# **Factor Structure of the Differential Ability Scales–Second Edition Core Subtests: Standardization Sample Confirmatory Factor Analyses**

Journal of Psychoeducational Assessment 2020, Vol. 38(7) 791-815 © The Author(s) 2020 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/0734282920914792 journals.sagepub.com/home/jpa



Gary L. Canivez<sup>1</sup>, Ryan J. McGill<sup>2</sup>, and Stefan C. Dombrowski<sup>3</sup>

#### Abstract

The present study examined the factor structure of the Differential Ability Scales–Second Edition (DAS-II) core subtests from the standardization sample via confirmatory factor analysis (CFA) using methods (bifactor modeling and variance partitioning) and procedures (robust model estimation due to nonnormal subtest score distributions) recommended but not included in the DAS-II Introductory and Technical Handbook. CFAs were conducted with the three DAS-II standardization sample age groups (lower early years [age = 2:6-3:5 years], upper early years [age = 3:6-6:1] years], school age [7:0-17:1] years]) using standardization sample raw data provided by NCS Pearson, Inc. Although most DAS-II core subtests were properly associated with the theoretically proposed group factors, both the higher order and bifactor models indicated that the g factor accounted for large portions of total and common variance, whereas the group factors (Verbal, Nonverbal, Spatial) accounted for small portions of total and common variance. The DAS-II core battery provides strong measurement of general intelligence, and clinical interpretation should be primarily, if not exclusively, at that level.

#### **Keywords**

DAS-II, confirmatory factor analysis, higher order structure, bifactor structure, structural validity

The Differential Ability Scales-Second Edition (DAS-II; Elliott, 2007a) is a popular battery of cognitive tests to assess intelligence of children and adolescents aged 2 to 17 years and, although becoming somewhat dated (the norms are now more than 12 years old), is still currently used by practitioners and included in omnibus interpretive systems such as Cross-Battery Assessment. The DAS-II is a revision of the DAS (Elliott, 1990), an adaptation of the British Ability Scales (Elliott et al., 1979) that was standardized for use in the United States.

**Corresponding Author:** Gary L. Canivez, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920, USA. Email: glcanivez@eiu.edu

<sup>&</sup>lt;sup>1</sup>Eastern Illinois University, Charleston, USA <sup>2</sup>William & Mary, Williamsburg, VA, USA <sup>3</sup>Rider University, Lawrenceville, NJ, USA

There are three age-related levels: lower early years (2:6–3:5 years), upper early years (3:6–6:11 years), and school age (7:0–17:11 years), and the three levels contain different configurations of 10 core subtests appropriate for each age. These subtests combine to yield a *General Conceptual Ability* (GCA) score, a higher-order composite score thought to measure psychometric g (Spearman, 1927). There are also three first-order composite scores called *cluster* scores (Verbal Ability [V], Nonverbal Reasoning Ability [NV], and Spatial Ability [SP]) that are hypothesized to reflect more specific and diverse aptitudes. In addition, the DAS-II provides users with nine supplementary subtests across the various age brackets, which contribute to the measurement of three diagnostic cluster scores (Processing Speed, Working Memory, and School Readiness). However, these indicators do not contribute to the measurement of the GCA or the three primary cluster scores and thus were not the focus of the present investigation.

Although the *Introductory and Technical Handbook* (Elliott, 2007b) indicated that the DAS-II development was not driven by a single theory of cognitive ability, the content and structure of the DAS-II were heavily influenced by the Cattell–Horn–Carroll (CHC) model of cognitive abilities (Carroll, 1993, 2003; Cattell & Horn, 1978; Horn, 1991; Schneider & McGrew, 2018). This model also served to guide assessment of DAS-II structural validity and serves as the primary method for score interpretation.

The *Introductory and Technical Handbook* suggests that users should interpret DAS-II scores in a stepwise fashion beginning with the GCA and then proceed to more specific measures (e.g., clusters and subtests). However, Elliott (2007b) suggested that the profile of strengths and weak-nesses generated at the cluster and subtest levels *is of more value* than the information provided by the GCA, especially in cases where considerable variability across the cluster scores is observed and detailed procedures for evaluating scatter among the cluster and subtest scores are outlined in the *Introductory and Technical Handbook*. According to Elliott, "the most satisfactory description of a child's abilities is nearly always at the level of profile analysis" (p. 87). However, such prescriptive statements are rarely justified in applied practice and require adherence to standards of empirical evidence (Marley & Levin, 2011). More recently, McGill et al. (2018) reported on the absence of supportive evidence and negative evidence for such profile analyses since the seminal review by Watkins (2000).

Interpretation of test scores and comparisons must be guided by strong replicated empirical evidence deriving from structural validity, relationships with external variables including incremental validity and diagnostic and treatment utility, as noted in the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014). An important starting point for such evidence resides in the test structure as structural validity is a requisite property of broader construct validity (Keith & Kranzler, 1999).

The DAS-II *Introductory and Technical Handbook* did not report results of exploratory factor analysis (EFA) in examining construct validity nor was there disclosure of proportions of variance accounted for by the higher order *g* factor and the proposed first-order group factors, subtest *g* loadings, subtest specificity estimates, or incremental predictive validity estimates for the factors and subtest scores. Without this information, clinicians are unable to independently determine the relative importance of factor and subtest scores relative to the GCA score. Factor or subtest scores that fail to capture meaningful portions of non-*g* true score variance will likely be of limited clinical utility. The omission of incremental predictive validity results is particularly troubling because users are encouraged to interpret the DAS-II beyond the GCA level but DAS-II cluster scores, like all such scores, conflate general intelligence variance and group factor variance. Youngstrom et al. (1999) examined the incremental validity of the original DAS and found that interpretation beyond the GCA was not supported.

#### Structural Validity Investigations of the DAS-II

Confirmatory factor analyses (CFAs) of the DAS-II hierarchical structure were reported in the DAS-II *Introductory and Technical Handbook* (Elliott, 2007b), and figures 8.1, 8.2, 8.3, and 8.4 illustrate the standardized validation models for the seven core and diagnostic subtests (2:6–3:5 years) featuring two first-order factors, 11 core and diagnostic subtests (4:0–5:11 years) with five first-order factors, 14 core and diagnostic subtests (6:0–12:11 years) with seven first-order factors, and 12 core and diagnostic subtests featuring six first-order factors, respectively. In these models, several first-order factors not available in the actual DAS-II were specified (e.g., auditory processing, visual–verbal memory, and verbal short-term memory). In addition, the auditory processing and visual–verbal memory factors in the final validation models for ages 6 to 17 years were each produced from a single indicator variables is possible in CFA, variables assessed by a single measure should not be interpreted as factors due to the fact that they do not possess any shared variance from multiple indicators (Brown, 2015).

Keith et al. (2010) examined measurement invariance of the DAS-II core and diagnostic subtest structure and reported support for a six-factor hierarchical model that corresponded closely with CHC theory with general intelligence at the apex. However, the final validation model required the specification of a cross-loading for the Verbal Comprehension measure on Crystallized Ability and Fluid Reasoning factors. Although Keith et al. provided the results of residualized subtest factor loadings in their DAS-II CFA analyses, the clinical utility of these results are limited due to the fact they were derived from a hypothesized first-order latent structure that deviates significantly from the structure suggested in the *Introductory and Technical Handbook* (Elliott, 2007b). Neither Elliott (2007b) nor Keith et al. reported univariate or multivariate skewness or kurtosis estimates among the scales used as indicators in their CFA models, which could have implications for proper model estimation if data were nonnormally distributed.<sup>1</sup> Also missing from CFAs conducted by Elliott and Keith et al. were comparisons of rival bifactor structures as explanations of DAS-II.

Until recently, independent factor analytic investigations of the DAS-II, as well as the validation study results reported in the *Introductory and Technical Handbook*, relied singularly on application of CFA procedures to various configurations of the core and supplementary subtests to produce different hierarchical models consistent with CHC theory. However, use of these results to ascertain what the core battery measures is problematic as those models are not structurally equivalent (Cattell, 1978). In recognition of these limitations, Canivez and McGill (2016) conducted hierarchical EFA and variance decomposition using the Schmid and Leiman (1957) procedure. Canivez and McGill found that although the DAS-II core subtests measured the general intelligence dimension (estimated by the GCA) well, as evidenced by high omegahierarchical ( $\omega_{\rm H}$ ) coefficients, the DAS-II group factors (V, NV, SP) did *not* contribute sufficient portions of unique variance as evidenced by low and inadequate omega-hierarchical subscale ( $\omega_{\rm HS}$ ) coefficients. These results suggest that clinical interpretation of the DAS-II should likely be restricted to the GCA level and any interpretation of other scores or comparisons beyond the GCA should be done with caution and in light of additional external validation evidence.

To further elaborate on the DAS-II structure, Dombrowski, McGill, Canivez, and Peterson (2019) utilized similar EFA procedures to examine the total battery using the standardization sample data from the 5- to 8-year-old age range to determine the degree to which the DAS-II theoretical structure proposed in the *Introductory and Technical Handbook*, later refined by Keith et al. (2010), could be replicated. Results suggested a six-factor solution that was generally consistent with the CHC-based structure suggested by the publisher, with desired simple structure

attained. However, two subtests (Picture Similarities and Early Number Concepts) did not saliently load on any group factors. Dombrowski, McGill, and Morgan (2019) used Monte Carlo simulation by resampling the standardization sample correlation matrices 1,000 times and analyzed the structure of the DAS-II total battery using maximum likelihood CFA. Both studies generally supported the theoretical structure posited in the *Introductory and Technical Handbook*. However, like Canivez and McGill (2016), large portions of subtest variance across both studies were apportioned to the general intelligence dimension resulting in a large  $\omega_{\rm H}$  coefficient, but small portions of unique variance were apportioned to the first-order group factors as evidenced by the small  $\omega_{\rm HS}$  coefficients, indicating inadequate unique measurement by the group factors.

Alternatively, Dombrowski, Golay, et al. (2018) used Bayesian structural equation modeling (BSEM) to examine the latent factor structure from the DAS-II core subtests using the 7 to 17 years age range standardization sample of raw data. This allowed the estimation of small, non-zero parameters often set to zero in traditional CFA that can inflate factor covariances and potentially distort model results. Results revealed the plausibility of the hypothesized three-factor model, consistent with publisher theory, expressed as either a higher order (HO) or a bifactor (BF; Holzinger & Swineford, 1937) model. However, best BSEM model fit was obtained from an alternative structure, a two-group factor (V, SP) bifactor (BF) model with Matrices (MAT) and Sequential and Quantitative Reasoning (SQR) loading on g only and no NV group factor. As with Canivez and McGill (2016); Dombrowski, McGill, Canivez, and Peterson (2019); and Dombrowski, McGill, and Morgan (2019), the general intelligence factor dominated subtest variance and had high omega-hierarchical ( $\omega_{\rm H}$ ) coefficients, but the DAS-II group factors (V, NV, SP) did not contribute sufficient portions of unique variance as shown by low and inadequate omega-hierarchical subscale ( $\omega_{\rm HS}$ ) coefficients.

Although the latent structure of the DAS-II core subtests was examined via CFA in the *Introductory and Technical Handbook*, the standardized solutions for these analyses were not provided and alternative, rival models (e.g., bifactor) were not evaluated. Inspection of the goodness-of-fit results in table 8.4 indicated that a three-factor hierarchical model provided the most optimal solution for the core subtests across the age spans with fairly robust improvements in fit when compared with competing one-factor and hierarchical two-factor models. Generally, CFAs have supported a hierarchical model with general intelligence at the apex and three first-order factors for the core subtests, but bifactor structure has thus far only been examined by Dombrowski, Golay, et al. (2018) using a recently rediscovered approach to latent variable modeling and those results did not support a three-factor structure for the core tests.

#### Purpose of the Current Study

CFAs for the DAS-II core subtests reported in the *Introductory and Technical Handbook* are not sufficiently explicated, did not recognize or account for nonnormal distributions, and did not disclose portions of variance accounted for by the various factors. Furthermore, Elliott did not provide standardized parameter estimates for the core subtest models or examine rival bifactor model representations, which might provide better and more parsimonious fit to the standardization sample data. Accordingly, the purpose of the present investigation was to extend the results of the Canivez and McGill (2016) EFA study and to address the limitations of CFA reported in the DAS-II *Introductory and Technical Handbook*. Specifically, the present study examined the factor structure of the DAS-II core subtests through CFA and disclosure of variance contributions of latent factors using the normative sample raw data across the three test levels (i.e., lower early years, upper early years, and school age). It is believed that the results furnished by the present investigation will be instructive for determining how the DAS-II core battery should be interpreted in clinical practice.

### Method

#### Participants

Participants were members of the DAS-II standardization sample and included a total of 3,460 individuals ranging in age from 2 to 17 years. Age groups included lower early years (2:6- to 3:5-year-olds; N = 352), upper early years (3:6- to 6:11-year-olds; N = 920), and school age (7:0- to 17:11-year-olds; N = 2,188). Detailed demographic characteristics are provided in the DAS-II *Introductory and Technical Handbook* (Elliott, 2007b). The standardization sample was obtained using stratified proportional sampling across key demographic variables of age, sex, race/ethnicity, parent educational level, and geographic region; and examination of the demographic results reported in the *Introductory and Technical Handbook* reveal a close correspondence across the stratification variables to the October 2002 U.S. census estimates.

Table S1 (see supplemental material) presents DAS-II Core Subtest correlation matrices and descriptive statistics for the three DAS-II age groups indicating some departure from normal distribution (Onwuegbuzie & Daniel, 2002; West et al., 1995). Univariate skewness estimates from the three age groups ranged from -0.911 to 0.733. Univariate kurtosis estimates from the three age groups, however, ranged from 0.556 to 3.195. Mardia's (1970) multivariate kurtosis estimates for the lower early years age 2:6 to 3:5 sample (Z = 11.37), upper early years age 3:6 to 6:11 sample (Z = 26.49), and the school age 7:0 to 17:11 sample (Z = 33.86) indicated statistically significant (p < .05) multivariate nonnormality for all three age groups (Cain et al., 2017) that has implications for CFA model estimation and fit statistics.

#### Instrument

The DAS-II uses different combinations of the 10 core subtests to produce the GCA score at different points of the age span. Whereas the GCA score is composed of four subtests at ages 2:6 through 3:5 years, six core subtests are used from ages 3:6 through 17:11 years. The core subtests combine to form three primary cognitive clusters at the first-order level and each is composed of two subtests. Verbal (V) ability and Nonverbal (NV) Reasoning Ability clusters are provided for all ages, but an additional Spatial (SP) Ability cluster is available from ages 3:6 through 17:11 years. Additional combinations of supplemental diagnostic subtests are provided, which can be combined to yield additional first-order clusters (e.g., Working Memory, Processing Speed, and School Readiness); however, these measures are not utilized to calculate the higher order GCA composite or its lower order cognitive clusters. In addition, the diagnostic measures cannot be used to substitute for the core subtests.

#### Procedure and Analyses

NCS Pearson, Inc. provided standardization sample raw data for independent analyses. EFA models suggested by Canivez and McGill (2016) and those promoted by Elliott (2007b) and the publisher (see table 8.4) were examined and compared. Whereas Elliott only reported oblique models for the core subtests, the present study examined oblique, higher order, *and* bifactor structures to determine fit to these data.

CFAs with maximum likelihood estimation were conducted using EQS 6.3 (Bentler & Wu, 2016). Each of the three latent group factors produced by DAS-II core subtests (V, NV, SP) have only two observed indicators and thus are empirically underidentified. Consequently, to ensure identification of CFA bifactor models, those subtests were constrained to equality (Little et al., 1999). Given the significant multivariate kurtosis observed in all three age groups, robust maximum likelihood estimation with the Satorra and Bentler (S-B; 2001) corrected chi-square was

applied. Byrne (2006) indicated "the S-B  $\chi^2$  has been shown to be the most reliable test statistic for evaluating mean and covariance structure models under various distributions and sample sizes" (p. 138). Because Elliott (2007b) did not disclose univariate or multivariate normality estimates or apply the S-B–corrected  $\chi^2$ , present results may differ from those presented in the *Introductory and Technical Handbook*. Some models reported in table 8.4 included cross-loading the Picture Similarities and Matrices subtests on multiple factors, but these were not presently examined given problems of cross-loading and its abandoning of desired simple structure. It should be noted that previous EFA studies (Canivez & McGill, 2016; Dombrowski, Golay, et al., 2018; Dombrowski, McGill, Canivez, & Peterson, 2019) *did not* support specification of this parameter, whereas Keith et al. (2010) did not include cross-loadings of Picture Similarities and Matrices subtests in initial calibration, reference variable, or final validation CFA models. Furthermore, those parameters deviate from the theoretical structure of the test, which is based upon desired simple structure.

Given that the large sample size could unduly influence the  $\chi^2$  value (Kline, 2016), approximate fit indices were used to aid model evaluation and selection. Although criterion values for approximate fit indices are not universally accepted (McDonald, 2010), the comparative fit index (CFI), Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA) were used to evaluate overall global model fit. Higher values indicated better fit for the CFI and TLI, whereas lower values indicated better fit for the RMSEA. Combinatorial heuristics of Hu and Bentler (1999) were applied where CFI and TLI  $\geq$  .90 and RMSEA  $\leq$  .08 were criteria for adequate model fit; whereas CFI and TLI  $\geq$  .95 and RMSEA  $\leq$  .06 were criteria for well-fitting models. Marsh et al. (2004), however, cautioned overgeneralizing such heuristics that could result in the incorrect rejection of an acceptable model (Type I error). The Akaike Information Criterion (AIC) was also considered. Because AIC does not have a meaningful scale, the model with the smallest AIC value was preferred as most likely to replicate (Kline, 2016). Superior models required adequate to good overall fit and indication of meaningfully better fit ( $\Delta CFI >$ .01,  $\Delta RMSEA > .015$ ,  $\Delta AIC > 10$ ) than alternative models (Burnham & Anderson, 2004; Chen, 2007; Cheung & Rensvold, 2002). Local fit was also considered in addition to global fit as models should never be retained "solely on global fit testing" (Kline, 2016, p. 461). Statistical power sufficient to detect even small differences is provided by the large sample size as well as more precise model parameter estimates.

Coefficients omega-hierarchical ( $\omega_{H}$ ) and omega-hierarchical subscale ( $\omega_{HS}$ ) were estimated and provide a model-based estimate of the proportion of true score variance that would be obtained in a unit-weighted score composed of subtests associated with a specific factor (Reise, 2012; Rodriguez et al., 2016a, 2016b; Watkins, 2017). The  $\omega_{\rm H}$  coefficient is the unique general intelligence factor variability estimate with variability from the group factors removed. The  $\omega_{HS}$ coefficient is the unique group factor variability estimate with variability from all other group and general factors removed (Brunner et al., 2012; Reise, 2012). Omega estimates ( $\omega_{\rm H}$  and  $\omega_{\rm HS}$ ) are calculated from CFA bifactor solutions or decomposed variance estimates from higher order models and were obtained using the Omega program (Watkins, 2013), which is based on the works of Zinbarg et al. (2005, 2006) and the Brunner et al. (2012) tutorial. Although standards for omega coefficients acceptability for clinical use are not universally accepted, it has been suggested that  $\omega_{H}$  and  $\omega_{HS}$  coefficients should exceed .50, but .75 might be preferred (Reise, 2012; Reise et al., 2013; Rodriguez et al., 2016a, 2016b). Reise et al. (2013) and Rodriguez et al. (2016a, 2016b) illustrated meaningful attribution of indicators to latent general or group factor measurement when the majority of unique variability was present in the factor and thus the minimum criterion of .50. The Hancock and Mueller (2001) construct reliability or construct replicability coefficient (H) supplemented omega coefficients and estimated the latent construct adequacy represented by the indicators, using a criterion value of .70 (Hancock & Mueller, 2001; Rodriguez et al., 2016). H coefficients were produced by the Omega program (Watkins, 2013).

### Results

### Lower Early Years (Age = 2:6-3:5)

Table 1 presents fit statistics for the only two models that could be tested. Both unidimensional *g* (Model 1, see Figure 1) and the oblique V and NV (Model 2, see Figure 2) models fit the standardization data well. No statistically significant or meaningful differences between these two models were noted in fit statistics, so given the extremely high V–NV covariance (.936), the unidimensional *g* model was determined the best representation for parsimonious explanation of DAS-II measurement for this age group. Higher-order and bifactor models would be mathematically equivalent to Model 2.

### Upper Early Years (Age = 3:6-6:11)

Table 1 presents fit statistics for models tested for the 3:6 to 6:11 age group. The combinatorial heuristics of Hu and Bentler (1999) indicated that Model 1 (g) was inadequate with too low TLI and too high RMSEA. Model 2 (oblique V and NV) provided adequate fit to standardization sample data, but Model 3 (oblique V, NV, SP) fit the standardization sample data well and better than Model 2 (higher TLI and CFI and lower RMSEA and AIC). However, given the significant covariance among the three group factors, it was necessary to explicate higherorder and bifactor representations of that model. Model 4 (higher-order with V and NV) produced adequate to good fit but contained a local fit problem of linear dependence of the lower-order Verbal factor disturbance that needed to be fixed to zero to allow model estimation. Model 5a (bifactor with V and NV) provided better fit than Model 4 (higher TLI and CFI and lower RMSEA and AIC), but Matrices and Picture Similarities subtests had negative standardized path coefficients on the NV group factor so was respecified as Model 5b (as per Dombrowski, Golay, et al., 2018) with Matrices (MAT) and Picture Similarities (PS) subtests containing only standardized g path coefficients and no NV group factor standardized path coefficients. Due to only having two indicators per group factor, Model 6 (higher-order [see Figure 3] and bifactor [see Figure 4] representations with V, NV, and SP) was mathematically equivalent and both provided good fits to standardization sample data, and neither produced local fit problems. As such, both higher-order (Figure 3) and bifactor (Figure 4) representations of Model 6 are further explicated in Tables 2 and 3 to illustrate decomposed sources of variance and model-based reliability estimates.

The general intelligence dimension accounted for most of the DAS-II subtest variance and substantially smaller portions of subtest variance were uniquely associated with the three DAS-II group factors (V, NV, SP). Omega-hierarchical and omega-hierarchical subscale coefficients estimated using bifactor results from Table 4 found the  $\omega_{\rm H}$  coefficient for general intelligence (.748) was high and indicated a unit-weighted composite score based on the six subtest indicators would account for 74.8% true score variance. The  $\omega_{HS}$  coefficients for the three DAS-II group factors (V, NV, SP) were considerably lower ranging from .072 (NV) to .210 (V). Thus, unit-weighted composite scores for the three DAS-II first-order factors possess too little unique true score variance to recommend confident clinical interpretation (Reise, 2012; Reise et al., 2013). Table 4 also presents H coefficients that reflect correlations between the latent factors and optimally weighted composite scores (Rodriguez et al., 2016). The H coefficient for the general factor (.785) indicated the general factor was well defined by the six DAS-II subtest indicators, but the H coefficients for the three DAS-II group factors ranged from .096 to .281 and thus were not adequately defined by their subtest indicators. Results were identical or nearly identical for the higher-order representation of DAS-II (see Table 3).

| Measurement Models                                      | $S\text{-}B~\chi^2$ | df | Þ     | TLI   | CFI   | RMSEA | RMSEA 90% CI   | AIC       |
|---|---------------------|----|-------|-------|-------|-------|----------------|-----------|
| Age = 2:6-3:5 (N = 352)                                 |                     |    |       |       |       |       |                |           |
| l One factor (g)  | 1.94                | 2  | .3786 | 1.000 | 1.000 | .000  | [0.000, 0.105] | 9,992.94  |
| 2 Two oblique factors (V, NV)                           | 0.66                | T  | .4170 | 1.000 | 1.000 | .000  | [0.000, 0.131] | 9,993.71  |
| Age = 3:6-6:11 (N = 920)                                |                     |    |       |       |       |       |                |           |
| l One factor (g)  | 91.16               | 9  | .0001 | .835  | .901  | .100  | [0.081, 0.118] | 39,670.84 |
| 2 Two oblique factors (V, NV)                           | 36.33               | 8  | .0001 | .936  | .966  | .062  | [0.042, 0.083] | 39,608.50 |
| 3 Three oblique factors (V, NV, SP)                     | 4.36                | 6  | .6287 | 1.000 | 1.000 | .000  | [0.000, 0.036] | 39,573.54 |
| 4ª Higher order (V, NV)                                 | 31.78               | 7  | .0001 | .936  | .970  | .062  | [0.041, 0.085] | 39,610.50 |
| 5a <sup>b</sup> Bifactor (V, NV)                        | 3.61                | 4  | .4616 | 1.000 | 1.000 | .000  | [0.000, 0.048] | 39,578.87 |
| 5b <sup>c</sup> Bifactor (V, NV)                        | 6.76                | 7  | .4544 | 1.000 | 1.000 | .000  | [0.000, 0.040] | 39,578.51 |
| 6 <sup>d</sup> Higher order and bifactor                | 4.36                | 6  | .6287 | 1.000 | 1.000 | .000  | [0.000, 0.036] | 39,573.54 |
| (V, NV, SP)   |                     |    |       |       |       |       |                |           |
| Age = 7:0-17:11 (N = 2,188)                             |                     |    |       |       |       |       |                |           |
| l One factor (g)  | 326.63              | 9  | .0001 | .781  | .868  | .127  | [0.115, 0.139] | 92,378.96 |
| 2 Two oblique factors (V, NV)                           | 76.30               | 8  | .0001 | .939  | .968  | .062  | [0.050, 0.075] | 92,821.48 |
| 3 Three oblique factors (V, NV, SP)                     | 9.03                | 6  | .1721 | .997  | .999  | .015  | [0.000, 0.034] | 92,036.27 |
| 4 <sup>e</sup> Higher order (V, NV)                     | 51.90               | 7  | .0001 | .960  | .981  | .054  | [0.041, 0.068] | 92,088.29 |
| 5 <sup>f</sup> Bifactor (V, NV)                         | 2.92                | 4  | .5710 | 1.000 | 1.000 | .000  | [0.000, 0.028] | 92,035.95 |
| 6 <sup>g</sup> Higher order and bifactor (V,<br>NV, SP) | 9.03                | 6  | .1716 | .997  | .999  | .015  | [0.000, 0.034] | 92,036.29 |
| 7 <sup>h</sup> Bifactor (V, SP)                         | 8.97                | 7  | .2552 | .998  | .999  | .011  | [0.000, 0.030] | 92,038.29 |

Table 1. CFA Fit Statistics for DAS-II Core Subtests for the Total Standardization Samples.

Note. CFA = confirmatory factor analysis; DAS-II = Differential Ability Scales-Second Edition; S-B = Satorra-Bentler; TLI =

Tucker–Lewis index; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; AIC = Akaike's Information Criterion; g = general intelligence; V = Verbal; NV = Nonverbal; SP = Spatial.

Bold text reflects best and preferred model.

<sup>a</sup>Factor 2 (Verbal) disturbance was linearly dependent on other parameters so EQS set disturbance variance to zero for model estimation. <sup>b</sup>Matrices and Picture Similarities subtests had negative path coefficients on the NV group factor.

<sup>c</sup>Model respecified with Matrices and Picture Similarities subtests with only g paths and no NV group factor paths.

<sup>d</sup>Higher order model AIC presented in the table, bifactor model AIC was slightly higher at 39,579.54 but not meaningfully different. <sup>e</sup>EQS condition code noted the NV and g factors were linearly dependent on other parameters.

<sup>f</sup>Matrices and Sequential and Quantitative Reasoning subtests had small negative path coefficients (-.01 and -.05, respectively) on NV group factor.

<sup>g</sup>Higher order model AIC presented in the table, bifactor model AIC was slightly higher at 92,042.29 54 but not meaningfully different than higher order model, and Matrices and Sequential and Quantitative Reasoning group factor standardized path values were 0 and thus not statistically significant.

<sup>h</sup>Bifactor model respecified with Matrices and Sequential and Quantitative Reasoning subtests having only g paths and no NV group factor paths (equivalent to removing Matrices and Sequential and Quantitative Reasoning subtests NV group factor paths in Model 5).



**Figure 1.** Unidimensional measurement model with standardized coefficients, for the 4 DAS-II core subtests for ages 2:6 to 3:5, N = 352.

All standardized coefficients are statistically significant, p < .05.



**Figure 2.** Two oblique factors measurement model with standardized coefficients, for the four DAS-II core subtests for ages 2:6 to 3:5, N = 352.

All standardized coefficients are statistically significant, p < .05.



**Figure 3.** Higher-order measurement model with standardized coefficients, for the six DAS-II core subtests for ages 3:6 to 6:11, N = 920.

Note. DAS-II = Differential Ability Scales–Second Edition.

All standardized coefficients are statistically significant, p < .05.

## School Age (Age = 7:0-17:11)

Table 1 presents fit statistics for tested models for the 7:0 to 17:11 age group. Examination of fit statistics indicated that Model 1 (g) was inadequate (too low TLI and CFA, too high RMSEA). Model 2 (oblique V and NV) provided adequate to good fit but Model 3 (oblique V, NV, SP) fit the standardization sample data well and was superior to Models 1 and 2 (higher TLI and CFI and lower RMSEA and AIC). Due to significant covariance of the three group factors (V, NV, SP), higher-order and bifactor models were necessary. Model 4 (higher-order with V and NV)



**Figure 4.** Bifactor measurement model with standardized coefficients, for the six DAS-II core subtests for ages 3:6 to 6:11, N = 920.

Note. DAS-II = Differential Ability Scales–Second Edition.

All standardized coefficients are statistically significant, p < .05.

|                                 | General |                | Verbal |                | Nonverbal |                | Spatial |                |      |                |      |
|---------------------------------|---------|----------------|--------|----------------|-----------|----------------|---------|----------------|------|----------------|------|
| DAS-II subtest                  | Ь       | S <sup>2</sup> | Ь      | S <sup>2</sup> | Ь         | S <sup>2</sup> | Ь       | S <sup>2</sup> | h²   | u <sup>2</sup> | ECV  |
| Verbal Comprehension            | .627    | .393           | .405   | .164           |           |                |         |                | .557 | .443           | .706 |
| Naming Vocabulary               | .623    | .388           | .402   | .162           |           |                |         |                | .550 | .450           | .706 |
| Picture Similarities            | .641    | .411           |        |                | .242      | .059           |         |                | .469 | .531           | .875 |
| Matrices                        | .542    | .294           |        |                | .205      | .042           |         |                | .336 | .664           | .875 |
| Pattern Construction            | .670    | .449           |        |                |           |                | .437    | .191           | .640 | .360           | .702 |
| Copying                         | .550    | .303           |        |                |           |                | .359    | .129           | .431 | .569           | .701 |
| Total variance                  |         | .373           |        | .054           |           | .017           |         | .052           | .497 | .503           |      |
| ECV                             |         | .750           |        | .109           |           | .034           |         | .107           |      |                |      |
| ω                               |         | .831           |        | .713           |           | .572           |         | .696           |      |                |      |
| ω <sub>H</sub> /ω <sub>HS</sub> |         | .748           |        | .210           |           | .072           |         | .208           |      |                |      |
| Relative $\omega$               |         | .900           |        | .294           |           | .125           |         | .299           |      |                |      |
| Factor correlation              |         | .865           |        | .458           |           | .267           |         | .456           |      |                |      |
| Н                               |         | .785           |        | .280           |           | .096           |         | .277           |      |                |      |
| PUC                             |         | .800           |        |                |           |                |         |                |      |                |      |

**Table 2.** Sources of Variance in the DAS-II Core Subtests for the Total Standardization Sample Ages 3:6 to 6:11 (N = 920) According to a CFA Higher-Order Model (Figure 3).

Note. DAS-II = Differential Ability Scales–Second Edition; CFA = confirmatory factor analysis; b = standardized loading of subtest on factor;  $S^2$  = variance explained in the subtest;  $h^2$  = communality;  $u^2$  = uniqueness; ECV = explained common variance;  $\omega$  = omega;  $\omega_H$  = omega-hierarchical (general factor);  $\omega_{HS}$  = omega-hierarchical subscale (group factors); H = construct reliability or replicability index; PUC = percentage of uncontaminated correlations.

|                                 | General |                | Verbal |                       | Nonv | erbal          | Spatial |                       |      |                |      |
|---------------------------------|---------|----------------|--------|-----------------------|------|----------------|---------|-----------------------|------|----------------|------|
| DAS-II subtest                  | Ь       | S <sup>2</sup> | Ь      | <b>S</b> <sup>2</sup> | Ь    | S <sup>2</sup> | Ь       | <b>S</b> <sup>2</sup> | h²   | u <sup>2</sup> | ECV  |
| Verbal Comprehension            | .627    | .393           | .400   | .160                  |      |                |         |                       | .553 | .447           | .711 |
| Naming Vocabulary               | .623    | .388           | .408   | .166                  |      |                |         |                       | .555 | .445           | .700 |
| Picture Similarities            | .641    | .411           |        |                       | .226 | .051           |         |                       | .462 | .538           | .889 |
| Matrices                        | .542    | .294           |        |                       | .222 | .049           |         |                       | .343 | .657           | .856 |
| Pattern Construction            | .670    | .449           |        |                       |      |                | .392    | .154                  | .603 | .397           | .745 |
| Copying                         | .550    | .303           |        |                       |      |                | .399    | .159                  | .462 | .538           | .655 |
| Total variance                  |         | .373           |        | .054                  |      | .017           |         | .052                  | .496 | .504           |      |
| ECV                             |         | .752           |        | .110                  |      | .034           |         | .105                  |      |                |      |
| ω                               |         | .831           |        | .713                  |      | .572           |         | .693                  |      |                |      |
| ω <sub>H</sub> /ω <sub>HS</sub> |         | .748           |        | .210                  |      | .072           |         | .205                  |      |                |      |
| Relative $\omega$               |         | .900           |        | .295                  |      | .125           |         | .296                  |      |                |      |
| Factor correlation              |         | .865           |        | .458                  |      | .268           |         | .453                  |      |                |      |
| Н                               |         | .785           |        | .281                  |      | .096           |         | .271                  |      |                |      |
| PUC                             |         | .800           |        |                       |      |                |         |                       |      |                |      |

**Table 3.** Sources of Variance in the DAS-II Core Subtests for the Total Standardization Sample Ages 3:6 to 6:11 (N = 920) According to a CFA Bifactor Model (Figure 4).

Note. DAS-II = Differential Ability Scales–Second Edition; CFA = confirmatory factor analysis; b = standardized loading of subtest on factor;  $S^2$  = variance explained in the subtest;  $h^2$  = communality;  $u^2$  = uniqueness; ECV = explained common variance;  $\omega$  = omega;  $\omega_H$  = omega-hierarchical (general factor);  $\omega_{HS}$  = omega-hierarchical subscale (group factors); H = construct reliability or replicability index; PUC = percentage of uncontaminated correlations.

|  | General          |      | Verbal           |      | Nonverbal |                | Spatial |                |      |                |      |
|--|------------------|------|------------------|------|-----------|----------------|---------|----------------|------|----------------|------|
| DAS-II subtest                           | b S <sup>2</sup> |      | b S <sup>2</sup> |      | Ь         | S <sup>2</sup> | Ь       | S <sup>2</sup> | h²   | u <sup>2</sup> | ECV  |
| Word Definitions                         | .654             | .428 | .469             | .220 |           |                |         |                | .648 | .352           | .660 |
| Verbal Similarities                      | .662             | .438 | .473             | .224 |           |                |         |                | .662 | .338           | .662 |
| Matrices                                 | .772             | .596 |                  |      | .004      | .000           |         |                | .596 | .404           | .999 |
| Sequential and Quantitative<br>Reasoning | .820             | .672 |                  |      | .024      | .001           |         |                | .673 | .327           | .999 |
| Pattern Construction                     | .705             | .497 |                  |      |           |                | .334    | .112           | .609 | .391           | .817 |
| Recall of Designs                        | .638             | .407 |                  |      |           |                | .302    | .091           | .498 | .502           | .817 |
| Total variance                           |                  | .506 |                  | .074 |           | .000           |         | .034           | .614 | .386           |      |
| ECV                                      |                  | .824 |                  | .120 |           | .000           |         | .055           |      |                |      |
| ω  |                  | .893 |                  | .791 |           | .776           |         | .712           |      |                |      |
| $\omega_{H}/\omega_{HS}$                 |                  | .834 |                  | .268 |           | .000           |         | .130           |      |                |      |
| Relative $\omega$                        |                  | .933 |                  | .339 |           | .000           |         | .183           |      |                |      |
| Factor correlation                       |                  | .913 |                  | .518 |           | .015           |         | .361           |      |                |      |
| Н  |                  | .871 |                  | .363 |           | .001           |         | .184           |      |                |      |
| PUC                                      |                  | .800 |                  |      |           |                |         |                |      |                |      |

**Table 4.** Sources of Variance in the DAS-II Core Subtests for the Total Standardization Sample Ages 7:0 to 17:11 (N = 2,188) According to a CFA Higher-Order Model (Figure 5).

Note. DAS-II = Differential Ability Scales–Second Edition; CFA = confirmatory factor analysis; b = standardized loading of subtest on factor;  $S^2$  = variance explained in the subtest;  $h^2$  = communality;  $u^2$  = uniqueness; ECV = explained common variance;  $\omega$  = omega;  $\omega_H$  = omega-hierarchical (general factor);  $\omega_{HS}$  = omega-hierarchical subscale (group factors); H = construct reliability or replicability index; PUC = percentage of uncontaminated correlations.



**Figure 5.** Higher-order measurement model with standardized coefficients, for the six DAS-II core subtests for ages 7:0 to 17:11, N = 2,188. *Note.* DAS-II = Differential Ability Scales–Second Edition.

All standardized coefficients are statistically significant, p < .05.

produced good fit but contained a local fit problem where the NV and g factors were linearly dependent on other parameters. Model 5 (bifactor with V and NV) fit the standardization sample data well and not only was superior to Model 4 (higher TLI and CFI and lower RMSEA and AIC) but also contained local fit problems where Matrices (MAT) and Sequential and Quantitative Reasoning (SQR) subtests had small negative path coefficients (-.01 and -.05, respectively) on the NV group factor. Due to only having two indicators per group factor, Model 6 (higher-order [see Figure 5] and bifactor [see Figure 6] representations with V, NV, and SP) were mathematically equivalent and provided good fit to standardization sample data. The higher-order version of Model 6 contained a local fit problem of a standardized path coefficient of 1.0 between g and NV (see Figure 5) and the bifactor version of Model 6 contained a local fit problem of standardized path coefficients of 0 between the NV group factor and MAT and SQR subtests. As a result, the NV group factor was deleted from the bifactor model and Model 7 (see Figure 7) estimated to represent a bifactor model with only the V and SP group factors and MAT and SQR subtests contained only standardized path coefficients with g. Both higher-order (Figure 5) and bifactor (Figures 6 and 7) representations of Models 6 and 7 are further explicated in Table 4 (higherorder) and Table 5 (bifactor) to illustrate decomposed sources of variance and model-based validity estimates.

In both the higher-order (Model 5) and bifactor (Model 6) models, the general intelligence dimension accounted for most of the DAS-II subtest variance and substantially smaller portions of subtest variance were uniquely associated with the three DAS-II group factors (V, NV, SP). Omega-hierarchical and omega-hierarchical subscale coefficients estimated using bifactor results from Table 5 found the  $\omega_{\rm H}$  coefficient for general intelligence (.834) was high, and indicated a



**Figure 6.** Bifactor measurement model with standardized coefficients, for the six DAS-II core subtests for ages 7:0 to 17:11, N = 2,188.

*Note*. DAS-II = Differential Ability Scales–Second Edition.

All standardized coefficients are statistically significant, p < .05 except nonverbal to Matrices and Nonverbal to Sequential and Quantitative Reasoning paths.

unit-weighted composite score based on the six subtest indicators would account for 83.4% true score variance. The  $\omega_{\text{HS}}$  coefficients for the three DAS-II group factors (V, NV, SP) were considerably lower ranging from .000 (NV) to .268 (V). Thus, unit-weighted composite scores for the three DAS-II first-order factors possess too little unique true score variance to recommend clinical interpretation (Reise, 2012; Reise et al., 2013). Table 5 also presents *H* coefficients that reflect correlations between the latent factors and optimally weighted composite scores (Rodriguez et al., 2016). The *H* coefficient for the general factor (.871) indicated the general factor was well defined by the six DAS-II subtest indicators and essentially unidimensional, but the *H* coefficients for the three group factors ranged from .000 to .365 and thus were not adequately defined by their subtest indicators. Results were nearly identical for the higher-order representation of DAS-II (see Table 4).

### Discussion

The present study provided an independent analysis of the factor structure of the DAS-II core subtests with the three age groups in the standardization sample using best practice CFA methods. The DAS-II *Introductory and Technical Handbook* (Elliott, 2007b) does not report the CFA procedures and analyses necessary to adequately support reported construct validity. Lack of disclosure of univariate and multivariate nonnormality among DAS-II core subtests in the standardization sample and apparent lack of robust model estimation in CFA reported in the DAS-II *Introductory and Technical Handbook* resulted in misestimation of model fit statistics



**Figure 7.** Final bifactor measurement model with standardized coefficients, for the six DAS-II core subtests for ages 7:0 to 17:11, N = 2,188.

Note. DAS-II = Differential Ability Scales–Second Edition. All standardized coefficients are statistically significant, p < .05.

| Table  | 5.   | Sources | of V  | ariance | in the  | DAS-II | Core   | Subtest | s for  | the  | Total  | Standar | dization | Sample | Ages |
|--------|------|---------|-------|---------|---------|--------|--------|---------|--------|------|--------|---------|----------|--------|------|
| 7:0 to | 17:1 | (N =    | 2,188 | 3) Acco | rding t | o a CF | A Bifa | ctor Mo | del (F | igur | re 6). |         |          |        |      |

|                                       | Gen  | eral                  | Verbal |                       | Nonverbal |                | Spatial |                       |      |                |      |
|---------------------------------------|------|-----------------------|--------|-----------------------|-----------|----------------|---------|-----------------------|------|----------------|------|
| DAS-II subtest                        | Ь    | <b>S</b> <sup>2</sup> | Ь      | <b>S</b> <sup>2</sup> | Ь         | S <sup>2</sup> | Ь       | <b>S</b> <sup>2</sup> | h²   | u <sup>2</sup> | ECV  |
| Word Definitions                      | .655 | .429                  | .472   | .223                  |           |                |         |                       | .652 | .348           | .658 |
| Verbal Similarities                   | .662 | .438                  | .469   | .220                  |           |                |         |                       | .658 | .342           | .666 |
| Matrices                              | .772 | .596                  |        |                       | .000      | .000           |         |                       | .596 | .404           | .999 |
| Sequential and Quantitative Reasoning | .820 | .672                  |        |                       | .000      | .000           |         |                       | .672 | .328           | .999 |
| Pattern Construction                  | .705 | .497                  |        |                       |           |                | .319    | .102                  | .599 | .401           | .830 |
| Recall of Designs                     | .638 | .407                  |        |                       |           |                | .320    | .102                  | .509 | .491           | .799 |
| Total variance                        |      | .507                  |        | .074                  |           | .000           |         | .034                  | .614 | .386           |      |
| ECV                                   |      | .825                  |        | .120                  |           | .000           |         | .055                  |      |                |      |
| ω                                     |      | .893                  |        | .792                  |           | .776           |         | .713                  |      |                |      |
| ω <sub>H</sub> / ω <sub>HS</sub>      |      | .834                  |        | .268                  |           | .000           |         | .132                  |      |                |      |
| Relative $\omega$                     |      | .933                  |        | .338                  |           | .000           |         | .185                  |      |                |      |
| Factor correlation                    |      | .913                  |        | .517                  |           | .000           |         | .363                  |      |                |      |
| Н                                     |      | .871                  |        | .363                  |           | .000           |         | .185                  |      |                |      |
| PUC                                   |      | .800                  |        |                       |           |                |         |                       |      |                |      |

Note. DAS-II = Differential Ability Scales–Second Edition; CFA = confirmatory factor analysis; b = standardized loading of subtest on factor;  $S^2$  = variance explained in the subtest;  $h^2$  = communality;  $u^2$  = uniqueness; ECV = explained common variance;  $\omega$  = omega;  $\omega_H$  = omega-hierarchical (general factor);  $\omega_{HS}$  = omega-hierarchical subscale (group factors); H = construct reliability or replicability index; PUC = percentage of uncontaminated correlations.

and parameter estimates. Furthermore, the lack of reporting portions of variance captured by the various dimensions prohibits users of the DAS-II from determining which scores contain sufficient unique true score variance necessary for individual decision making. The present study attempted to overcome these shortcomings using the standardization sample raw data provided by NCS Pearson, Inc. for independent assessment.

Results of the present study paralleled quite well the EFA results from the DAS-II core subtests reported by Canivez and McGill (2016) that indicated that the DAS-II core subtests measured general intelligence well, and although the subtests generally had associations with theoretically linked first-order factors (V, NV, SP), the unique contributions of true score variance in the first-order group factors were universally low, prohibiting confident individual clinical interpretation. The present CFA results for the three DAS-II age groups (lower early years [2:6–3:5 years], upper early years [3:6–6:11 years], school age [7:0–17:11 years]) showed that although most subtests were generally aligned with their theoretical first-order group factors (V, NV, SP), most of the reliable subtest variance was associated with an overall, general intelligence factor (g), regardless of model expression (higher-order vs. bifactor). The dominance of the general intelligence factor and the limited unique measurement of the three group factors is evidenced by the subtest variance apportions where the general factor accounted for more than 6.84 times as much common subtest variance (3:6–6:11 years) and 6.88 times as much common subtest variance (7:0-17:11 years) as any individual DAS-II group factor and about 3 times as much common subtest variance (3:6-6:11 years) and about 4.7 times as much common subtest variance (7:0-17:11 years) as all three DAS-II group factors combined. Similar results were reported by Cucina and Howardson (2017) with the original DAS (Elliott, 1990).

The omega coefficients ( $\omega_{\rm H}$  and  $\omega_{\rm HS}$ ) and construct reliability or construct replicability coefficients (*H*) from CFA results of the bifactor models (and higher-order models) indicated that although the broad *g* factor allows for confident individual interpretation of the GCA, the  $\omega_{\rm HS}$  and *H* estimates for the three DAS-II group factors were unacceptably low (see Tables 3–5), and thus extremely limited for measuring unique cognitive constructs (Brunner et al., 2012; Hancock & Mueller, 2001; Reise, 2012; Rodriguez et al., 2016) supposedly quantified by the DAS-II cluster scores. Most disconcerting is the observation that for ages 7:0 to 17:11, the NV factor appears completely absent (a result suggested by Dombrowski, Golay, et al., 2018), yet an NV cluster score is provided for interpretation by the publisher. Such results indicate "to interpret subscale scores as representing the precise measurement of some latent variable that is unique or different from the general factor, clearly, is misguided" (Rodriguez et al., 2016, p. 225). Had variance apportions been reported in the DAS-II *Introductory and Technical Handbook*, this problem would have been disclosed and users of the DAS-II would be better able to decide whether there was little to nothing to report beyond the GCA.<sup>2</sup>

The present results, like those reported by Cucina and Howardson (2017) with the original DAS, challenge the CHC-inspired interpretive model preferred by the test publisher model (see also Canivez & Youngstrom, 2019), in that, the portions of unique variance conveyed by the broad ability clusters (V, NV, SP) are quite small and thus likely to be of little consequence but the variance contributed by g is quite large and of primary importance. Thus, it appears these results provide ample support for Carroll's conceptualization of the structure of intelligence but not Cattell and Horn or McGrew who have de-emphasized psychometric g and focused on the group factors (Horn & Blankson, 2005; Horn & Noll, 1997; McGrew, 2018). An additional theoretical implication is the preference for the bifactor model when there is an attempt to estimate or account for domain-specific abilities (Murray & Johnson, 2013), something explicitly done with DAS-II interpretations of V, NV, and SP scores and their comparisons. Users of the DAS-II must consider the empirical *evidence* of how well the group factor cluster scores (domain-specific) *uniquely* measure their represented construct independent of the general intelligence (g) factor

(GCA) score (Chen et al., 2006, 2012). Bifactor models contain a general factor but permit multidimensionality, which some consider an advantage relative to the higher-order model for determining the group factor contributions independent of the general intelligence factor (Reise et al., 2010).

Reynolds and Keith (2013) have questioned the appropriateness of the bifactor model and stated that "we believe that higher-order models are theoretically more defensible, more consistent with relevant intelligence theory (e.g., Jensen, 1998), than are less constrained hierarchical [bifactor] models" (p. 66). However, Gignac (2006, 2008) argued, in comparing bifactor and higher-order models, that general intelligence is the most substantive factor of a battery of cognitive tests, so g should be modeled *directly* and that it is the higher-order model that requires explicit theoretical justification for the full mediation of general intelligence by the group factors. Carroll (1993, 1995) empirically illustrated that variation in subtest scores reflect both general and a more specific group factor variances. So, although subtest scores may appear reliable, in the majority of cases, that reliability estimate is primarily due to the influence of the general factor and not the specific group factor (Carretta & Ree, 2001). Others have argued that Spearman's (1927) and Carroll's (1993) conceptualizations of intelligence are better represented by the bifactor model (Beaujean, 2015; Brunner et al., 2012; Frisby & Beaujean, 2015 Gignac, 2006, 2008; Gignac & Watkins, 2013; Gustafsson & Balke, 1993). For example, Beaujean (2015) suggests that Spearman's conception of general intelligence was of a factor "that was directly involved in all cognitive performances, not indirectly involved through, or mediated by, other factors" (p. 130) and also noted that "Carroll was explicit in noting that a bi-factor model best represents his theory" (p. 130).

The question of whether the general factor of intelligence *actually* represents a legitimate psychological dimension continues to be adjudicated and there are respected intelligence scholars that contend that g is nothing more than a statistical artifact. Recently, Kovacs and Conway (2016, 2019a, 2019b) presented their process overlap theory (POT), which argues for combination of, and attempts to merge, psychometric aspects of intelligence with cognitive psychology and neuroscience. They suggest that g is an emergent property (not the cause) of domain-general executive functions. Their effort was to provide a unified theory for general intelligence but Gottfredson (2016) pointed out a number of misconceptions and misattributions of g theory noting "the g theory they portray is not the one to which g theorists actually subscribe" (p. 210). Gottfredson welcomed the attempt to merge the disparate fields but illustrated how Kovacs and Conway are consistent with g theory, and not contrary to it, based on different levels of analysis.

Even so, the substantially greater total and common variance associated with general intelligence among DAS-II core subtests is a result that has been observed in numerous other studies examining the latent factor structure of intelligence or cognitive ability tests using both EFA and CFA procedures (Bodin et al., 2009; Canivez, 2008, 2014; Canivez & Watkins, 2010a, 2010b; Canivez et al., 2009, 2016, 2017; DiStefano & Dombrowski, 2006; Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013; Dombrowski et al., 2009; Gignac & Watkins, 2013; Nelson & Canivez, 2012; Nelson et al., 2007, 2013; Watkins, 2006, 2010; Watkins & Beaujean, 2014; Watkins et al., 2006, 2013). These results continue to support the dominance of psychometric g and are consistent with the literature regarding the practical importance of general intelligence (Deary, 2013; Gottfredson, 2008; Jensen, 1998; Lubinski, 2000; Ree et al., 2003). Although it appears that in the case of highly gifted (precocious) individuals, there are additional effects of spatial abilities and intraindividual differences (higher verbal or higher quantitative abilities) related to excelling in humanities or science, technology, engineering, and math (STEM) domains (Kell, Lubinski, & Benbow, 2013; Kell, Lubinski, Benbow, & Steiger, 2013; Lubinski, 2016; Makel et al., 2016) such that g accounts for *less* variance in these circumstances, g still typically accounts for the most variance. This phenomenon is described by Spearman's law of diminishing returns, which was specifically examined in the DAS-II by Reynolds et al. (2011), and they did indeed find that there was less *g* variance related to most subtests in the high- versus low-ability group, but for most subtests, there was still *more g* variance than broad ability variance associated with most DAS-II subtests. As such, the principal interpretation of DAS-II core subtests should be of the GCA, the estimate of *g*; although, perhaps in intellectually gifted individuals, other factors might be of value. The dominance of *g* variance captured by the DAS-II core subtests is a likely reason that methods to determine how many factors to extract and retain in EFA such as parallel analysis and minimum average partials suggest the DAS-II might be sufficiently represented by only one factor and the inability to locate the posited NV factor consistently in the present results (Crawford et al., 2010).

Relatedly, the confidence intervals provided for the DAS-II factor scores are considerably smaller (due to conflated general intelligence variance) than they might be if only the *unique* true score variance of the factor scores was used. The poor incremental validity provided by intelligence test group factors in accounting for meaningful portions of achievement variance beyond that provided by the omnibus composite IQ score in many contemporary intelligence tests (e.g., Canivez, 2013; Canivez et al. 2014; Glutting et al., 2006; McGill, 2015) may be the result of small amounts of unique variance captured by first-order factors as observed in the present study. Youngstrom et al. (1999) found in the assessment of incremental validity of the DAS factor scores, as predictors of achievement beyond the GCA, that interpretation of broad factor scores was not supported. Although incremental validity of DAS-II cluster scores above and beyond the GCA does not yet appear to have been investigated, it is hard to imagine these specific group factors would provide useful incremental information when predicting performance in academic achievement or relations with other external criteria given the current results.

Another problem for DAS-II interpretation is the recommended practice of identification of factor-based cognitive strengths and weaknesses through ipsative comparisons because analyses of DAS-II factor score differences at the observed score level conflate g variance and specific group factor (Verbal, Nonverbal, Spatial) variance. The same is true of analyses of subtest-based processing strengths and weaknesses (PSWs). Because it is not possible to disaggregate these sources of variance for individuals, it is impossible to know how much of the variance in performance is due to the general factor, specific group factor, or the narrow subtest ability. These concerns are in addition to the long-standing problems identified for ipsative score comparisons (McDermott et al., 1990, 1992; McDermott & Glutting, 1997) and suggest that these interpretive practices should probably be eschewed. In addition, the longitudinal stability of such processing strengths and weaknesses (PSWs) (see Watkins & Canivez, 2004) or diagnostic and treatment utility of such DAS-II PSWs in particular, has yet to be demonstrated. Although these types of profile analysis methods remain popular in clinical practice, compelling empirical support for the validity of these practices is presently lacking (e.g., Glutting et al., 2003; Macmann & Barnett, 1997; McDermott et al., 1990, 1992; McDermott & Glutting, 1997; McGill et al., 2018; Miciak et al., 2014; Watkins, 2000; Watkins et al., 2007).

Finally, it should be noted that these results are not unique to the DAS-II. As a result, a host of independent CFA and EFA studies of other major tests of intelligence such as the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Bodin et al., 2009; Canivez, 2014; Keith, 2005; Watkins, 2006, 2010; Watkins et al., 2006), Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V; Canivez et al., 2016, 2017, 2020; Dombrowski, Canivez, & Watkins, 2017), Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV; Canivez & Watkins, 2010a, 2010b; Nelson et al., 2013), Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition (WPPSI-IV; Watkins & Beaujean, 2014), Woodcock-Johnson–Third Edition (WJ III; Cucina & Howardson, 2017; Dombrowski, 2013, 2014a, 2014b; Dombrowski & Watkins, 2013; Strickland et al., 2015), Woodcock-Johnson–Fourth Edition (WJ IV; Dombrowski, McGill, & Canivez, 2017, 2018a, 2018b), Stanford–Binet Intelligence Scale: Fifth Edition

(SB-5; Canivez, 2008; DiStefano & Dombrowski, 2006), Kaufman Assessment Battery for Children (KABC; Cucina & Howardson, 2017), Kaufman Assessment Battery for Children–Second Edition (KABC-2; McGill & Dombrowski, 2018), Kaufman Adolescent and Adult Intelligence (KAIT; Cucina & Howardson, 2017), and Reynolds Intellectual Assessment Scales (RIAS; Dombrowski et al., 2009; Nelson & Canivez, 2012; Nelson et al., 2007) have reached similar conclusions about what commercial ability tests measure. We encourage practitioners to consider these results along with the psychometric meta-analysis conducted by Dombrowski, McGill, and Morgan (2019) when making decisions about how these measures should be interpreted and utilized in clinical practice.

### Limitations

The results of the present study pertain only to the latent factor structure of the DAS-II core subtests and do not fully test all aspects of construct validity. In fact, as emphasized by Bonifay et al. (2017), bifactor (and other structures) must be examined for adequacy against external criteria in theoretical validation. Latent profile analysis might be useful to determine whether the DAS-II is able to identify various diagnostic groups that might be expected to differ from normative samples. As previously mentioned, studies examining relations of DAS-II scores with external criteria such as examinations of incremental predictive validity (Youngstrom et al., 1999) or effects of extreme cluster score variability on DAS-II prediction of academic achievement (Kotz et al., 2008) should also be conducted. In addition to observed scores, DAS-II latent factor scores could also be examined for contributions to the explanation of academic achievement (see Glutting et al., 2006; Kranzler et al., 2015). Diagnostic utility of DAS-II cluster scores should also be examined to determine whether they offer utility for correct classification of individuals within specific groups or differential treatment response (see Canivez, 2013b).

### Conclusion

The present CFA results reinforce the admonition of extreme caution for any interpretations of DAS-II scores beyond the GCA (Canivez & McGill, 2016; Dombrowski, Golay, et al., 2018, 2019), including assessments for PSW. Due to the very small portions of unique true score variance provided by cluster scores and the inability to locate the NV score consistently across the age span of the test, such scores and their comparisons are potentially misleading. Better measurement of posited DAS-II first-order dimensions as distinct from g will likely require the creation and inclusion of more or better indicators as has been suggested with other general intelligence tests (Canivez et al., 2016, 2017; Dombrowski, McGill, & Canivez, 2017, 2018a, 2018b). These results, in addition to the advantages of bifactor modeling in aiding our understanding of test structure (Canivez, 2016; Cucina & Byle, 2017; Gignac, 2008; Reise, 2012), indicate that comparisons of bifactor and higher order representations are likely needed to fully understand what cognitive tests such as the DAS-II measure. Given "the ultimate responsibility for appropriate test use and interpretation lies predominantly with the test user" (American Educational Research Association et al., 2014, p. 141), consideration of the present results and other independent DAS-II studies allow users to "know what their tests can do and act accordingly" (Weiner, 1989, p. 829).

#### Authors' Note

Preliminary results were presented at the 2017 Annual Convention of the American Psychological Association, Washington, D.C. Standardization data from the Differential Ability Scales–Second Edition (DAS-II). Copyright© 1998, 2000, 2004, 2007 NCS Pearson, Inc. and Colin D. Elliott. Normative data copyright© 2007 NCS Pearson, Inc. Used with permission. All rights reserved.

#### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **ORCID** iDs

Gary L. Canivez D https://orcid.org/0000-0002-5347-6534 Stefan C. Dombrowski D https://orcid.org/0000-0002-8057-3751

#### Notes

- 1. The estimation method used by Keith and colleagues was not disclosed.
- 2. Standardized path coefficients between Nonverbal Reasoning Ability (NV) and g reported in the *Introductory and Technical Handbook* ranged from .93 to .99 suggesting those constructs may be redundant.

#### Supplemental material

Supplemental material for this article is available online.

#### References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.
- Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. Journal of Intelligence, 3, 121–136. https://doi.org/10.3390/jintelligence3040121
- Bentler, P. M., & Wu, E. J. C. (2016). EQS for Windows. Multivariate Software, Inc.
- 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. https://doi. org/10.1080/09297040802603661
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5, 184–186. https://doi. org/10.1177/2167702616657069
- Brown, T. A. (2015). Confirmatory factor analysis for applied research (2nd ed.). Guilford.
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80, 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods & Research, 33, 261–304. https://doi.org/10.1177/0049124104268644
- Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Lawrence Erlbaum.
- Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*, 49, 1716– 1735. https://doi.org/10.3758/s13428-016-0814-1
- Canivez, G. L. (2008). Orthogonal higher-order factor structure of the Stanford-Binet Intelligence Scales-Fifth Edition for children and adolescents. *School Psychology Quarterly*, 23, 533–541. https://doi. org/10.1037/a0012884
- Canivez, G. L. (2013a). Incremental validity of WAIS-IV factor index scores: Relationships with WIAT– II and WIAT–III subtest and composite scores. *Psychological Assessment*, 25, 484–495. https://doi. org/10.1037/a0032092
- Canivez, G. L. (2013b). 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 assessments* (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, 29, 38–51. https://doi.org/10.1037/spq0000032
- Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247– 271). Hogrefe.
- Canivez, G. L., Konold, T. R., Collins, J. M., & Wilson, G. (2009). Construct validity of the Wechsler Abbreviated Scale of Intelligence and Wide Range Intelligence Test: Convergent and structural validity. *School Psychology Quarterly*, 24, 252–265. http://doi.org/10.1037/a0018030
- 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, 1475–1488. https://doi.org/10.1037/pas0000279
- Canivez, G. L., McGill, R. J., Dombrowski, S. C., Watkins, M. W., Pritchard, A. E., & Jacobson, L. A. (2020). Construct validity of the WISC–V in clinical cases: Exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment*, 27, 274–296. https://doi.org/10.1177/1073191118811609
- Canivez, G. L., & Watkins, M. W. (2010a). Investigation of the factor structure of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV): Exploratory and higher order factor analyses. *Psychological Assessment*, 22, 827–836. https://doi.org/10.1037/a0020429
- Canivez, G. L., & Watkins, M. W. (2010b). Exploratory and higher-order factor analyses of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV) adolescent subsample. *School Psychology Quarterly*, 25, 223–235. https://doi.org/10.1037/a0022046
- 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, 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, 458–472. https://doi.org/10.1037/pas0000358
- Canivez, G. L., Watkins, M. W., James, T., James, K., & Good, R. (2014). Incremental validity of WISC– IV<sup>UK</sup> factor index scores with a referred Irish sample: Predicting performance on the WIAT–IIUK. *British Journal of Educational Psychology*, 84, 667–684. https://doi.org/10.1111/bjep.12056
- Canivez, G. L., & Youngstrom, E. A. (2019). Challenges to the Cattell-Horn-Carroll Theory: Empirical, clinical, and policy implications. *Applied Measurement in Education*, 32, 232–248. https://doi.org/10. 1080/08957347.2019.1619562
- Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research. *International Journal of Selection and* Assessment, 9, 325–335. https://doi.org/10.1111/1468-2389.00184
- Carroll, J. B. (1993). Human cognitive abilities. Cambridge University Press.
- Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research*, 30, 429–452. https://doi.org/10.1207/s15327906mbr3003\_6
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Pergamon Press.
- Cattell, R. B. (1978). The scientific use of factor analysis in behavioral and life sciences. Plenum Press.
- Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 15, 139–164. https://doi. org/10.1111/j.1745-3984.1978.tb00065.x
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. https://doi.org/10.1080/10705510701301834
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal* of Personality, 80, 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41, 189–225. https://doi.org/10.1207/s15327906mbr4102\_5
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255. https://doi.org/10.1207/ S15328007SEM0902\_5

- Crawford, A. V., Green, S. B., Levy, R., Lo, W.-J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement*, 70, 885–901. https://doi.org/10.1177/0013164410379332
- Cucina, J. M., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence*, 5, 27–48. https://doi. org/10.3390/jintelligence5030027
- Cucina, J. M., & Howardson, G. N. (2017). Woodcock-Johnson–III, Kaufman Adolescent and Adult Intelligence Test (KAIT), Kaufman Assessment Battery for Children (KABC), and Differential Ability Scales (DAS) Support Carroll but not Cattell-Horn. *Psychological Assessment*, 29, 1001–1015. https:// doi.org/10.1037/pas0000389
- Deary, I. J. (2013). Intelligence. Current Biology, 23, 673–676. https://doi.org/10.1016/j.cub.2013.07.021
- DiStefano, C., & Dombrowski, S. C. (2006). Investigating the theoretical structure of the Stanford-Binet-Fifth Edition. *Journal of Psychoeducational Assessment*, 24, 123–136. https://doi. org/10.1177/0734282905285244
- Dombrowski, S. C. (2013). Investigating the structure of the WJ–III Cognitive at school age. School Psychology Quarterly, 28, 154–169. https://doi.org/10.1037/spq0000010
- Dombrowski, S. C. (2014a). Exploratory bifactor analysis of the WJ–III Cognitive in adulthood via the Schmid-Leiman procedure. *Journal of Psychoeducational Assessment*, 32, 330–341. https://doi. org/10.1177/0734282913508243
- Dombrowski, S. C. (2014b). Investigating the structure of the WJ–III Cognitive in early school age through two exploratory bifactor analysis procedures. *Journal of Psychoeducational Assessment*, 32, 483–494. https://doi.org/10.1177/0734282914530838
- Dombrowski, S. C., Canivez, G. L., & Watkins, M. W. (2017). Factor structure of the 10 WISC–V primary subtests across four standardization age groups. *Contemporary School Psychology*, 22, 90–104. https:// doi.org/10.1007/s40688-017-0125-2
- Dombrowski, S. C., Golay, P., McGill, R. J., & Canivez, G. L. (2018). Investigating the theoretical structure of the DAS-II core battery at school age using Bayesian structural equation modeling. *Psychology in the Schools*, 55, 190–207. https://doi.org/10.1002/pits.22096
- 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, 394–407. http://dx.doi.org/10.1037/ pas0000350
- Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018a). 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, 1–13. http://dx.doi.org/10.1037/arc0000039
- Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018b). 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
- Dombrowski, S. C., McGill, R. J., Canivez, G. L., & Peterson, C. H. (2019). Investigating the theoretical structure of the Differential Ability Scales–Second Edition through hierarchical exploratory factor analysis. *Journal of Psychoeducational Assessment*, 37, 94–104. https://doi.org/10.1177/0734282918760724
- Dombrowski, S. C., McGill, R. J., & Morgan, G. B. (2019). Monte Carlo modeling of contemporary intelligence test (IQ) factor structure: Implications for IQ assessment, interpretation, and theory. Assessment. Advance online publication. https://doi.org/10.1177/1073191119869828
- Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ– III full test battery: A school aged analysis. *Psychological Assessment*, 25, 442–455. https://doi. org/10.1037/a0031335
- Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, 27, 494–507. https://doi.org/10.1177/0734282909333179
- Elliott, C. D. (1990). Differential Ability Scales. The Psychological Corp.
- Elliott, C. D. (2007a). Differential Ability Scales-Second Edition. Harcourt Assessment.
- Elliott, C. D. (2007b). *Differential Ability Scales–Second Edition: Introductory and technical handbook.* Harcourt Assessment.

- Elliott, C. D., Murray, D. J., & Pearson, L. S. (1979). *British Ability Scales*. National Foundation for Educational Research.
- Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS– IV/WMS–IV standardization data. *Intelligence*, 51, 79–97. https://doi.org/10.1016/j.intell.2015.04.007
- Gignac, G. E. (2006). The WAIS–III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences*, 27, 73–86. https://doi. org/10.1027/1614-0001.27.2.73
- Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: *g* as superordinate or breadth factor? *Psychology Science Quarterly*, *50*, 21–43.
- Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS–IV. *Multivariate Behavioral Research*, 48, 639–662. https://doi.org/10.1080/00273171. 2013.804398
- Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC–IV in estimating reading and math achievement on the WIAI–II. *Journal of Special Education*, 40, 103–114. https://doi.org/10.1177/002 24669060400020101
- Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). Multifactored and cross-battery assessments: Are they worth the effort? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological* and educational assessment of children: Intelligence, aptitude, and achievement (2nd ed., pp. 343– 376). Guilford.
- Gottfredson, L. S. (2008). Of what value is intelligence? In A. Prifitera, D. Saklofske, & L. G. Weiss (Eds.), *WISC–IV clinical assessment and intervention* (2nd ed., pp. 545–564). Elsevier.
- Gottfredson, L. S. (2016). A g theorist on why Kovacs and Conway's Process Overlap Theory amplifies, not opposes, g theory. *Psychological Inquiry*, 27, 210–217. http://doi.org/10.1080/10478 40X.2016.1203232
- Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*, 407–434. https://doi.org/10.1207/s15327906mbr2804\_2
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. Du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 195–216). Scientific Software International.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2, 41–54. https://doi. org/10.1007/BF02287965
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock-Johnson technical manual* (Rev. ed., pp. 197–232). Riverside.
- Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 41–68). Guilford.
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf–Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). Guilford.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 5, 1–55. https://doi.org/10.1080/10705519909540118
- Jensen, A. R. (1998). The g factor: The science of mental ability. Praeger.
- Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581–614). Guilford.
- Keith, T. Z., & Kranzler, J. H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the Cognitive Assessment System does and does not measure. *School Psychology Review*, 28, 303–321.
- Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales–II: Consistency across ages 4 to 17. *Psychology in the Schools*, 47, 676–697. https://doi.org/10.1002/pits.20498

- Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science*, 24, 648–659. https://doi.org/10.1177/0956797612457784
- Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science*, 24, 1831–1836. https://doi.org/10.1177/0956797613478615
  Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford.
- Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the general conceptual ability score from the Differential Ability Scales as a function of significant and rare interfactor variability. *School Psychology Review*, 37, 261–278.
- Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, 27, 151–177. http://doi.org/10.1080/1047840X.2016.1153946
- Kovacs, K., & Conway, A. R. A. (2019a). A unified cognitive/differential approach to human intelligence: Implications for IQ testing. *Journal of Applied Research in Memory and Cognition*, 8, 255–272. https:// doi.org/10.1016/j.jarmac.2019.05.003
- Kovacs, K., & Conway, A. R. A. (2019b). What is IQ? Life beyond "general intelligence." Current Directions in Psychological Science, 28, 189–194. http://dx.doi.org/10.1177/0963721419827275
- Kranzler, J. H., Benson, N., & Floyd, R. G. (2015). Using estimated factor scores from a bifactor analysis to examine the unique effects of the latent variables measured by the WAIS–IV on academic achievement. *Psychological Assessment*, 27, 1402–1416. https://doi.org/10.1037/pas0000119
- Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, 4, 192–211. https://doi.org/10.1037/1082-989X.4.2.192
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points." *Annual Review of Psychology*, 51, 405–444. https://doi.org/10.1146/annurev. psych.51.1.405
- Lubinski, D. (2016). From Terman to today: A century of findings on intellectual precocity. *Review of Educational Research*, 86, 900–944. https://doi.org/10.3102/0034654316675476
- Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "Intelligent Testing" approach to the WISC–III. School Psychology Quarterly, 12, 197– 234.
- Makel, M. C., Kell, H. J., Lubinski, D., Putallaz, M., & Benbow, C. P. (2016). When lightning strikes twice: Profoundly gifted, profoundly accomplished. *Psychological Science*, 27, 1004–1018. https://doi. org/10.1177/0956797616644735
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519–530. https://doi.org/10.1093/biomet/57.3.519
- Marley, S. C., & Levin, J. R. (2011). When are prescriptive statements in educational research justified? *Educational Psychology Review*, 23, 197–206. https://doi.org/10.1007/s10648-011-9154-y
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis- testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320–341. https://doi.org/10.1207/ s15328007sem1103\_2
- McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment*, 8, 290–302. https://doi. org/10.1177/073428299000800307
- 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. https://doi.org/10.1177/002246699202500407
- McDermott, P. A., & Glutting, J. J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests—Or, more illusions of meaning? *School Psychology Review*, 26, 163– 176.
- McDonald, R. P. (2010). Structural models and the art of approximation. Perspectives on Psychological Science, 5, 675–686. https://doi.org/10.1177/1745691610388766
- McGill, R. J. (2015). Incremental criterion validity of the WJ-III COG clinical clusters: Marginal predictive effects beyond the general factor. *Canadian Journal of School Psychology*, 30, 51–63. https://doi. org/10.1177/0829573514560529

- McGill, R. J., & Dombrowski, S. C. (2018). Factor structure of the CHC model for the KABC-II: Exploratory factor analyses with the 16 core and supplemental subtests. *Contemporary School Psychology*, 22, 279–293. https://doi.org/10.1007/s40688-017-0152-z
- 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
- McGrew, K. S. (2018, April 12). Dr. Kevin McGrew and Updates to CHC Theory [Video webcast]. Invited podcast presentation for School Psyched! Podcast presented 12 April 2018. https://itunes.apple.com/ us/podcast/episode-64-dr-kevin-mcgrew-and-updates-to-chc-theory/id1090744241?i=100040872862 0&mt=2
- Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly*, 29, 21–37. https://doi.org/10.1037/spq0000037
- Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing bi-factor versus higherorder models of human cognitive ability structure. *Intelligence*, 41, 407–422. https://doi.org/10.1016/j. intell.2013.06.004
- Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychological Assessment*, 24, 129–140. https://doi.org/10.1037/a0024878
- Nelson, J. M., Canivez, G. L., Lindstrom, W., & Hatt, C. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology*, 45, 439–456. https://doi.org/10.1016/j.jsp.2007.03.003
- Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV) with a clinical sample. *Psychological Assessment*, 25, 618–630. https://doi.org/10.1037/a0032086
- Onwuegbuzie, A. J., & Daniel, L. G. (2002). Uses and misuses of the correlation coefficient. *Research in the Schools*, 9, 73–90.
- Ree, M. J., Carretta, T. R., & Green, M. T. (2003). The ubiquitous role of g in training. In H. Nyborg (Ed.), The scientific study of general intelligence: Tribute to Arthur R. Jensen (pp. 262–274). Pergamon Press.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696. https://doi.org/10.1080/00273171.2012.715555
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95, 129–140. https://doi.org/ 10.1080/00223891.2012.725437
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92, 544–559. https://doi.org/10.1080/00223891.2010.496477
- Reynolds, M. R., Hajovsky, D. B., Niileksela, C. R., & Keith, T. Z. (2011). Spearman's law of diminishing returns and the DAS-II: Do g effects on subtest scores depend on the level of g? *School Psychology Quarterly*, 26, 275–289.
- Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In D. H. Saklofske, V. L. Schwean, & C. R. Reynolds (Eds.), Oxford handbook of child psychological assessment (pp. 48–83). Oxford University Press.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98, 223–237. https://doi.org/10.10 80/00223891.2015.1089249f
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21, 137–150. https://doi.org/10.1037/met0000045
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514. https://doi.org/10.1007/BF02296192
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53–61. https://doi.org/10.1007/BF02289209
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDounough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford.

Spearman, C. (1927). The abilities of man. Cambridge University Press.

- Strickland, T., Watkins, M. W., & Caterino, L. C. (2015). Structure of the Woodcock-Johnson III cognitive tests in a referral sample of elementary school students. *Psychological Assessment*, 27, 689–697. https://doi.org/10.1037/pas0000052
- Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. School Psychology Quarterly, 15, 465–479. https://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. https://doi.org/10.1037/1040-3590.18.1.123
- 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. https://doi.org/10.1037/ a0020043
- Watkins, M. W. (2013). Omega [Computer software]. Ed & Psych Associates.
- Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The Clinical Neuropsychologist*, 31, 1113–1126. https://doi.org/10.1080/13854046.2017.131 7364
- 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. https://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. https://doi.org/10.1037/1040–3590.16.2.133
- Watkins, M. W., Glutting, J. J., & Lei, P. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, 14, 13–20. https://doi.org/10.1080/09084280701280353
- Watkins, M. W., Canivez, G. L., James, T., Good, R., & James, K. (2013). Construct validity of the WISC-IV-UK with a large referred Irish sample. *International Journal of School and Educational Psychology*, 1, 102–111. http://doi.org/10.1080/21683603.2013.794439
- 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, 975–983. https://doi.org/10.1177/0013164406288168
- Weiner, I. B. (1989). On competence and ethicality in psychodiagnostic assessment. *Journal of Personality Assessment*, 53, 827–831. https://doi.org/10.1207/s15327752jpa5304\_18
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). SAGE.
- Youngstrom, E. A., Kogos, J. L., & Glutting, J. J. (1999). Incremental efficacy of Differential Ability Scales factor scores in predicting individual achievement criteria. *School Psychology Quarterly*, 14, 26–39. https://doi.org/10.1037/h0088996
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ωh: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70, 123–133. https://doi.org/10.1007/s11336-003-0974-7
- Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ωh. *Applied Psychological Measurement*, *30*, 121–144. https://doi.org/10.1177/0146621605278814

# Appendix (Online Supplement)

Table A1

Standardization Sample Pearson Product-Moment Correlations and Descriptive Statistics for DAS–II Core Subtests

| Ages 2:6 to 3:5 ( <i>N</i> = 352)         | VC      | NV      | PS      | PC      |         |         |
|---|---------|---------|---------|---------|---------|---------|
| Verbal Comprehension (VC)                 | _       |         |         |         |         |         |
| Naming Vocabulary (NV)                    | .664    | _       |         |         |         |         |
| Picture Similarities (PS)                 | .484    | .505    | _       |         |         |         |
| Pattern Construction (PC)                 | .441    | .423    | .367    | _       |         |         |
| Μ   | 49.693  | 49.676  | 50.045  | 48.244  |         |         |
| SD  | 10.334  | 10.322  | 9.506   | 9.074   |         |         |
| Sk  | 0.255   | 0.068   | - 0.623 | 0.157   |         |         |
| Κ   | 3.070   | 2.695   | 1.178   | 1.812   |         |         |
| Ages 3:6 to 6:11 ( <i>N</i> = 920)        | VC      | NV      | PS      | MAT     | PC      | СР      |
| Verbal Comprehension (VC)                 | _       |         |         |         |         |         |
| Naming Vocabulary (NV)                    | .554    | _       |         |         |         |         |
| Picture Similarities (PS)                 | .399    | .412    | _       |         |         |         |
| Matrices (MAT)                            | .354    | .309    | .397    | _       |         |         |
| Pattern Construction (PC)                 | .412    | .427    | .422    | .374    | _       |         |
| Copying (CP)                              | .350    | .335    | .354    | .298    | .525    | _       |
| Μ   | 50.036  | 50.414  | 49.790  | 49.994  | 50.432  | 50.294  |
| SD  | 9.750   | 9.560   | 9.947   | 10.119  | 10.222  | 10.061  |
| Sk  | 0.733   | 0.096   | 0.496   | - 0.147 | - 0.214 | - 0.278 |
| Κ   | 3.195   | 0.851   | 2.564   | 0.628   | 1.542   | 1.383   |
| Ages 7:0 to 17:11 ( <i>N</i> = 2,188)     | WD      | VS      | MAT     | SQR     | PC      | RD      |
| Word Definitions (WD)                     | _       |         |         |         |         |         |
| Verbal Similarities (VS)                  | .655    | _       |         |         |         |         |
| Matrices (MAT)                            | .494    | .508    | _       |         |         |         |
| Sequential & Quantitative Reasoning (SQR) | .547    | .546    | .633    | _       |         |         |
| Pattern Construction (PC)                 | .450    | .458    | .548    | .581    | _       |         |
| Recall of Designs (RD)                    | .431    | .434    | .500    | .508    | .551    | _       |
| M   | 50.160  | 50.258  | 50.190  | 50.226  | 50.009  | 49.972  |
| SD  | 9.829   | 9.885   | 10.203  | 10.335  | 9.956   | 9.919   |
| Sk  | - 0.269 | - 0.911 | - 0.216 | 0.224   | 0.269   | - 0.391 |
| Κ   | 2.375   | 2.355   | 0.556   | 2.075   | 2.320   | 1.470   |

*Note.* Sk = skewness, K = kurtosis. Mardia's normalized multivariate kurtosis estimates: Ages 2:6 to 3:5 (11.37), Ages 3:6 to 6:11 (26.49), Ages 7:0 to 17:11 (33.86).