interpretation: Scientist-practitioner perspectives* (pp. 237–268). Academic Press.
- Sattler, J. M. (2018). *Assessment of children: Cognitive foundations and applications*. Jerome M. Sattler, Publisher, Inc.
- Schneider, W. J., & Kaufman, A. S. (2017). Let's not do away with comprehensive cognitive assessment just yet. *Archives of Clinical Neuropsychology*, 32(1), 8–20. <https://doi.org/10.1093/arclin/acw104>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Schweizer, K. (2012). On correlated errors. *European Journal of Psychological Assessment*, 28(1), 1–2. <https://doi.org/10.1027/1015-5759/a000094>
- Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52(3), 333–343. <https://doi.org/10.1007/BF02294360>
- Snowling, M. J., & Hulme, C. (2012). Annual research review: The nature and classification of reading disorders—a commentary on proposals for DSM-5. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 53(5), 593–607. <https://doi.org/10.1111/j.1469-7610.2011.02495.x>
- Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools*, 51(10), 1031–1045. <https://doi.org/10.1002/pits.21802>
- Speece, D. L. (1994). Cluster analysis in perspective. *Exceptionality*, 5(1), 31–44. [https://doi.org/10.1207/s15327035ex0501\\_3](https://doi.org/10.1207/s15327035ex0501_3)
- Speece, D. L. (2003). The methods of cluster analysis and the study of learning disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities*. (pp. 501–513). Guilford.
- Styck, K. M., & Watkins, M. W. (2016). Structural validity of the WISC-IV for students with learning disabilities. *Journal of Learning Disabilities*, 49(2), 216–224. <https://doi.org/10.1177/0022219414539565>
- Swanson, H. L., Olide, A. F., & Kong, J. E. (2018). Latent class analysis of children with math difficulties and/or learning disabilities: Are there cognitive differences? *Journal of Educational Psychology*, 110(7), 931–951. <https://doi.org/10.1037/edu0000252>
- Thorndike, R. M., & Thorndike-Christ, T. M. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Pearson.
- 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(2), 402–409. <https://doi.org/10.1177/2167702616672038>
- U.S. Department of Education. (2004). Individuals with Disabilities Education Improvement Act, Pub. L. 108 – 466. (Vol. 70, pp. 77968–77969). Federal Register. Author.
- Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.). Wiley.
- Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, 15(4), 465–479. <https://doi.org/10.1037/h0088802>
- Watkins, M. W. (2009). Errors in diagnostic decision making and clinical judgment. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of school psychology* (4th ed., pp. 210–229). Wiley.
- Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children-Fourth Edition among a national sample of referred students. *Psychological Assessment*, 22(4), 782–787. <https://doi.org/10.1037/a0020043>
- Watkins, M. W., & Canivez, G. L. (2004). Temporal stability of WISC-III subtest composite: Strengths and weaknesses. *Psychological Assessment*, 16(2), 133–138. <https://doi.org/10.1037/1040-3590.16.2.133>
- Watkins, M. W., Canivez, G. L., Dombrowski, S. C., McGill, R. J., Pritchard, A. E., Holingue, C. B., & Jacobson, L. A. (2021). Long-term stability of Wechsler Intelligence Scale for Children—fifth edition scores in a clinical sample. *Applied Neuropsychology: Child*. Advance online publication. <https://doi.org/10.1080/21622965.2021.1875827>
- 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). Guilford.
- Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children—Fourth Edition among referred students. *Educational and Psychological Measurement*, 66(6), 975–983. <https://doi.org/10.1177/0013164406288168>
- Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children—Fourth Edition*. Psychological Corporation.
- Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children—Fourth Edition technical and interpretive manual*. Psychological Corporation.
- Willson, V. L., & Rupley, W. H. (2013). Design for learning disabilities experimental and quasi-experimental research. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 593–606). Guilford.

- Wodarski, J. S., & Hopson, L. M. (2012). *Research methods for evidence-based practice*. Sage.
- Woo, S. E., Jebb, A. T., Tay, L., & Parrigon, S. (2018). Putting the “person” in the center: Review and synthesis of person-centered approaches and methods in organizational science. *Organizational Research Methods, 21*(4), 814–845. <https://doi.org/10.1177/1094428117752467>
- Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M.-L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice, 24*(4), 331–363. <https://doi.org/10.1111/cpsp.12207>
- Zirkel, P. A. (2017). RTI and other approaches to SLD identification under the IDEA: A legal update. *Learning Disability Quarterly, 40*(3), 165–173. <https://doi.org/10.1177/0731948717710778>
- Zumeta, R. O., Zirkel, P. A., & Danielson, L. (2014). Identifying specific learning disabilities: Legislation, regulation, and court decisions. *Topics in Language Disorders, 34*(1), 8–24. <https://doi.org/10.1097/TLD.0000000000000006>

## AUTHOR BIOGRAPHICAL STATEMENTS

**Marley W. Watkins** received his PhD in School Psychology from the University of Nebraska-Lincoln and has held positions with the Deer Valley Unified School District, Pennsylvania State University, Arizona State University, and Baylor University. He is currently Research Professor in the

Department of Educational Psychology at Baylor University. His research interests include professional issues, the psychometrics of assessment and diagnosis, individual differences, and computer applications. Dr. Watkins has published more than 170 peer-reviewed journal articles and made more than 125 presentations at professional conferences.

**Gary L. Canivez, PhD**, is Professor of Psychology at Eastern Illinois University and principally involved in the Specialist in School Psychology program. Before entering academia Dr. Canivez was a School Psychologist for 8 years, was on the Adjunct Faculty of Arizona State University and Northern Arizona University, and was President of the Arizona Association of School Psychologists. Dr. Canivez currently serves as Associate Editor of *Archives of Scientific Psychology* and served as Associate Editor for *Psychological Assessment*. He is a consulting editor and frequent reviewer for numerous other school psychology, assessment, and clinically oriented journals. Dr. Canivez is a member of the Society for the Study of School Psychology, a Fellow of the Division (5) of Quantitative and Qualitative Methods and Division (16) of School Psychology of the American Psychological Association, and a Charter Fellow of the Midwestern Psychological Association. The author of over 100 research and professional publications and over 200 professional presentations and continuing professional development workshops, Dr. Canivez has research interests in psychological assessment and measurement pertaining to intelligence, achievement, personality, and psychopathology; and his research has been supported by the National Institutes of Health/National Institute of Mental Health.