Psychological Assessment
2018, Vol. 30, No. 6, 793–808

1040-3590/18/$12.00 http://dx.doi.org/10.1037/pas0000526

Complementary Exploratory and Confirmatory Factor Analyses of the French WISC–V: Analyses Based on the Standardization Sample

Thierry Lecerf
University of Geneva and Swiss Distance Learning University

Gary L. Canivez
Eastern Illinois University

Interpretation of the French Wechsler Intelligence Scale for Children–Fifth Edition (French WISC–V; Wechsler, 2016a) is based on a 5-factor model including Verbal Comprehension (VC), Visual Spatial (VS), Fluid Reasoning (FR), Working Memory (WM), and Processing Speed (PS). Evidence for the French WISC–V factorial structure was established exclusively through confirmatory factor analyses (CFAs). However, as recommended by Carroll (1995); Reise (2012), and Brown (2015), factorial structure should derive from both exploratory factor analysis (EFA) *and* CFA. The first goal of this study was to examine the factorial structure of the French WISC–V using EFA. The 15 French WISC–V primary and secondary subtest scaled scores intercorrelation matrix was used and factor extraction criteria suggested from 1 to 4 factors. To disentangle the contribution of first- and second-order factors, the Schmid and Leiman (1957) orthogonalization transformation (SLT) was applied. Overall, no EFA evidence for 5 factors was found. Results indicated that the *g* factor accounted for about 67% of the common variance and that the contributions of the first-order factors were weak (3.6 to 11.9%). CFA was used to test numerous alternative models. Results indicated that bifactor models produced better fit to these data than higher-order models. Consistent with previous studies, findings suggested dominance of the general intelligence factor and that users should thus emphasize the Full Scale IQ (FSIQ) when interpreting the French WISC–V.

Public Significance Statement
The present study indicated that the factorial structure of the French Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V) consists of a general intelligence factor (*g* factor) and 4 first-order primary factors. Data were not consistent with the 5-factor model promulgated by the publisher. The general intelligence factor accounted for the largest portion of common variance, hence supported the primary interpretation of the FSIQ.

Keywords: WISC–V, exploratory factor analyses, confirmatory factor analysis, ESEM, Schmid-Leiman transformation

Supplemental materials: http://dx.doi.org/10.1037/pas0000526.supp

Assessment of cognitive abilities of French-speaking children is mainly performed using the Wechsler Intelligence Scale for Children and the latest version, the French WISC–V (Wechsler, 2016a) was recently published. While the United States (U.S.) WISC–V is composed of 16 intelligence subtests, the French WISC–V includes *only* 15 subtests. Unlike the U.S. WISC–V, the French WISC–V includes neither the Picture Concepts subtest nor Complementary Index score (Naming Speed, Symbol Translation, and Storage and Retrieval) subtests that were added to assess additional cognitive processes. Seven primary subtests are used to estimate the FSIQ: Block Design (BD), Similarities (SI), Matrix Reasoning (MR), Digit Span (DS), Coding (CD), Vocabulary (VC), and Figure Weights (FW). Three additional primary subtests (Visual Puzzles [VP], Picture Span [PS], and Symbol Search [SS]) permit estimation of the five first-order primary factor index scores: VC, VS, FR, WM, and PS. At the secondary level, five subtests (Information [IN], Letter-Number Sequencing [LNS], Cancellation [CA], Comprehension [CO], and Arithmetic [AR]) can replace one subtest in the estimation of the FSIQ or be used in the estimation of five ancillary index scores: Quantitative Reasoning (QR), Auditory Working Memory (AWM), Nonverbal (NV), General Ability (GA), Cognitive Proficiency (CP).

From a theoretical perspective, although the WISC–V attempted to reflect the current conceptualizations of intellectual measurement, Naglieri (2016, p. 665) argued “there is no unifying theory upon which the WISC–V was built” (Reynolds, & Keith, 2017). However, the Cattell-Horn-Carroll (CHC) theory of the cognitive

This article was published Online First December 28, 2017.

Thierry Lecerf, Faculty of Psychology and Educational Sciences, University of Geneva, and Department of Psychology, Swiss Distance Learning University; Gary L. Canivez, Department of Psychology, Eastern Illinois University.

Correspondence concerning this article should be addressed to Thierry Lecerf, Faculty of Psychology and Educational Sciences, University of Geneva, FPSE, 40 Bd du Pont d’Arve, 1205 Geneva, Switzerland. E-mail: thierry.lecerf@unige.ch

ability (McGrew, 2009; Schneider & McGrew, 2012) as well as neuropsychological theories were provided as theoretical foundations in the French WISC–V *Interpretive Manual* (Wechsler, 2016b). The main theoretical goal of the WISC–V was to split the Perceptual Reasoning (PR) factor into two distinct factors, VS and FR. The FSIQ and the first-order primary factors of the French WISC–V reflect the higher-order CHC model. The publisher of the French WISC–V (Pearson France & ECPA) provided evidence of the factorial structure exclusively through CFAs. Further, their CFA examined only higher-order models with the final model consisting of a general intelligence factor (*g* factor) and five first-order primary factors. In a higher-order model, there are no direct loadings or influences of *g* on the subtest indicators. It is thus assumed that the general intelligence factor is fully mediated by the five first-order broad abilities in influencing the subtest scores and hence *g* is a *superordinate* factor (Canivez, 2014; Gignac, 2008). The first-order factors are components of the global factor, rather than distinct components. The five-factor higher-order model preferred by the publisher of the French WISC–V (Model 5e, Wechsler, 2016b, Figure 5.1 p. 70), is exactly the one proposed in the United States version (Chen, Zhang, Raiford, Zhu, & Weiss, 2015; Wechsler, 2014), and includes three cross loadings of the AR subtest on VC, FR, and WM latent variables, while all other 14 subtests were associated with only one latent variable.

Like the U.S. WISC–V, several concerns (Canivez & Watkins, 2016; Canivez, Watkins, & Dombrowski, 2016, 2017; Dombrowski, Canivez, Watkins, & Beaujean, 2015) can also be raised regarding the French WISC–V factorial structure, and regarding the CFAs reported in the French WISC–V *Interpretive Manual*. The first, and perhaps the most important issue, is that the French test publisher's did not report EFAs.[1] Evidence of the factorial structure of the French WISC–V was established exclusively using CFAs. However, EFA and CFA are not redundant and must be used in a complementary manner (Brown, 2015; Carroll, 1993, 1995; Gorsuch, 1983; Morin, Arens, & Marsh, 2016). Reise (2012), for instance, suggested "the necessity to conduct exploratory [bifactor] analysis prior to considering confirmatory modeling" (p. 677). Concretely, it means that the basic model tested with CFAs must be the model obtained with EFAs, which determine the appropriate number of factors and which variables are indicators of the various latent variables. Ruscio and Roche (2012) indicated that poor decisions would be made if the number of factors that is retained is not correct. For instance, Canivez (2008) found that two factors were sufficient to describe the factorial structure of the Stanford-Binet Intelligence Scales, Fifth edition (SB5) while Dombrowski, Watkins, and Brogan (2009) concluded that only one factor was sufficient for the Reynolds Intellectual Assessment Scales (RIAS). These findings were consistent with Frazier and Youngstrom's (2007) conclusion that intelligence batteries were frequently "overfactored." Most importantly, these studies indicated that the general factor accounted for most of the total and the common variance as has also been found in other studies (Canivez & Watkins, 2010; Watkins, 2006; Watkins, Wilson, Kotz, Carbone, & Babula, 2006), and hence clinical interpretations that focus on first-order factors should be done with caution (Canivez, 2008; Golay & Lecerf, 2011).

Regarding the French WISC–V, this is particularly important because the new French WISC–V deviates from the French WISC–IV (Wechsler, 2005): three new subtests were introduced (FW, VP, PS) and two subtests were removed (PC, WR). The French WISC–V also differs from the U.S. WISC–V by not including the Picture Concepts subtest. Therefore, because the factors and the factorial structure depend on the subtests included within CFA, the basic factorial structure of the French WISC–V should not derive directly from the factorial structure of the French WISC–IV nor directly from the U.S. WISC–V.

Although CFA has largely replaced EFA, CFA has some disadvantages. For instance, fixing some cross loadings to zero might specify a too parsimonious model and poor model fit. Therefore, to avoid some limitations of EFA and CFA, Asparouhov and Muthén (2009) developed the exploratory structural equation modeling (ESEM) to assess the construct-relevant multidimensionality in measurement instruments. This method integrates EFA within the structural equation modeling (SEM) framework and provides fit statistics (Morin, Arens, et al., 2016, Morin, Arens, Tran, & Caci, 2016). Bifactor ESEM was developed to allow the estimation of one general factor (G) and specific group factors (S). The general factor reflects the variance that is shared across all indicators, while S factors reflect the residual covariance not explained by the general factor. While bifactor models would be more adequate for hierarchically organized constructs, bifactor ESEM would be more appropriate for conceptually adjacent construct. Although Morin, Arens, et al. (2016) suggested that bifactor ESEM could be useful with multidimensional measures, and particularly when hierarchically superior constructs are included, the present study examined EFA and CFA similar to that conducted in published studies of the WISC–V (Canivez et al., 2016, 2017). Nevertheless, bifactor ESEM results were also examined and briefly presented.

There are six additional issues that concern the test publisher's CFA. First, the publisher of the French WISC–V inexplicably used weighted least squares (WLS) estimation without specific justification as was done with the United States version (Canivez et al., 2016, 2017). The "use of an estimation method other than ML [maximum likelihood] requires explicit justification" (Kline, 2011, p. 154), because estimation method can substantially affect parameter estimation, but no such explanation was provided. Assuming that the standardization sample data were normally distributed and subtest indicators were on an interval scale of measurement, then maximum lilelihood (ML) estimation should probably have been used because it provides the most precise variance estimates (Kline, 2016; Ullman, & Bentler, 2013). Second, as suggested by Beaujean (2016), it is not known how the publisher set the scales for the latent variables (i.e., fixing a loading, fixing latent variable variance, or effects coding). Because *df* are very important in computing statistical indices and understanding results, it is necessary to know which parameters were fixed and which were freely estimated. Third, the publisher selected the best fitting model based solely on the χ^2 difference for all models with four and five first-order factors (Wechsler, 2016b, Table 5.3, p. 69). However, according to Cheung and Rensvold (2002) and Chen (2007), ΔCFI $>$ .01 and ΔRMSEA $>$ .015 indicate meaningful differences between models. Nor did the publisher use Akaike's information criterion (AIC) as suggested by Kline (2016). The use of the ΔCFI, for instance, could allow avoiding the confusion

[1] Similarly, no EFAs were reported in the U.S. WISC–V technical manual.

between statistical significance and practical significance (Kline, 2016, p. 465).

Fourth, under the CFA framework and SEM, an important step is to determine whether the model is consistent with the data. Most frequently, this evaluative step focused on *global* or *model-fit* statistics (Byrne, 2001). However, as recommended by Kline (2016, p. 461–462), a model should never be retained "*solely on global fit testing,*" local fit assessment should always be conducted. Regarding the favored French WISC–V measurement model (labeled 5e in the French WISC–V *Interpretive Manual*), local fit assessment revealed two main problems: (a) the measurement model included a standardized path coefficient of VC on AR (.02) that was not only not statistically significant, but also not meaningful; and (b) the standardized loading between *g* and Fluid intelligence (Gf) was 1.00, a finding consistent with previous studies (Weiss, Keith, Zhu, & Chen, 2013), that suggests that the French WISC–V may be overfactored, or a result of statistical artifact. Indeed, Bayesian structural equation modeling with the French WISC–IV suggested that the correlation between *g* and Gf was not 1.00, but around .88 (Golay, Reverte, Rossier, Favez, & Lecerf, 2013). Inappropriate zero cross loadings could account for the unitary loading between *g* and Gf, because the correlations between first-order factors can be overestimated when using standard CFA estimators (Asparouhov, & Muthén, 2009; Marsh et al., 2010; Morin, Arens, Tran, & Caci, 2016). Other authors have suggested that with classical ML-CFA, statistical power could explain why it may not always be possible to distinguish FR from the general factor (Matzke, Dolan, & Molenaar, 2010). Thus, the equivalence of *g* and Gf in higher-order models may be better accounted for by statistical artifacts than the mere equivalence of the two constructs. Nevertheless, this perfect correlation between *g* and Gf constitutes a problem and may pose a threat to discriminant validity of test scores interpretation and use. Taken together, local fit assessment of the publisher preferred model (Model 5e) suggested it was not adequate.

Fifth, the publisher did not examine rival bifactor models (Holzinger & Swineford, 1937); only higher-order measurement models were reported in the French WISC–V *Interpretive Manual*. However, higher-order and bifactor models are commonly used to fit multidimensional data (Canivez, 2016; Canivez, Watkins, & Dombrowski, 2017; Carroll, 1993; Gignac, 2016; Mansolf, & Reise, 2017; Reise, 2012). While higher-order and bifactor models are very similar, and are nested models (Mansolf & Reise, 2017), they underlie distinctly different theories of the structure of cognitive abilities. The higher-order model is a hierarchical structure, while the bifactor model is nonhierarchical, because in this latter model, all subtest scores load directly onto a general factor and also onto one (or more) of the first-order group factors. The *g* factor has no effect on the first-order group factors, which are modeled orthogonally to each other (i.e., uncorrelated first-order factors) and to the *g* factor. In contrast, in the higher-order model, the *g* factor is indirectly related to subtest scores, and explains the correlations between first-order factors. Consequently, while the *g* factor is conceptualized as a superordinate factor in higher-order models, it is conceptualized as a *breadth* factor in bifactor models (Gignac, 2008).

The omission of bifactor model examination is especially astonishing because some authors regularly prefer them to higher-order models. One reason bifactor models are preferred over the higher-order model is that they are more in line with Carroll's *three-stratum* model, which is considered as foundational to the WISC–V. Indeed, like Spearman (1927); Carroll (1993) favored the bifactor model. A second reason is that bifactor models allow for a better partitioning of general and group factor variances. The bifactor model permits easy identification of the relative importance of the first-order group factors and the general factor on each test score, and also permits the estimation of omega-hierarchical (ω_H) and omega-hierarchical subscale (ω_{HS}). These coefficients are model-based coefficients that represent the proportion of variance in a unit-weighted composite score that is attributable to a factor. The ω_H estimates the proportion of variance explained by the general factor independent of group factors, while the ω_{HS} estimates the proportion of variance uniquely explained by the group factors with effects of the general and other group factors removed (Zinbarg, Revelle, Yovel, & Li, 2005). For instance, if a test battery has an ω_H of .80, this means that the unit-weighted score for *g* based on its indicators accounts for 80% true score variance. Thus, omega coefficients permit the determination of the importance of the general factor and group factors in the interpretation of composite and subtest scores. In contrast, one substantial limitation of higher-order models is that the first-order (broad abilities) influence on subtest scores is confounded with *g*'s influence. The lack of distinction between *g*'s direct influence on subtests scores may lead to overestimation of the broad abilities' influence.

From a theoretical perspective, although bifactor and higher-order models are very similar, choices between a bifactor model or a higher-order model are necessary, because they imply different theoretical conceptions of the general factor for instance (i.e., breadth vs. superordinate). Several studies have indicated that bifactor models provided better fit to the Wechsler Intelligence Scales than higher-order models (Beaujean, Parkin, & Parker, 2014; Canivez et al., 2016, 2017; Gignac, 2016; Golay & Lecerf, 2011). Consequently, some studies were conducted to determine whether the bifactor model represents really a more adequate description of cognitive abilities than higher-order model. Murray and Johnson (2013), on the basis of simulation studies, found that standard indices of goodness of fit (AIC, Tucker-Lewis index [TLI], and Bayesian information criterion [BIC]) were biased in favor of the bifactor model, because the data involve unmodeled complexity. Murray and Johnson suggested that although the bifactor model fits better, it does not necessarily indicate that it is a better description of ability structure. They suggested that the choice between bifactor and higher-order models should not solely depend on model fit. Morgan, Hodge, Wells, and Watkins (2015) also conducted Monte Carlo simulations to address this issue. They found that when the simulated data were consistent with the bifactor model, the fit indexes (root-mean-square error of approximation [RMSEA], TLI, etc.) favored the bifactor model. In contrast, when the simulated data were consistent with a higher-order model, the fit indexes favored several times incorrectly the bifactor model. In sum, Murray and Johnson (2013) and Morgan et al. (2015) demonstrated that bifactor model may provide better fit, because fit indices are biased.

Gignac (2016) suggested that the source of the bias in favor of the bifactor model is the proportionality constraint imposed in a higher-order model. This constraint means that with the higher-order model, the general factor loading and the specific factor loadings ratio (g/s) are constrained to be equal across all indicators within a group factor. The bifactor model does not impose this constraint. Gignac (2016) assumed that the bifactor model fits data better than the higher-order model when data are inconsistent with the proportionality constraint (Molenaar, 2016). This hypothesis implies that comparisons between higher-order and bifactor models should be based on indices that take into account the penalty for model complexity (TLI, AIC, and BIC).

However, Mansolf and Reise (2017) disagreed with Gignac's (2016) proportionality interpretation. They distinguished higher-order and bifactor models in terms of tetrad constraints. They suggested that while all models impose *rank constraints*, unique tetrad constraints are imposed in a higher-order model, but not in a bifactor one. When tetrad constraints are violated, goodness-of-fit statistics are biased in favor of the bifactor model. This tetrad constraint hypothesis could explain the results of Murray and Johnson (2013) and Gignac (2016). Following this hypothesis, Yang, Spirtes, Scheines, Reise, and Mansoff (2017) were able to reduce the statistical bias in favor of the bifactor model, by developing an algorithm for purifying or removing impure indicators. From an empirical point of view, it is important to note that a recent study indicated that the fit statistics were not systematically biased in favor of the bifactor model. In a study conducted with the WISC–IV$^{\text{U.K.}}$ (U.K. WISC–IV edition) with referred Irish children, meaningful difference between higher-order and bifactor models were not observed (Canivez, Watkins, Good, James, & James, 2017). These findings suggested that the decisions to adopt a higher-order or a bifactor model should rely on theory and on the conceptualization of *g*, not solely on fit indices.

Given the above concerns, independent examination of the structural validity of the French WISC–V was necessary to assess its construct validity. The present study addressed three goals. The first goal was to estimate the number of factors in the French WISC–V using best practices in EFA (Velicer, Eaton, & Fava, 2000). Incorrect specification of the correct number of factors can lead to poor score pattern reproduction and interpretation. Based on Canivez et al. (2016) findings with the U.S. WISC–V, it was hypothesized that the factor structure of the French WISC–V would be better described with four factors. By conducting EFA, the second goal was to ascertain the exact nature of the constructs assessed by each subtest score by estimating the relationship between every latent variable and subtest score. By applying the SLT (Schmid and Leiman, 1957), the present study also allowed the determination of the proportion of variance explained by the general factor and the proportion of variance explained by the group factors (Gignac, 2007). Based on previous studies, it was assumed that the *g* factor would explain the largest portions of total and common variance. Following the SLT procedure, the third and final goal was to test the competing theories of superordinate general intelligence versus breadth general intelligence. Using CFA, bifactor models were tested and compared with higher-order models, but unlike publisher reported analyses the present study used ML estimation in CFA.

Method

Participants

French WISC–V standardization sample raw data were requested from the publisher but access to this data set to conduct these independent analyses was denied. Absent raw data, the summary statistics (correlations and descriptive statistics) in the French WISC–V *Interpretive Manual* (Wechsler, 2016b, Table 5.1, p. 62) were used to conduct EFA and CFA. The French WISC–V standardization sample (ages 6:0 through 16:11 years) is described in the French WISC–V *Interpretive Manual* and included 1,049 participants. The sample was stratified according to age, sex, parental education level (with 6 levels), and geographic region (with 5 regions); and detailed demographic characteristics are reported. The total sample was representative of the French population, according to the general census of the population made by the Institut National de la Statistique et des Etudes Economiques (INSEE) in 2010, and was divided in 11 age groups (6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16). Each group was composed of 80 to 104 participants. Standardized scores were computed for each age group separately (M = 10; SD = 3). Because no data were directly collected in this study, no ethics committee approval was received.

Instrument

The French WISC–V is an individually administered intelligence test for children (6 to 16:11 years old). The French WISC–V FSIQ is based on the sum of seven primary subtests: BD, SI, VC, MR, FW, DS, and CD. The primary index scale level is composed of the 10 primary subtests, which are used for the estimation of the five index scores: VC, VS, FR, WM, and PS. In addition to the seven primary subtests used for the FSIQ, VP, PS, and SS are added for the estimation of the five primary indices. The FSIQ and the five indices are based on a mean of 100 and standard deviation of 15. Five ancillary index scores may be computed: QR, AWM, NV, GA, and CP.

Procedure and Analyses

EFAs were conducted using the intercorrelation matrix for the 15 primary and secondary subtests reported in Table 5.1 in the French WISC–V *Interpretive Manual* (Wechsler, 2016b, p. 62). The published matrix includes correlations rounded to only two decimals, but Carroll (1993) noted, "Little precision is lost by using two-decimal values" (p. 82). CFAs were conducted with covariance matrices reproduced from the correlation matrix and subtest standard deviations published in Table 5.1 in the *Interpretive Manual* (Wechsler, 2016b, p. 62).

Several criteria were examined to determine the number of factors to retain and included eigenvalue >1 (Kaiser, 1960), the scree test (Cattell, 1966), standard error of scree (SE_{scree}; Zoski & Jurs, 1996), parallel analysis (PA; Horn, 1965), minimum average partials (MAP; Frazier & Youngstrom, 2007; Velicer, 1976), the BIC (Schwarz, 1978), and the sample size adjusted BIC (SSBIC; Sclove, 1987). Criteria were estimated with SPSS 24 for Macintosh or with specific software. The SE_{scree} was used as programmed by Watkins (2007), while random eigenvalues for PA were produced by Monte Carlo PCA for Parallel Analysis software (Watkins, 2000) with 100 iterations to provide stable estimates. According to Glorfeld (1995), a modified PA was used to reduce the tendency to PA to overextract, and hence the eigenvalue at the 95th percentile was used as estimated by the CIeigenvalue program (Watkins, 2011). Most frequently, PA suggests retaining too few factors particularly in the presence of a strong general factor (Crawford et al., 2010).

Principal axis EFAs were conducted to analyze the factorial structure of the French WISC–V using SPSS 24 for Macintosh.[2] Retained factors were subjected to promax rotation (k = 4; Gorsuch, 1983). For a factor to be considered viable at least two subtests required salient loadings (≥.30; Child, 2006). Then, to disentangle the contribution of first and second order factors, the SLT was applied. This procedure has been extensively used and advocated by Carroll (1993). The SLT has been used in numerous studies with the WISC–IV (Watkins, 2006), the RIAS (Dombrowski et al., 2009), the WISC–V (Canivez et al., 2016), the Wechsler Abbreviated Scale of Intelligence (WASI) and the Wide Range Intelligence Test (WRIT; Canivez, Konold, Collins, & Wilson, 2009), the SB5 (Canivez, 2008), the French Wechsler Adult Intelligence Scale—Third Edition (Golay & Lecerf, 2011), and the French WISC–IV (Lecerf et al., 2011). The SLT allows for deriving a hierarchical factor model from higher-order models and decomposes the variance of each subtest score into the general factor first and then the first-order factor. The first-order factors are modeled orthogonally to each other and to the general factor (Gignac, 2006; Gorsuch, 1983). The SLT approximates the bifactor model and was produced using the MacOrtho program (Watkins, 2004). This procedure permits disentangling the common variance explained by the general factor and the residual common variance explained by the first-order factors.

The ω_H and ω_{HS} (Reise, 2012; Reise, Bonifay, & Haviland, 2013) were estimated and several authors have suggested that these coefficients are more adequate than the alpha coefficient for test scores reliability assessment (Brunner, Nagy, & Wilhelm, 2012; Gignac & Watkins, 2013). ω_H estimates the reliability of the hierarchical general intelligence factor independently of the variance of group factors. ω_{HS} estimates the reliability of group factors with general intelligence and other group factor variance removed. Omega estimates were obtained with the Omega program developed by Watkins (2013). According to Reise et al. (2013), Omega coefficients should be, at minimum, higher than .50, but .75 is better.

Finally, complementary CFAs were conducted with R-package "Lavaan" (version 05–22 with the option "mimic = MPLUS") in Rstudio for Macintosh version 1.0.136 (R Development Core Team, 2015). All higher-order models reported in the French WISC–V *Interpretive Manual* were examined with ML estimation and alternative bifactor models of all higher-order models were also tested. Multiple indicators of approximate fit were considered to assess competing models (Hu & Bentler, 1999). The chi-square (χ^2) statistic, the RMSEA, and the standardized-root-mean square residuals (SRMR), which expresses the degree of fit between the covariance matrix of the observed data and the covariance matrix predicted by the model, were used as primary fit indices (Byrne, 2001). The TLI, which is relatively unrelated to sample size, and the comparative fit index (CFI) were also used to evaluate model fit. Contemporary criteria for evaluating fit were applied with values of >.95 for CFI and TLI, and <.06 for RMSEA and <.08 for SRMR representing good model fit (Hu & Bentler, 1999). The AIC and the BIC were also used to compare models: the smaller AIC suggests the better model most likely to replicate (Kline, 2016). In the French WISC–V *Interpretive Manual*, $\Delta\chi^2$ was the sole criterion used to compare models. However, because it has been demonstrated that $\Delta\chi^2$ is sensitive to large sample size, the ΔCFI and ΔRMSEA were also examined. For a model to be considered superior, it had to exhibit adequate to good overall fit and display meaningfully better fit (ΔCFI > .01 and ΔRMSEA > .015) than alternative models (Chen, 2007; Cheung & Rensvold, 2002). AIC and BIC were also reported in the French WISC–V *Interpretive Manual*, but not used for model comparisons.

Results

EFAs

Examination of multiple criteria to determine the number of factors to retain (Ruscio & Roche, 2012; Velicer et al., 2000) found MAP to suggest one factor, visual scree and Horn's parallel analysis (HPA) suggested two factors, eigenvalue >1 and SE_{scree} suggested three factors, BIC and SSBIC suggested four factors, and the publisher (theory) proposed structure promoted five factors. Although none of the objective extraction criteria suggested more than four factors, EFA began with the extraction of five factors based on the suggested factorial structure proposed by the publisher of the French WISC–V. Extraction of five factors was also examined because it is better to overextract than underextract to examine performance of smaller factors. As stated by Wood, Tataryn, & Gorsuch, (1996, p. 354), "avoid underextraction, even at the risk of overextraction." Subsequently, models with four, three, and two factors were sequentially examined.

Results of EFA with five extracted factors and promax rotation (k = 4) are provided in supplementary Table S1. Data indicated that the fifth factor included only one salient pattern loading (AR), which does not satisfy the basic requirement that each factor should be marked by at least two salient factor pattern coefficients. SI, VC, IN, and CO loaded on a VC factor; BD, VP, MR, and FW loaded on a PR factor; CD, SS, and CA loaded on a PS factor; and DS, LNS, and PS loaded on a WM factor. Separate VS and FR factors did not emerge when forcing extraction of five factors.

Table 1 presents the results of extraction of four factors with promax rotation (k = 4). In this model, SI, VC, IN, and CO loaded on a VC factor; BD, VP, MR, and FW loaded on a PR factor (AR also had a secondary cross-loading [.302] on this factor); CD, SS, and CA loaded on a PS factor; and DS, LNS, and AR loaded on a WM factor. Given standard error, the PS subtest could be considered having achieved a salient factor pattern coefficient on the WM factor (.291). The g loadings ranged from .327 (CA) to .724 (SI) and all were within the fair to good range (except CD, SS, and CA) based on Kaufman's (1994) criteria (≥.70 = good, .50–.69 = fair, <.50 = poor).

Results of EFA with three-factors and two-factors with promax rotation (k = 4) are provided in supplementary Table S2. Results of three-factor extraction found the PR and the WM factors merged, while VC and PS factors were still distinct. In the two-factor extraction, VC, PR, and WM factors merged and a separate PS factor emerged. These models display a fusion of theoretically meaningful constructs that is a likely result of underextraction,

[2] Although not reported here, analyses were also conducted with R-package ("psych" version 1.6.9), and results were similar.

Table 1
French Wechsler Intelligence Scale for Children—Fifth Edition (French WISC–V) Four Oblique Factor Solution for the Total Standardization Sample

French WISC–V subtest	General	F1: Verbal Comprehension	F2: Perceptual Reasoning	F3: Processing Speed	F4: Working Memory	h^2
SI	.724	**.672** (.481)	.135 (.609)	−.065 (.304)	.058 (.616)	.623
VC	.666	**.918** (.809)	−.054 (.503)	−.034 (.271)	−.080 (.531)	.664
IN	.716	**.611** (.749)	.149 (.607)	.005 (.350)	.045 (.604)	.577
CO	.621	**.707** (.698)	−.091 (.463)	.120 (.364)	.008(.513)	.499
BD	.652	−.013 (.478)	**.733** (.719)	.121 (.444)	−.088 (.514)	.530
VP	.719	−.058 (.521)	**.922** (.825)	.015 (.416)	−.086 (.573)	.687
MR	.683	.087 (.557)	**.589** (.710)	−.057 (.337)	.120 (.598)	.519
FW	.628	.130 (.534)	**.462** (.632)	−.090 (.281)	.167 (.565)	.431
AR	.708	.089 (.583)	*.302* (.661)	.031 (.407)	**.374** (.682)	.520
DS	.686	−.036 (.553)	.035 (.588)	−.074 (.336)	**.818** (.781)	.615
PS	.592	.021 (.458)	.225 (.546)	.173 (.440)	.291 (.561)	.372
LNS	.716	.038 (.591)	−.084 (.579)	.055 (.430)	**.809** (.801)	.646
CD	.446	.055 (.307)	−.057 (.344)	**.698** (.697)	.012 (.356)	.488
SS	.491	−.014 (.314)	.015 (.401)	**.758** (.768)	.015 (.393)	.589
CA	.327	−.024 (.203)	.106 (.291)	**.467** (.488)	−.047 (.247)	.242
Eigenvalue		6.49	1.46	1.02	.80	
Factor correlations		F1: VC	F2: PR	F3: PS	F4: WM	
Verbal Comprehension (VC)						
Perceptual Reasoning (PR)		.690				
Processing Speed (PS)		.404	.507			
Working Memory (WM)		.727	.752	.497		

Note. N = 1,049. French WISC–V subtests: SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; PS = Picture Span; LNS = Letter-Number Sequencing; CD = Coding; SS = Symbol Search; CA = Cancellation. Salient pattern coefficients (≥.30) presented in bold (structure coefficient). h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (g loadings).

thereby producing unsatisfactory representations (Gorsuch, 1983; Wood et al., 1996).

Hierarchical EFA: Four French WISC–V Factors SLT Bifactor

Based on the present EFA results, the four-factor EFA was the most reasonable fit to theory and psychometric standards and was subjected to higher-order EFA and the SLT procedure.Table 2 presents results from SLT of the four extracted factors. Results indicated that all subtests were consistently associated with their theoretically proposed factor, except PS and AR. AR had relatively similar residual group factor loadings on WM (.170) and on PR (.151). PS had relatively similar residual group factor loadings on PS (.144), WM (.132), and PR (.113).

The g factor accounted for 35.7% of the total variance and 67.0% of the common variance. This finding is consistent with the presence of a general intelligence factor. Regarding subtests, the g factor accounted for between 8.4% (CA) and 50.3% (LNS) of individual subtest variability. According to Kaufman's (1994) criteria, only the LNS subtest had a "good" g loading (≥.70). This LNS loading is relatively close to the one reported by Canivez et al. (2016) with the U.S WISC–V (.69). However, in the U.S. WISC–V, the higher g loading was for VC (.774).

At the group factor level, smaller portions of additional common variance were provided by VC (10.2%), PR (6.8%), WM (4.2%), and PS (11.9%). The combination of the general factor and group factors measured 53.3% of the common variance; hence, 46.7% of the French WISC–V variance is unique (a combination of specific and error variance). CA and PS were heavily influenced by unique variance (76.2% and 62.9%, respectively). The ω_H and ω_{HS} coefficients were estimated from the SLT results. The ω_H coefficient for the general factor was high (.831). The ω_{HS} coefficients for the four group factors were lower and ranged from .108 (WM) to .468 (PS). Thus, the four French WISC–V group factors suggested by EFA would produce unit-weighted composites that likely possess too little unique true score variance for confident clinical interpretation (Reise, 2012; Reise et al., 2013).

For comparison purposes, bifactor ESEM analyses were conducted with Mplus 7.4 (with ML estimation and bigeomin rotation, orthogonal rotation; Muthén & Muthén). Model comparisons indicated that the best model included a general factor and three group factors (CFI = .992, TLI = .984): VC (SI [.415], VO [.576], IN [.365], CO [.438]); PS (CD [.591], SS [.631], CA [.388], PS [.127]); and VS/WM (BD [.262], VP [.321], MR [.126], DS [−.316], LNS [−.337]). CO and BD also loaded weakly on PS (.092 and .081, respectively). French WISC–V g loadings varied from .294 (CA) to .76 (VP). These findings are not consistent with the favored model reported in the French WISC–V *Interpretive Manual*. It is important to note that ESEM with an oblique rotation revealed a similar model with a general factor and 3 group factors: VC, PS, and WM/VS.

CFAs

CFAs were conducted to replicate and extend the data analyses reported in the French WISC–V *Interpretive Manual* (Wechsler,

Table 2
Sources of Variance in the French Wechsler Intelligence Scale for Children—Fifth Edition (French WISC–V) for the Total Standardization Sample According to the Schmid-Leiman Bifactor Model (Orthogonalized Higher-Order Factor Model) With Four First-Order Factors

	General		F1: Verbal Comprehension		F2: Perceptual Reasoning		F3: Processing Speed		F4: Working Memory			
French WISC–V subtest	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	*b*	S^2	h^2	u^2
Similarities	.666	.444	**.409**	**.167**	.068	.005	−.054	.003	.026	.001	.619	.381
Vocabulary	.592	.350	**.558**	**.311**	−.027	.001	−.028	.001	−.036	.001	.665	.335
Information	.657	.432	**.371**	**.138**	.075	.006	.004	.000	.020	.000	.575	.425
Comprehension	.556	.309	**.430**	**.185**	−.046	.002	.100	.010	.004	.000	.506	.494
Block Design	.613	.376	−.008	.000	**.367**	**.135**	.101	.010	−.040	.002	.522	.478
Visual Puzzles	.684	.468	−.035	.001	**.461**	**.213**	.012	.000	−.039	.002	.683	.317
Matrix Reasoning	.655	.429	.053	.003	**.295**	**.087**	−.047	.002	.054	.003	.524	.476
Figure Weights	.602	.362	.079	.006	**.231**	**.053**	−.075	.006	.076	.006	.433	.567
Arithmetic	.683	.466	.054	.003	.151	.023	.026	.001	**.170**	**.029**	.522	.478
Digit Span	.690	.476	−.022	.000	.018	.000	−.062	.004	**.371**	**.138**	.618	.382
Picture Span	.566	.320	.013	.000	.113	.013	*.144*	*.021*	**.132**	**.017**	.371	.629
Letter-Number Sequencing	.709	.503	.023	.001	−.042	.002	.046	.002	**.367**	**.135**	.642	.358
Coding	.391	.153	.033	.001	−.029	.001	**.582**	**.339**	.005	.000	.494	.506
Symbol Search	.434	.188	−.009	.000	.008	.000	**.632**	**.399**	.007	.000	.588	.412
Cancellation	.289	.084	−.015	.000	.053	.003	**.389**	**.151**	−.021	.000	.238	.762
Total variance		.357		.054		.036		.063		.022	.533	.467
Explained common variance		.670		.102		.068		.119		.042		
Picture Span on WM	ω_H = .831		ω_{HS} = .287		ω_{HS} = .179		ω_{HS} = .468		ω_{HS} = .108			
Picture Span on PS	ω_H = .833		ω_{HS} = .186		ω_{HS} = .178		ω_{HS} = .374		ω_{HS} = .220			

Note. N = 1,049. WM = Working Memory; PS = Processing Speed; b = loading of subtest on factor; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; ω_H = omega-hierarchical; ω_{HS} = omega-hierarchical subscale. Bold type indicates coefficients and variance estimates consistent with the theoretically proposed factor. Italic type indicates coefficients and variance estimates associated with an alternate factor (where residual cross-loading b was larger than for the theoretically assigned factor).

2016b, pp. 67–69). Therefore, the procedure and the models reported in the French WISC–V *Interpretive Manual* were used as "baseline." CFAs reported in the French WISC–V *Interpretive Manual* were conducted using the raw data; however the present analyses were conducted on the reproduced covariance matrix derived from the French WISC–V subtest correlation matrix and descriptive statistics because Pearson France declined the request to provide raw data for independent analyses. In addition, CFAs in the present study used ML estimation rather than WLS estimation used by the publisher. For all higher-order models, the *df* obtained in the present study were identical to those reported in the French WISC–V *Interpretive Manual*. In addition, the approximate fit values obtained in the present study were similar to those reported in the French WISC–V *Interpretive Manual* (CFI, TLI, RMSEA). This first step is crucial because it provided evidence that present results conducted with the correlation matrix instead of the raw data replicated those reported in the French WISC–V *Interpretive Manual*. Table 3 presents fit statistics for higher-order models and alternative bifactor representations for direct comparison.

Higher-order models. Table 3 illustrates the progressively improved model fit for models with one through five first-order factors. Results indicated that approximate fit indices (CFI, TLI, RMSEA, and SRMR) were relatively similar for all higher-order models, except for models with one, two, and three group factors. Models 1, 2, and 3 were inadequate based on CFI and TLI ($<$.95) and RMSEA ($>$.06). This finding indicated that the structure suggested by HPA (2 factors) or MAP (1 factor) did not fit these data. For higher-order models with 4 and 5 first-order factors, RMSEA ranged from .052 to .058 and SRMR ranged from .033 to .034. Overall, fit statistics for all higher-order models with 4 or 5 group factors fit these data well, but numerous problems were observed in many models with nonsignificant standardized paths, factors with negative variance, and standardized paths ≥1.0 (local assessment, see Table 3). Thus, local fit assessment indicated that Models 4a, 4c, and 5a should be retained. ΔCFI and ΔTLI indicated that there were no meaningful differences (ΔCFI $>$ .01 and ΔTLI $>$ .010), between the four-factor higher-order models and the five-factor higher-order models. The AIC was slightly lower for Model 5c with five first-order factors. In the French WISC–V *Interpretive Manual*, the AIC was also lower for the Model 5c. In this model, all loadings were statistically significant. SI, VC, IN, and CO loaded on VC; BD and VP loaded on VS; MR, FW, and AR loaded on FR; DS, LNS, PS, and AR loaded on WM; and CD, SS, and CA loaded on PS. There is only one cross-loading for AR (WM and FR). However, FR variance was negative in this model and the AR loading on WM was weak (.22), although statistically significant. Likewise, Model 4c corresponded to the results from the present EFA and was psychometrically appropriate. In this model, all loadings were statistically significant. SI, VC, IN, and CO loaded on VC; BD, VP, MR, FW, and AR loaded on PR; DS, LNS, PS, and AR loaded on WM; and CD, SS, and CA loaded on PS. There is only one cross-loading for AR (WM and PR). Finally, examination of Model 5e, the favored model in the French WISC–V *Interpretive Manual*, indicated that the Arithmetic loading with VC was not statistically significant (.025). Hence, Model 5e was *not* the best model for the French WISC–V.

Bifactor models. Bifactor models examined in the present study included all subtests loading directly onto a general factor

Table 3
CFA Fit Statistics for French WISC–V 15 Subtests for the Total Standardization Sample

Models	χ^2	*df*	AIC	BIC	RMSEA	90% CI RMSEA	SRMR	TLI	CFI
Higher-order models									
Model 1 (1 factor + *g*)	1136.59	90	72,942.18	73,165.19	.105	[.100, .111]	.063	.817	.843
Model 2 (2 factors + *g*)	915.18	86	72,728.77	72,971.60	.096	[.089, .100]	.058	.848	.876
Model 3 (3 factors + *g*)	598.30	84	72,415.89	72,668.63	.076	[.069, .081]	.043	.904	.923
Model 4a (4 factors + *g*)	334.19	82	72,155.78	72,418.43	.054	[.048, .060]	.034	.952	.962
Model 4b (4 factors + *g*)[a]	368.82	82	72,190.42	72,453.06	.058	[.052, .064]	.034	.945	.957
Model 4c (4 factors + *g*)	306.53	81	72,130.12	72,397.72	.052	[.045, .058]	.033	.956	.966
Model 4d (4 factors + *g*)[b]	303.62	80	72,129.21	72,401.77	.052	[.046, .058]	.033	.956	.967
Model 5a (5 factors + *g*)	330.16	80	72,155.76	72,428.31	.055	[.049, .061]	.034	.951	.963
Model 5b (5 factors + *g*)[c]	304.98	80	72,130.57	72,403.13	.052	[.046, .058]	.034	.956	.966
Model 5c (5 factors + *g*)[c]	299.71	79	72,127.30	72,404.81	.052	[.045, .058]	.034	.956	.967
Model 5d (5 factors + *g*)[d]	320.39	79	72,147.98	72,425.50	.054	[.048, .060]	.034	.952	.964
Model 5e (5 factors + *g*)[e]	299.55	78	72,129.14	72,411.61	.052	[.046, .058]	.034	.955	.967
Bifactor models									
Model 2 (2 factors + *g*)[f]	355.30	75	72,190.89	72,488.22	.060	[.054, .066]	.029	.941	.958
Model 3 (3 factors + *g*)[g]	257.37	75	72,092.96	72,390.30	.048	[.042, .055]	.030	.962	.973
Model 4a (4 factors + *g*)[h]	**213.06**	**75**	**72,048.66**	**72,345.99**	**.042**	**[.035, .049]**	**.027**	**.971**	**.979**
Model 4b (4 factors + *g*)[i]	233.75	76	72,067.34	72,359.72	.044	[.038, .051]	.029	.967	.976
Model 4c (4 factors + *g*)[j]	207.04	74	72,044.63	72,346.92	.041	[.035, .048]	.027	.972	.980
Model 4d (4 factors + *g*)[k]	206.60	73	72,046.19	72,353.44	.042	[.035, .049]	.027	.971	.980
Model 5a (5 factors + *g*)[l]	237.93	77	72,069.52	72,356.94	.045	[.038, .051]	.029	.967	.976
Model 5b (5 factors + *g*)[m]	249.18	76	72,082.77	72,375.15	.046	[.040, .053]	.029	.964	.974
Model 5c (5 factors + *g*)[n]	232.76	75	72,068.36	72,365.69	.044	[.038, .051]	.029	.967	.976
Model 5d (5 factors + *g*)[o]	237.93	77	72,069.52	72,356.94	.045	[.039, .052]	.029	.967	.976
Model 5e (5 factors + *g*)[p]	232.69	75	72,068.28	72,365.62	.045	[.038, .051]	.029	.967	.976

Note. $N = 1{,}049$. *df* = degrees of freedom; AIC = Akaike's information criterion; BIC = Bayesian information criterion; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square; TLI = Tucker-Lewis index; CFI = comparative fit index; *g* = general intelligence. Bold text indicates best fitting model.
[a] Working Memory (WM) loading on g > 1, hence negative WM variance. [b] Arithmetic loading on Verbal Comprehension (VC) not significant. [c] Fluid Reasoning (FR) loading on g > 1, hence negative FR variance. [d] Arithmetic loading on VC significant but weak (.16). [e] Arithmetic loading on VC not significant; FR loading on g > 1, hence negative FR variance. [f] Visual Puzzles loading on F2 not significant, some loadings weak and negative (Matrix Reasoning [MR]–F2; Figure Weights [FW]–F2). [g] Arithmetic loading on F1 not significant, some loadings (Digit Span [DS]–F1; Letter Number Sequencing [LNS]–F1) were negative. [h] FW loading on Perceptual Reasoning (PR) was weak (.10), Arithmetic and Picture Span loadings on WM were weak. [i] Loading of FW on WM was not significant, loadings of Arithmetic and Picture Span on WM were weak, negative loading of MR on WM. [j] Loading of Picture Span on WM was not significant, Arithmetic loading on WM and PR were significant but weak (.14 and .10). [k] Arithmetic loading on VC and Picture Span loading on WM not significant. [l] FW loading on FR (.14) and MR loading on FR (.14) were weak, Arithmetic loading on WM (.14) was weak. [m] MR and Arithmetic loadings on FR not significant, Picture Span loading on WM was weak. [n] MR loading on FR not significant, Arithmetic loading on FR (.06) was weak, Picture Span loading on WM (.12) was weak. [o] FW (.14) and MR (.14) loadings on FR were weak, Picture Span loading on WM (.12) was weak. [p] FW was fixed, model did not converge, Arithmetic loading on VC and MR loading on FR not significant, Arithmetic loading on FR (.06) was weak.

and also onto one group factor. All bifactor models demonstrated good fit to these data (see Table 3) according to RMSEA, SRMR, TLI, and CFI (except TLI for bifactor Model 2). For models with four and five factors, CFI ranged from .974 (Model 5b) to .980 (Models 4c and 4d); TLI ranged from .964 (Model 5b) to .972 (Model 4c); RMSEA ranged from .041 (Model 4c) to .046 (Model 5b); and SRMR ranged from .027 to .029; thus, there were no meaningful differences between these models (ΔRMSEA, ΔCFI and ΔTLI; Chen, 2007; Cheung & Rensvold, 2002). According to AIC, the best model is the bifactor Model 4c (Figure 1), which corresponds to the result of the present EFA with SI, VC, IN, and CO loading on VC; BD, VP, MR, AR, and FW loading on PR; AR, DS, PS, and LNS loading on WM; and CD, SS, and CA loading on PS. In bifactor Model 4c, the PS loading on WM was not statistically significant and AR loadings on WM and on PR were weak. The loadings of MR and FW on PR were also weak (.206 and .130, respectively). However, it is important to note that in contrast with higher-order models, nonsignificant paths are not so dramatic with the bifactor model, because it indicates that the variance of the subtest score was explained by the general factor rather than by the first-order factors. The modification of Model 4c with reestimation of parameters after dropping the nonsignificant PS to WM path is provided in supplementary Figure S3.

Overall, this finding was consistent with SLT, because subtests loadings on the first-order factors were lower than subtest loadings on the general factor, except for CD, SS, and CA. For instance, in the bifactor Model 4c (supplemental material Figure S3), the *g* loadings for LNS (.703), AR (.701), SI (.686), IN (.684), DS (.679), MR (.678), and VP (.671) were very high. This result suggested that these subtests scores were primarily measures of the *g* factor. Conversely, CD (.396), SS (.437), and CA (.299) showed less salient direct loadings from *g*.

The second best bifactor model was Model 4a, which had a very close AIC estimate to Model 4c (ΔAIC = 4.03) and is illustrated in Figure 2. This model offers the advantage of simple structure and subtest alignment that matches earlier theoretical postulation. Given the lack of meaningful difference between Model 4a and Model 4c an argument could be made that this should be the

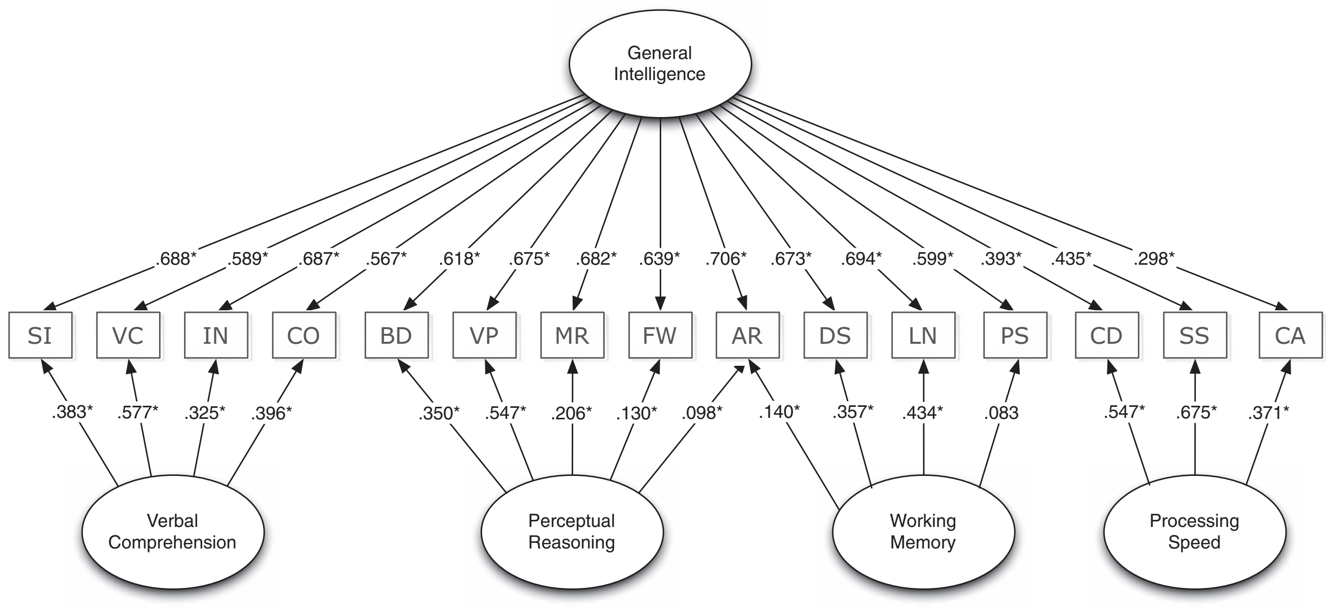

Figure 1. Bifactor measurement model (Model 4c bifactor), with standardized coefficients, for the French WISC–V standardization sample (N = 1,049) 15 subtests. SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; LN = Letter-Number Sequencing; PS = Picture Span; CD = Coding; SS = Symbol Search; CA = Cancellation. * $p < .05$.

preferred model for the French WISC–V. In comparing Model 4a and Model 4c, when AR is cross-loaded (Model 4c) it effectively diminishes the model variance of the PS subtest with WM where it is not statistically significant and thus forcing it to be dropped. When AR is not cross-loaded (Model 4a) then PS has a small yet statistically significant relation with WM.

Table 4 presents the decomposed variance sources for bifactor Model 4a and parallels that of Table 2. The g factor accounted for 37.2% of the total variance and 69.6% of the common variance. This finding is consistent with the presence of a general intelligence factor. Regarding subtests, the g factor accounted for between 8.8% (CA) and 53.1% (AR) of individual subtest variability. At the group factor level, smaller portions of additional common variance were provided by VC (9.6%), PR (5.3%), WM (4.4%), and PS (11.2%). The combination of the general factor and group factors measured 53.5% of the common variance; hence 46.5% of the French WISC–V variance is unique (a combination of specific and error variance). CA and PS were heavily influenced by unique variance (77.4% and 63.9%, respectively). The ω_H and ω_{HS} coefficients were estimated from bifactor Model 4a results. The ω_H coefficient for the general factor was high (.844). The ω_{HS} coefficients for the four group factors were lower and ranged from .100 (WM) to .464 (PS). Thus, the four French WISC–V group factors suggested by CFA would produce unit-weighted composite scores that likely possess too little unique true score variance for confident clinical interpretation (Reise, 2012; Reise et al., 2013).

Although higher-order and bifactor models achieved adequate fit to these WISC–V data (i.e., TLI and CFI >.95, RMSEA and SRMR <.06), the results of the present investigation suggested that bifactor models fit better than the corresponding higher-order model (for Models 4a, 4c, and 5a, ΔCFI and ΔTLI >.01); the difference was not meaningful using ΔRMSEA (<.015). Likewise, the bifactor models produced lower AIC than their corresponding higher-order models. This result suggested that there are some benefits to examine and describe bifactor models (Reise, 2012).

Discussion

According to the French WISC–V publisher, their CFA supported a model with one second-order factor (g) and five first-order factors, and included three AR cross loadings, matching that of the U.S. WISC–V (Wechsler, 2014). However, there are numerous concerns regarding this French WISC–V factor structure based on the CFAs reported in the French WISC–V *Interpretive Manual,* as well as those reported in the U.S. WISC–V *Technical and Interpretive Manual,* and undisclosed and nonstandard methods (Beaujean, 2016; Canivez et al., 2016, 2017). Therefore, the French WISC–V factor structure was independently examined in the present study using best practices in EFA, ESEM, and CFA. For theoretical reasons, the WISC–V structure was examined with both bifactor and higher-order models. While the publisher denied access to the French WISC–V standardization sample raw data for independent analyses, the availability of the 15 French WISC–V subtest correlation matrix in the *Interpretive Manual* permitted examination of EFA and reproduction of the covariance matrix for use in CFA.

EFA results indicated that a model with five factors was inadequate because the fifth factor contained only one subtest indicator with a salient pattern coefficient (AR). Instead, EFA indicated that a four-factor solution was most plausible, that included the familiar

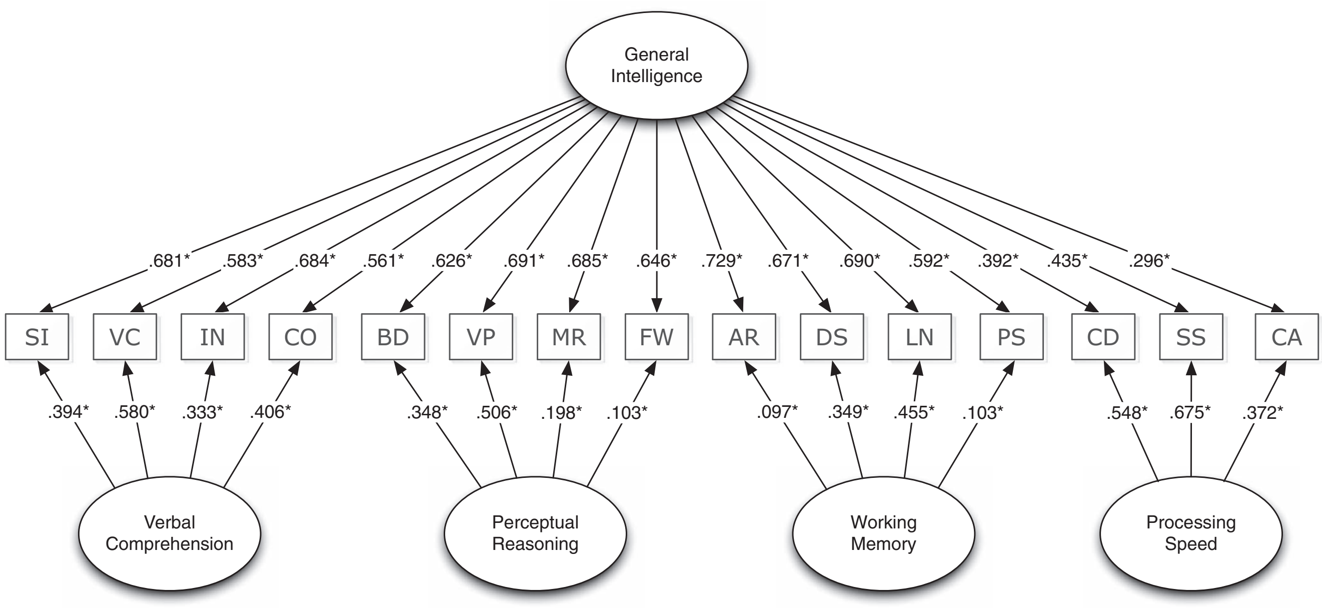

Figure 2. Bifactor measurement model (Model 4a bifactor), with standardized coefficients, for the French WISC–V standardization sample ($N = 1{,}049$) 15 subtests. SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; LN = Letter-Number Sequencing; PS = Picture Span; CD = Coding; SS = Symbol Search; CA = Cancellation. $^{*}\ p < .05$.

VC, WM, PS, and PR as found in the U.S. WISC–V (Canivez et al., 2016). Likewise, ESEM with bigeomin rotations (Morin, Arens, et al., 2016, Morin, Arens, Tran, et al., 2016) *did not* identify five group factors, but rather indicated that a bifactor model with three factors was the most plausible and included VC, PS, and a mixture of VS and WM. The SLT applied to the four-factor EFA indicated that the *g* factor explained most of the common and total variance in the current French WISC–V as previously observed by Canivez et al. (2016). CFA results suggested that the bifactor model with four first-order group factors provided better fit to these French WISC–V data than the higher-order model, replicating the results of Canivez et al. (2017) with the U.S. WISC–V.

EFA indicated that in the five-factor model, the fifth factor contained only the AR subtest, while in the U.S. WISC–V the fifth factor contained only the FW subtest (Canivez et al., 2016). In the French WISC–V the AR score clearly was neither associated with the FR factor nor with the WM factor. In the four first-order factor model, AR *was* associated with WM and PR. Extracting more factors than appropriate may have stripped variance away from legitimate factors to support the fifth factor. These findings could suggest that AR is mainly a measure of another factor, which may be quantitative reasoning (Gq), as was suggested with the U.S. WISC–III (Watkins, & Ravert, 2013) and the French WISC–IV (Lecerf, Rossier, Favez, Reverte, & Coleaux, 2010). Problems with continued inclusion of AR in the WISC without additional quantitative reasoning tasks have been noted (Canivez & Kush, 2013).

In addition, neither the five- nor four-factor models showed evidence for the distinction between VS and PR factors. There was no separation of BD and VP into a VS factor and MR and FW into a FR factor. These four subtests combined into the familiar PR factor observed in earlier Wechsler scales and the U.S. WISC–V (Canivez et al., 2016). This finding indicated that the separation of FR and VS was unsuccessful in both the U.S. WISC–V and in the French WISC–V. Therefore, separate VSI and FRI scores are likely misleading. If separate VSI and FRI scores are important and to be used in clinical assessments it is necessary to develop tasks which more clearly separate the visual-spatial and the fluid reasoning components (if this can actually be accomplished). For instance, because the MR subtest requires inductive reasoning (Gf-I) with perceptual patterns, the overlap might be too great to allow a clear distinction between fluid reasoning and visual processing. Thus, *mechanically* interpreting the distinction between Gf and Gv factors using VSI and FRI cannot be recommended.

Bifactor ESEM with bigeomin rotations (Morin, Arens, et al., 2016, Morin, Arens, Tran, et al., 2016) revealed similar results, as models with five, six, and seven factors did not converge. Results indicated that a bifactor model with three group factors was the most plausible, and included VC, PS, and a mixture of VS and WM. Indeed, with an oblique rotation, LNS, DS, AR, and PS loaded positively on this third factor, while VP and BD loaded negatively. As this bipolar factor is rotated, it was assumed that the positive loadings of LNS, DS, AR, and PS and the negative loadings of VP and BD may result from the contribution of Working Memory Capacity (WMC). BD and VP involve WMC. MR loaded weakly on this factor only with an orthogonal rotation, while FW did not load on this VS/WM factor either with orthogonal or with oblique rotation. As mentioned previously, this result might suggest that MR relies on FR and visual processing, like BD

Table 4

Sources of Variance in the French Wechsler Intelligence Scale for Children—Fifth Edition (French WISC–V) for the Total Standardization Sample According to the CFA Bifactor Model With Four First-Order Factors (Model 4a Bifactor)

French WISC–V subtest	General		Verbal Comprehension		Perceptual Reasoning		Working Memory		Processing Speed				
	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2	h^2	u^2	ECV
Similarities	.681	.464	.394	.155							.619	.381	.749
Vocabulary	.583	.340	.580	.336							.676	.324	.503
Information	.684	.468	.333	.111							.579	.421	.808
Comprehension	.561	.315	.406	.165							.480	.520	.656
Block Design	.626	.392			.348	.121					.513	.487	.764
Visual Puzzles	.691	.477			.506	.256					.734	.266	.651
Matrix Reasoning	.685	.469			.198	.039					.508	.492	.923
Figure Weights	.646	.417			.103	.011					.428	.572	.975
Arithmetic	.729	.531					.097	.009			.541	.459	.983
Digit Span	.671	.450					.349	.122			.572	.428	.787
Letter-Number Sequencing	.690	.476					.455	.207			.683	.317	.697
Picture Span	.592	.350					.103	.011			.361	.639	.971
Coding	.392	.154							.548	.300	.454	.546	.338
Symbol Search	.435	.189							.675	.456	.645	.355	.293
Cancellation	.296	.088							.372	.138	.226	.774	.388
Total variance		.372		.051		.028		.023		.060	.535	.465	
ECV		.696		.096		.053		.044		.112			
	ω_H = .844		ω_{HS} = .270		ω_{HS} = .131		ω_{HS} = .100		ω_{HS} = .464				

Note. N = 1,049. b = standardized loading of subtest on facto; S^2 = variance explained; h^2 = communality; u^2 = uniqueness; ECV = explained common variance; ω_H = omega-hierarchical (general factor); ω_{HS} = omega-hierarchical subscale (group factors).

and VP. It can also be hypothesized that MR loaded on this factor because it relies on WM. Several studies have shown strong correlations between WM and FR. Chuderski and Necka (2012), for instance, revealed that storage capacity (i.e., WMC) predicted fluid intelligence more than did executive control. However, in the present study, FW, which is also considered a fluid intelligence task (a quantitative reasoning task), did not load on this factor. Therefore, it might be assumed that the BD, VP, and MR loadings might reflect the contribution of visual processing rather than the contribution of working memory. The implication of both EFA and ESEM results is that the French WISC–V appears to be overfactored as promoted by the publisher and that the role of broad and specific abilities in subtest scores are inflated or overestimated. This calls into question the interpretation procedures promoted the French WISC–V publisher.

The SLT applied to the four-factor EFA and the examination of model-based reliability coefficients (ω_H and ω_{HS} coefficients) indicated that the *g* factor accounted for the largest portion of French WISC–V variance. These findings were consistent with results obtained by Canivez et al. (2016) with the U.S. WISC–V, and also with other Wechsler scales and with other intelligence test batteries like SB5, WASI, WRIT, and RIAS (Canivez, 2008; Canivez et al., 2009; Canivez & Watkins, 2010; Cucina & Howardson, 2017; Dombrowski et al., 2009; Golay & Lecerf, 2011; Watkins et al., 2006). The four first-order factor model results with the French WISC–V were very similar to those obtained with the U.S. WISC–V. Canivez et al. (2016) found that VC explained 9.2% (vs. 10.2% here), PR 5.6% (vs. 6.8), WM 6.5% (vs. 10.8%), and PS 11.6% (vs. 10.8) of the common variance. The SLT with four first-order factors conducted on the French WISC–V were also relatively consistent with results obtained with the previous version of the French WISC, the French WISC–IV (Wechsler, 2005). The *g* factor explained 67.0% of the common variance in the current French WISC–V, while it was 60.3% in the French WISC–IV (Lecerf et al., 2011). The portion of variance accounted for by the *g* factor is thus higher in the new French WISC–V. Furthermore, in the previous French WISC–IV the common variance explained by the VC was 14.3% (vs. 10.2% in the French WISC–V), 12.9% for PS (vs. 11.9%), 7.2% for WM (4.2%), and 5.4% for PR (vs. 6.8%).

Model-based reliability coefficients (ω_H and ω_{HS} coefficients) estimated in both French WISC–V EFA and CFA indicated that when *g* variance is removed, the unique contributions of the broad abilities were quite limited. The ω_H coefficient for the general factor was high, and hence a unit-weighted composite score based on these indicators would be satisfactory for confident interpretation. The ω_{HS} coefficients for the four group factors were considerably lower, failing to achieve the recommended minimum standard of .50 (Reise, 2012; Reise et al., 2013). These findings were consistent with Canivez et al. (2016) who found in the U.S. WISC–V that ω_{HS} ranged from .109 (PR) to .516 (PS). This indicates that unit-weighted composite scores derived from subtest indicators for VC, PR, WM, and PS likely contain too little unique true-score variance for confident interpretation (Reise, 2012; Reise et al., 2013). Thus, ω_{HS} in the present study were also not high enough in the French WISC–V to allow individual interpretation, even for PS. This supports a perspective more consistent with Carroll's three-stratum model than with the Cattell-Horn *extended Gf-Gc* model. Indeed, while Cattell-Horn excluded the *g* factor and considered it a statistical artifact, Carroll demonstrated the importance of the *g* factor. Likewise, Carroll suggested that subtest score is explained first by *g*, then by one or more broad ability, then by one or more narrow ability, and finally by unique variance. Although several broad abilities exist independently of the *g* factor, it appears that they are difficult to measure with appropriate level of precision. That is one reason why some authors defend Carroll's model rather than the Cattell-Horn model (Cucina & Howardson, 2017).

The present CFA indicated that the French WISC–V bifactor model provided better fit to these data than the higher-order model (ΔTLI, ΔCFI, and AIC) and that the bifactor model with four rather than five first-order group factors better described the latent structure of the French WISC–V. This finding was consistent with analyses conducted on the U.S. WISC–V (Canivez et al., 2017), but was different than that reported by Chen et al. (2015). However, it is important to consider Murray and Johnson's (2013); Gignac's (2016), and Mansolf and Reise's (2017) suggestions that bifactor models might benefit from statistical bias, due to the proportionality constraint or tetrad constraints. Therefore, although the bifactor model fits better, it does not necessarily indicate that it is a better description of ability structure. However, Murray and Johnson concluded that when there is an attempt to estimate or account for domain-specific abilities, something specifically recommended by the French WISC–V publisher, the "bifactor model factor scores should be preferred" (Murray & Johnson, 2013, p. 420). Our preference for a bifactor model is based also on theoretical perspective. In the bifactor model, the general factor and the first-order factors *directly* influence the subtest scores and is consistent with Carroll (1993) and Spearman (1927) perspectives. The *g* factor has no direct effects on the first-order factors, which are modeled orthogonally to each other. The bifactor model tests the presence of a global construct underlying all indicators and the coexistence of specific factors (group factor). Thus, the bifactor model is a breadth factor that permits multidimensionality by determining how broad abilities perform independent of the *g* factor. With the higher-order model, the broad abilities fully mediate the effect of *g* on the subtest scores. The *g* factor influences the subtest scores through the first-order factors. Thus, the bifactor model was considered more appropriate, because the conceptualization of the general factor as a breadth factor is preferable to its conceptualization as superordinate factor (Gignac, 2008). In our view, the relationship between the *g* factor and each subtest is not mediated by the first-order broad abilities.

In summary, the results of the present study indicated that the French WISC–V is overfactored when including five first-order factors. Results indicated that the higher-order model preferred by the publisher of the French WISC–V incorrectly concludes that the broad abilities provide useful information distinct from *g*. By reporting only higher-order models, the French WISC–V publishers overestimate the role of broad and specific abilities in subtest scores. This "overfactoring" could be due to the variance general factor's omission, and/or due to failing to consider use of EFA to inform latent structure and forcing their preconceived five-factor model. In contrast, the present results indicated that the French WISC–V is primarily a measure of *g*, because it accounts for substantially larger portions of common and total subtest variance, and supports the primary interpretation of the FSIQ. Although the FSIQ is not strictly equivalent to the *g* factor, the FSIQ is a good estimator of this general factor. These findings are consistent with Spearman (1927) and Carroll's (1993) conceptualizations of *g*. Given the overwhelming dominance of the general factor, the present results indicated that interpretation of first-order factors is quite limited and problematic given the conflation of general and group factor variance in index scores. Further, some studies have shown that the temporal stability of index scores was not sufficiently high (Kieng, Rossier, Favez, & Lecerf, 2017; Watkins & Smith, 2013).

Limitations

While critically important, EFA and CFA cannot by themselves fully determine construct validity of the French WISC–V so studies of relations with external criteria are needed. Methods such as incremental predictive validity (Canivez, 2013a; Canivez, Watkins, James, Good, & James, 2014; Glutting, Watkins, Konold, & McDermott, 2006) could help determine if reliable achievement variance is incrementally accounted for by the French WISC–V factor index scores beyond that accounted for by the FSIQ (or through latent factor scores [see also Kranzler, Benson, & Floyd, 2015]). Diagnostic utility (Canivez, 2013b) studies should also be examined to determine if differential patterns of French WISC–V factor index scores correctly identify individuals of differing clinical disorders. However, given the small portions of true score variance uniquely contributed by the group factors of the French WISC–V it is inconceivable that they would provide substantial value. Another limitation is that the present study examined EFA and CFA for the full French WISC–V standardization sample. It is possible that different age groups within the French WISC–V standardization sample might produce different results so examination of structural invariance across age (and other variables) would be useful. Further, these results also pertain to the standardization normative sample and may not generalize to clinical populations or independent samples of nonclinical groups.

Conclusion

From a practical point of view, the present results have several important implications for the interpretation of French WISC–V subtests and the factor index scores. The higher-order model proposed by the test publisher is not adequate, which could be quite problematic from a clinical point of view and may lead to errors in interpreting the scores (Silverstein, 1993). Practitioners must be aware that they are taking some risks when interpreting factor index scores, particularly VS and FR. Thus, it is recommended that practitioners use several reading grids of interpretation and not rely only on the interpretation promoted by the publisher. The present results suggested that primary interpretation of the French WISC–V should focus on the global FSIQ rather than on the first-order group factors, because the *g* factor accounts for the largest part of the common variance. It is encouraging that the publisher suggested interpretation the FSIQ at the first step, challenging the steps of interpretation recommended with the WISC–IV (Kaufman, Raiford, & Coalson, 2016). Factor index scores, however, conflate *g* variance and unique group factor variance, which cannot be disentangled for individuals. The factor index scores cannot be considered to reflect only broad ability measurement; indeed, they include a strong contribution of the general intelligence factor.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16,* 397–438. http://dx.doi.org/10.1080/10705510903008204

Beaujean, A. A. (2016). Reproducing the Wechsler Intelligence Scale for Children–Fifth ed.: Factor model results. *Journal of Psychoeducational Assessment, 34,* 404–408. http://dx.doi.org/10.1177/0734282916642679

Beaujean, A. A., Parkin, J., & Parker, S. (2014). Comparing Cattell-Horn-Carroll factor models: Differences between bifactor and higher order factor models in predicting language achievement. *Psychological Assessment, 26,* 789–805. http://dx.doi.org/10.1037/a0036745

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 80,* 796–846. http://dx.doi.org/10.1111/j.1467-6494.2011.00749.x

Byrne, B. M. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Canivez, G. L. (2008). Orthogonal higher order factor structure of the Stanford-Binet Intelligence Scales—Fifth ed. for children and adolescents. *School Psychology Quarterly, 23,* 533–541. http://dx.doi.org/10.1037/a0012884

Canivez, G. L. (2013a). Incremental criterion validity of WAIS–IV factor index scores: Relationships with WIAT–II and WIAT–III subtest and composite scores. *Psychological Assessment, 25,* 484–495. http://dx.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). New York, NY: Oxford University Press.

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

Canivez, G. L. (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). Gottingen, Germany: 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://dx.doi.org/10.1037/a0018030

Canivez, G. L., & Kush, J. C. (2013). WAIS–IV and WISC–IV structural validity. *Journal of Psychoeducational Assessment, 31,* 157–169. http://dx.doi.org/10.1177/0734282913478036

Canivez, G. L., & Watkins, M. W. (2010). 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. http://dx.doi.org/10.1037/a0020429

Canivez, G. L., & Watkins, M. J. (2016). Review of the Wechsler Intelligence Scale for Children—Fifth Edition: Critique, commentary, and independent analyses. In A. S. Kaufman, S. E. Raiford, & D. Coalson (Eds.), *Intelligent testing with the WISC–V* (pp. 683–702). Hoboken, NJ: Wiley.

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. http://dx.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. http://dx.doi.org/10.1037/pas0000358

Canivez, G. L., Watkins, M. W., Good, R., James, K., & James, T. (2017). Construct validity of the Wechsler Intelligence Scale for Children—Fourth UK edition with a referred Irish sample: Wechsler and Cattell-Horn-Carroll model comparisons with 15 subtests. *British Journal of Educational Psychology, 87,* 383–407. http://dx.doi.org/10.1111/bjep.12155

Canivez, G. L., Watkins, M. W., James, T., Good, R., & James, K. (2014). Incremental validity of WISC–IV^{UK} factor index scores with a referred Irish sample: Predicting performance on the WIAT–II^{UK}. *British Journal of Educational Psychology, 84,* 667–684. http://dx.doi.org/10.1111/bjep.12056

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571312

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429–452. http://dx.doi.org/10.1207/s15327906mbr3003_6

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276. http://dx.doi.org/10.1207/s15327906mbr0102_10

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural equation modeling: A *Multidisciplinarity Journal, 14,* 464–504. http://dx.doi.org/10.1080/10705510701301834

Chen, H., Zhang, O., Raiford, S. E., Zhu, J., & Weiss, L. G. (2015). Factor invariance between genders on the Wechsler Intelligence Scale for Children–Fifth Edition. *Personality and Individual Differences, 86,* 1–5. http://dx.doi.org/10.1016/j.paid.2015.05.020

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Child, D. (2006). *The essentials of factor analysis* (3rd ed.). New York, NY: Continuum.

Chuderski, A., & Necka, E. (2012). The contribution of working memory to fluid reasoning: Capacity, control, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1689–1710. http://dx.doi.org/10.1037/a0028465

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. http://dx.doi.org/10.1177/0013164410379332

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*. Psychological Assessment, *29,* 1001–1015.

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence, 53,* 194–201. http://dx.doi.org/10.1016/j.intell.2015.10.009

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. http://dx.doi.org/10.1177/0734282909333179

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169–182. http://dx.doi.org/10.1016/j.intell.2006.07.002

Gignac, G. (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. http://dx.doi.org/10.1027/1614-0001.27.2.73

Gignac, G. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences, 42,* 37–48. http://dx.doi.org/10.1016/j.paid.2006.06.019

Gignac, G. (2008). Higher-order models versus direct hierarchical models: G as superordinate or breadth factor? *Psychology Science, 50,* 21–43.

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55,* 57–68. http://dx.doi.org/10.1016/j.intell.2016.01.006

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. http://dx.doi.org/10.1080/00273171.2013.804398

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55,* 377–393. http://dx.doi.org/10.1177/0013164495055003002

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. *The Journal of Special Education, 40,* 103–114. http://dx.doi.org/10.1177/00224669060400020101

Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS–III). *Psychological Assessment, 23,* 143–152. http://dx.doi.org/10.1037/a0021230

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC–IV factor structure through Bayesian structural equation modeling. *Psychological Assessment, 25,* 496–508. http://dx.doi.org/10.1037/a0030676

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2,* 41–54. http://dx.doi.org/10.1007/BF02287965

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185. http://dx.doi.org/10.1007/BF02289447

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariancestructure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151. http://dx.doi.org/10.1177/001316446002000116

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York, NY: John Wiley & Sons, Inc.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC–V*. Hoboken, NJ: Wiley.

Kieng, S., Rossier, J., Favez, N., & Lecerf, T. (2017). Long-term stability of the French WISC–IV: Standard and CHC index scores. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée, 67,* 51–60. http://dx.doi.org/10.1016/j.erap.2016.10.001

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

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. http://dx.doi.org/10.1037/pas0000119

Lecerf, T., Golay, P., Reverte, I., Senn, D., Favez, N., & Rossier, J. (2011). *Orthogonal higher-order structure and confirmatory factor analysis of* *the French Wechsler Children Intelligence Scale—Fourth Edition (WISC–IV)*. Paper presented at the 12th European Congress of Psychology, Istanbul, Turkey.

Lecerf, T., Rossier, J., Favez, N., Reverte, I., & Coleaux, L. (2010). The four- vs. alternative six-factor structure of the French WISC–IV: Comparison using confirmatory factor analyses. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie, 69,* 221–232. http://dx.doi.org/10.1024/1421-0185/a000026

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence, 61,* 120–129. http://dx.doi.org/10.1016/j.intell.2017.01.012

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471–491. http://dx.doi.org/10.1037/a0019227

Matzke, D., Dolan, C. V., & Molenaar, D. (2010). The issue of power in the identification of "g" with lower-order factors. *Intelligence, 38,* 336–344. http://dx.doi.org/10.1016/j.intell.2010.02.001

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1–10. http://dx.doi.org/10.1016/j.intell.2008.08.004

Molenaar, D. (2016). On the distortion of model fit in comparing the bifactor model and the higher-order factor model. *Intelligence, 57,* 60–63. http://dx.doi.org/10.1016/j.intell.2016.03.007

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. J. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3,* 2–20. http://dx.doi.org/10.3390/jintelligence3010002

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling, 23,* 116–139. http://dx.doi.org/10.1080/10705511.2014.961800

Morin, A. J. S., Arens, A. K., Tran, A., & Caci, H. (2016). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the Composite Scale of Morningness. *International Journal of Methods in Psychiatric Research, 25,* 277–288. http://dx.doi.org/10.1002/mpr.1485

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41,* 407–422. http://dx.doi.org/10.1016/j.intell.2013.06.004

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Naglieri, J. A. (2016). Independent WISC–V test review: Theoretical and practical considerations. In A. S. Kaufman, S. E. Raiford, & D. Coalson (Eds.), *Intelligent testing with the WISC–V* (pp. 663–668). Hoboken, NJ: Wiley.

R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47,* 667–696. http://dx.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. http://dx.doi.org/10.1080/00223891.2012.725437

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure? *Intelligence, 62,* 31–47. http://dx.doi.org/10.1016/j.intell.2017.02.005

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24,* 282–292. http://dx.doi.org/10.1037/a0025697

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22,* 53–61. http://dx.doi.org/10.1007/BF02289209

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 99–144). New York, NY: Guilford Press.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6,* 461–464. http://dx.doi.org/10.1214/aos/1176344136

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343. http://dx.doi.org/10.1007/BF02294360

Silverstein, A. B. (1993). Type I, Type II, and other types of errors in pattern analysis. *Psychological Assessment, 5,* 72–74. http://dx.doi.org/10.1037/1040-3590.5.1.72

Spearman, C. E. (1927). *The abilities of man: Their nature and measurement*. New York, NY: MacMillan.

Ullman, J. B., & Bentler, P. M. (2013). Structural equation modeling. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology* (pp. 661–690). Hoboken, NJ: Wiley.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321–327. http://dx.doi.org/10.1007/BF02293557

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). Boston, MA: Kluwer. http://dx.doi.org/10.1007/978-1-4615-4397-8_3

Watkins, M. W. (2000). Monte Carlo PCA for parallel analysis [Computer software]. State College, PA: Ed. & Psych Associates.

Watkins, M. W. (2004). MacOrtho [Computer software]. State College, PA: Ed. & Psych Associates.

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

Watkins, M. W. (2007). SE_{scree} [Computer software]. State College, PA: Ed. & Psych Associates.

Watkins, M. W. (2011). CIeigenvalue [Computer software]. Phoenix, AZ: Ed. & Psych Associates.

Watkins, M. W. (2013). Omega [Computer software]. Phoenix: AZ: Ed. & Psych Associates.

Watkins, M. W., & Ravert, C. M. (2013). Subtests, factors, and constructs: What is being measured by tests of intelligence? In J. C. Kush (Ed.), *Intelligence quotient: Testing, role of genetics and the environment and social outcomes* (pp. 55–68). Hauppauge, NY: Nova Science.

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

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, 6,* 975–983, http://dx.doi.org/10.1177/0013164406288168

Wechsler, D. (2005). *WISC–IV. Echelle d'intelligence de Wechsler pour enfants-4e édition*. Paris, France: Pearson France-ECPA.

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

Wechsler, D. (2016a). *WISC–V. Echelle d'intelligence de Wechsler pour enfants-5e édition*. Paris, France: Pearson France-ECPA.

Wechsler, D. (2016b). *WISC–V. Echelle d'intelligence de Wechsler pour enfants-5e édition, manual d'interprétation*. Paris, France: Pearson France-ECPA.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013). WISC–IV and clinical validation of the four- and five-factor interpretative approaches. *Journal of Psychoeducational Assessment, 31,* 114–131. http://dx.doi.org/10.1177/0734282913478032

Wood, J. M., Tataryn, J. M., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 354–365. http://dx.doi.org/10.1037/1082-989X.1.4.354

Yang, R., Spirtes, P., Scheines, R., Reise, S. P., & Mansoff, M. (2017). Finding pure submodels for improved differentiation of bifactor and second-order models. *Structural Equation Modeling, 24,* 402–413. http://dx.doi.org/10.1080/10705511.2016.1261351

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. http://dx.doi.org/10.1007/s11336-003-0974-7

Zoski, K. W., & Jurs, S. (1996). An objective counterpart to the visual scree test for factor analysis: The standard error scree. *Educational and Psychological Measurement, 56,* 443–451. http://dx.doi.org/10.1177/0013164496056003006

Received January 26, 2017
Revision received July 17, 2017
Accepted July 20, 2017 ■

Correction to Lecerf and Canivez (2018)

In the article "Complementary Exploratory and Confirmatory Factor Analyses of the French WISC–V: Analyses Based on the Standardization Sample" by Thierry Lecerf and Gary L. Canivez (*Psychological Assessment*, 2018, Vol. 30, No. 6, pp. 793–808. http://dx.doi.org/10.1037/pas0000526), a production error resulted in the deletion of subtests in the "French WISC–V subtest" column and the misalignment of factor names in the "Eigenvalue" column of Table 1. The table should read as follows:

Table 1
French Wechsler Intelligence Scale for Children—Fifth Edition (French WISC–V) Four Oblique Factor Solution for the Total Standardization Sample

French WISC–V subtest	General	F1: Verbal Comprehension	F2: Perceptual Reasoning	F3: Processing Speed	F4: Working Memory	h^2
SI	.724	**.672** (.481)	.135 (.609)	−.065 (.304)	.058 (.616)	.623
VC	.666	**.918** (.809)	−.054 (.503)	−.034 (.271)	−.080 (.531)	.664
IN	.716	**.611** (.749)	.149 (.607)	.005 (.350)	.045 (.604)	.577
CO	.621	**.707** (.698)	−.091 (.463)	.120 (.364)	.008 (.513)	.499
BD	.652	−.013 (.478)	**.733** (.719)	.121 (.444)	−.088 (.514)	.530
VP	.719	−.058 (.521)	**.922** (.825)	.015 (.416)	−.086 (.573)	.687
MR	.683	.087 (.557)	**.589** (.710)	−.057 (.337)	.120 (.598)	.519
FW	.628	.130 (.534)	**.462** (.632)	−.090 (.281)	.167 (.565)	.431
AR	.708	.089 (.583)	*.302* (.661)	.031 (.407)	**.374** (.682)	.520
DS	.686	−.036 (.553)	.035 (.588)	−.074 (.336)	**.818** (.781)	.615
PS	.592	.021 (.458)	.225 (.546)	.173 (.440)	.291 (.561)	.372
LNS	.716	.038 (.591)	−.084 (.579)	.055 (.430)	**.809** (.801)	.646
CD	.446	.055 (.307)	−.057 (.344)	**.698** (.697)	.012 (.356)	.488
SS	.491	−.014 (.314)	.015 (.401)	**.758** (.768)	.015 (.393)	.589
CA	.327	−.024 (.203)	.106 (.291)	**.467** (.488)	−.047 (.247)	.242
Eigenvalue		6.49	1.46	1.02	.80	
Factor correlations		F1: VC	F2: PR	F3: PS	F4: WM	
Verbal Comprehension (VC)						
Perceptual Reasoning (PR)		.690				
Processing Speed (PS)		.404	.507			
Working Memory (WM)		.727	.752	.497		

Note. N = 1,049. French WISC–V subtests: SI = Similarities; VC = Vocabulary; IN = Information; CO = Comprehension; BD = Block Design; VP = Visual Puzzles; MR = Matrix Reasoning; FW = Figure Weights; AR = Arithmetic; DS = Digit Span; PS = Picture Span; LNS = Letter-Number Sequencing; CD = Coding; SS = Symbol Search; CA = Cancellation. Salient pattern coefficients (≥.30) presented in bold (structure coefficient). h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (g loadings).

The online version of this article has been corrected.

http://dx.doi.org/10.1037/pas0000638

Table S1

French Wechsler Intelligence Scale for Children–Fifth Edition (French WISC–V) Five Oblique Factor Solution for the Total Standardization Sample (N = 1,049)

French WISC–V Subtest	General	F1: Verbal Comprehension	F2: Perceptual Reasoning	F3: Processing Speed	F4: Working Memory	F5: Inadequate	h^2
SI	.723	**.676** (.782)	.142 (.604)	-.065 (.298)	.078 (.611)	-.034 (.487)	.626
VC	.664	**.919** (.810)	-.048 (.496)	-.032 (.266)	-.067 (.522)	-.025 (.434)	.666
IN	.715	**.610** (.748)	.120 (.596)	.008 (.344)	.002 (.588)	.087 (.533)	.577
CO	.620	**.706** (.698)	-.084 (.457)	.120 (.359)	.017 (.507)	-.016 (.408)	.499
BD	.650	-.007 (.478)	**.718** (.720)	.120 (.439)	-.073 (.508)	.000 (.460)	.530
VP	.717	-.051 (.521)	**.889** (.824)	.016 (.410)	-.083 (.563)	.035 (.531)	.684
MR	.682	.094 (.558)	**.594** (.713)	-.059 (.330)	.149 (.597)	-.038 (.479)	.527
FW	.626	.134 (.533)	**.416** (.623)	-.086 (.274)	.107 (.549)	.124 (.511)	.429
AR	.741	.003 (.582)	.079 (.650)	.019 (.401)	.073 (.665)	**.780** (.890)	.803
DS	.685	-.021 (.553)	.033 (.579)	-.074 (.329)	**.783** (.780)	.036 (.541)	.613
PS	.593	.023 (.459)	.267 (.553)	.169 (.440)	**.387** (.581)	-.149 (.363)	.403
LNS	.714	.054 (.590)	-.082 (.568)	.055 (.423)	**.752** (.794)	.054 (.559)	.636
CD	.445	.054 (.305)	-.061 (.339)	**.691** (.694)	.002 (.352)	.030 (.294)	.485
SS	.491	-.013 (.313)	.004 (.396)	**.758** (.770)	-.008 (.386)	.046 (.333)	.594
CA	.326	-.024 (.203)	.121 (.293)	**.463** (.488)	-.011 (.251)	-.052 (.186)	.244
Eigenvalue		6.49	1.46	1.02	0.80	0.72	
Factor correlations		F1: VC	F2: PR	F3: PS	F4: WM	F5	
Verbal Comprehension (VC)		—					
Perceptual Reasoning (PR)		.680	—				
Processing Speed (PS)		.396	.496	—			
Working Memory (WM)		.715	.733	.486	—		
F5		.596	.648	.391	.670	—	

Note. French WISC–V Subtests: SI = Similarities, VC = Vocabulary, IN = Information, CO = Comprehension, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, FW = Figure Weights, AR = Arithmetic, DS = Digit Span, PS = Picture Span, LNS = Letter–Number Sequencing, CD = Coding, SS = Symbol Search, CA = Cancellation. Salient pattern coefficients (≥ .30) presented in bold (Structure Coefficient); h^2 = Communality. General structure coefficients are based on the first unrotated factor coefficients (*g* loadings).

Table S2

French Wechsler Intelligence Scale for Children–Fifth Edition (French WISC–V) Two and Three Oblique Factor Solutions for the Total Standardization Sample (N = 1,049)

French WISC–V	Two Oblique Factors				Three Oblique Factors				
Subtest	g^1	F1: *g*	F2: PS	h^2	g^1	F1: PR/WM	F2: VC	F3: PS	h^2
SI	.725	**.849** (.758)	-.159 (.331)	.591	.726	.200 (.644)	**.662** (.777)	-.075 (.311)	.620
VC	.651	**.791** (.686)	-.182 (.274)	.493	.665	-.078 (.536)	**.872** (.796)	-.047 (.276)	.639
IN	.718	**.782** (.739)	-.076 (.375)	.549	.718	.203 (.639)	**.598** (.745)	-.004 (.356)	.574
CO	.614	**.635** (.625)	-.016 (.350)	.391	.624	-.087 (.502)	**.716** (.701)	.116 (.368)	.502
BD	.645	**.478** (.620)	.247 (.522)	.425	.649	**.703** (.687)	-.099 (.459)	.103 (.445)	.483
VP	.701	**.580** (.685)	.183 (.517)	.492	.711	**.871** (.772)	-.139 (.502)	.004 (.421)	.605
MR	.682	**.653** (.684)	.055 (.431)	.470	.685	**.723** (.719)	.046 (.547)	-.069 (.344)	.521
FW	.629	**.644** (.640)	-.008 (.363)	.409	.630	**.624** (.652)	.113 (.528)	-.100 (.287)	.438
AR	.711	**.639** (.705)	.115 (.483)	.506	.709	**.577** (.710)	.152 (.591)	.040 (.417)	.515
DS	.673	**.652** (.677)	.042 (.418)	.459	.671	**.536** (.669)	.192 (.579)	-.016 (.357)	.464
PS	.595	**.422** (.569)	.254 (.497)	.367	.593	**.434** (.585)	.071 (.466)	.182 (.448)	.369
LNS	.703	**.623** (.696)	.126 (.485)	.495	.700	**.430** (.672)	.259 (.615)	.095 (.438)	.490
CD	.446	-.031 (.358)	**.675** (.657)	.433	.448	-.063 (.367)	.063 (.312)	**.705** (.697)	.488
SS	.493	-.058 (.392)	**.780** (.747)	.560	.493	.013 (.422)	-.013 (.318)	**.767** (.769)	.592
CA	.329	-.031 (.263)	**.510** (.492)	.243	.328	.079 (.294)	-.051 (.200)	**.464** (.485)	.238
Eigenvalue		6.49	1.46			6.49	1.46	1.02	
Factor Correlations		F1	F2			F1	F2	F3	
	F1	–			F1	–			
	F2	.576	–		F2	.733	–		
					F3	.545	.418	–	

Note. French WISC–V Subtests: SI = Similarities, VC = Vocabulary, IN = Information, CO = Comprehension, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, FW = Figure Weights, AR = Arithmetic, DS = Digit Span, PS = Picture Span, LNS = Letter–Number Sequencing, CD = Coding, SS = Symbol Search, CA = Cancellation, *g* = general intelligence, PS = Processing Speed, PR = Perceptual Reasoning, WM = Working Memory, VC = Verbal Comprehension, h^2 = Communality. [1]General structure coefficients based on first unrotated factor coefficients (*g* loadings). Salient pattern coefficients (≥ .30) presented in bold (Structure Coefficient) based on principal factors extraction with promax rotation ($k = 4$).

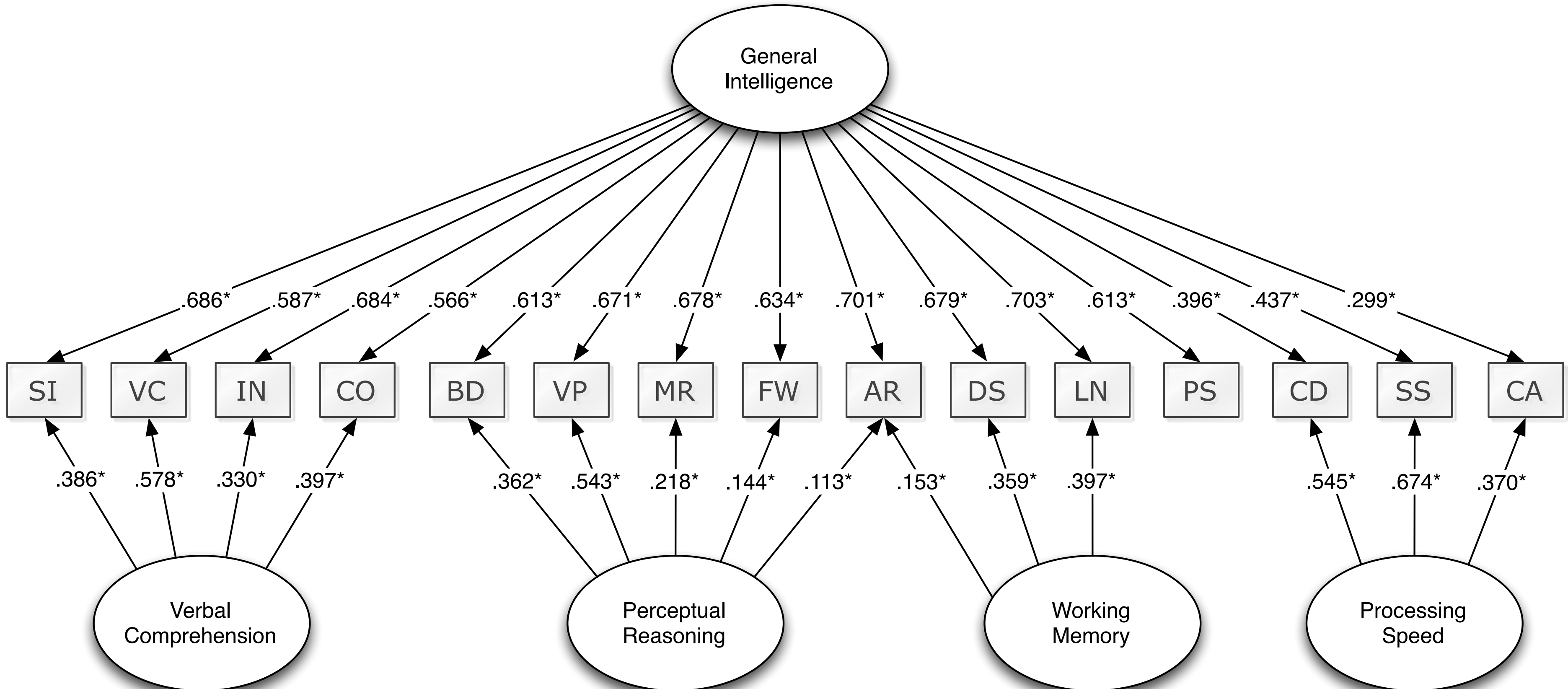

Figure S3. Bifactor measurement model (Model 4c Bifactor without PS––WM path), with standardized coefficients, for the French WISC–V standardization sample (N = 1,049) 15 Subtests. SI = Similarities, VC = Vocabulary, IN = Information, CO = Comprehension, BD = Block Design, VP = Visual Puzzles, MR = Matrix Reasoning, FW = Figure Weights, AR = Arithmetic, DS = Digit Span, LN = Letter-Number Sequencing, PS = Picture Span, CD = Coding, SS = Symbol Search, CA = Cancellation. *p < .05.