The Four- vs. Alternative Six-Factor Structure of the French WISC-IV. Comparison Using Confirmatory Factor Analyses

LECERF, Thierry, *et al*.

Reference

LECERF, Thierry, *et al*. The Four- vs. Alternative Six-Factor Structure of the French WISC-IV. Comparison Using Confirmatory Factor Analyses. *Swiss journal of psychology*, 2010, vol. 69, no. 4, p. 221-232

DOI : 10.1024/1421-0185/a000026

Comparing the Usual Four- and an Alternative Six-Factor Structure of the French WISC-IV Using Confirmatory Factor Analyses

Thierry Lecerf

Faculty of Psychology and Educational Sciences, University of Geneva, and Institute of Psychology, University of Lausanne

Jérôme Rossier

Institute of Psychology, University of Lausanne

Nicolas Favez, Isabelle Reverte, & Laurence Coleaux

Faculty of Psychology and Educational Sciences, University of Geneva,

Authors' Note

This research was funded through a grant from the Swiss National Science Foundation to T. Lecerf, N. Favez, and J. Rossier (grant no. 100014-118248). Correspondence should be addressed to Thierry Lecerf, FPSE, University of Geneva, 40, Bd du Pont d'Arve, 1205, Genève, Switzerland, E-mail : Thierry.Lecerf@unige.ch

Abstract

Exploratory and confirmatory factor analyses reported in the French technical manual of the WISC-IV provided evidence supporting a structure with four indices: Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Working Memory (WMI), and Processing Speed (PSI). Although the WISC-IV is more attuned to contemporary theory, it is still not in total accordance with the dominant theory: the Cattell-Horn-Carroll (CHC) theory of cognitive ability. This study was designed to determine whether the French WISC-IV is better described with the four factors solution or whether an alternative model based on the CHC theory is more appropriate. The intercorrelations matrix reported in the French technical manual was submitted to confirmatory factor analysis. Comparison of competing models suggests that a model based on the CHC theory fit the data better than the current WISC-IV structure. It appears that the French WISC-IV measures six factors: crystallized intelligence (Gc), fluid intelligence (Gf), short-term memory (Gsm), processing speed (Gs), quantitative knowledge (Gq), and visual processing (Gv). We recommend that clinicians interpret the subtests of the French WISC-IV in relation to this CHC model in addition to the four indices.

Keywords: WISC-IV, CHC theory, confirmatory factor analysis

Comparing the Usual Four- and an Alternative Six-Factor Structure of the French WISC-IV Using Confirmatory Factor Analyses

There are several tests of intelligence and IQ for children available at the moment (KABC-II, Raven, etc.), but the Wechsler batteries have dominated the field of individual testing since 1949 and are the most widely used over the world for the cognitive assessment of children (Flanagan & McGrew, 1998; Grégoire, 2006; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Zhu & Weiss, 2005). The latest version of the WISC – the WISC-IV– represents a considerable revision and is more psychometrically and theoretically grounded (Grégoire, 2006). However, this most recent version still has some limitations and above all with respect to relationships with theory. Although the new factorial structure of the WISC-IV is more theoretically grounded, it is not completely in the line with modern theories of cognitive abilities, or with the *Cattell-Horn-Carroll* (CHC) theory in particular, which stems mainly from analyses of adult samples (Alfonso, Flanagan, & Radwan, 2005). Nevertheless, some authors have suggested that models based on the CHC theory provide a better fit to the standardization data than the four-factor structure for children's data (Flanagan & Kaufman, 2004; Keith et al., 2006). Following these studies, the purpose of this investigation was to answer two questions raised regarding the factorial structure of the French WISC-IV and regarding the constructs measured by each subtest. As most studies concerning this aspect were conducted on the North American or on the Canadian version of the WISC-IV (Prifitera, Saklofske, & Weiss, 2005; Prifitera, Weiss, Saklofske, & Rolfhus, 2005), the main purpose of this study was to analyze the structure underlying the French version of the WISC-IV in order to assess whether the results obtained in North America can be generalized to other cultural settings.

The "Standard" Factorial Structure of the WISC-IV and the CHC Model

The interpretation of the WISC has moved from a two-factor (WISC; Wechsler, 1949) to a four-factor solution (WISC-III; Wechsler, 1991, for the North American version, and WISC-IV; Wechsler, 2005, for the French version), but has nevertheless provided some continuity to its users (i.e., the calculation of the Full Scale IQ). According to the technical manual, the current interpretation of the WISC-IV is based on four indices: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Processing Speed Index (PSI), and Working Memory Index (WMI). While the VCI and the PSI factors have retained their names from WISC-III, the name of the third factor – "Perceptual Organization" – has been changed to Perceptual Reasoning factor. Furthermore, the "Freedom from Distractibility" index of the WISC-III was renamed Working Memory index (note that this fourth factor was not present in the French WISC-III). The introduction of this four-factor solution and the changes in the names of the factors represent an attempt to align the factorial structure of the WISC-IV with contemporary theory. However, it is not explicitly and completely related to the dominant modern theory, the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. For instance, it has been suggested that the Perceptual Reasoning Index (PRI) confound fluid intelligence (Gf) and visual processing (Gv), two broad abilities included in the CHC model (Grégoire, 2006). Therefore, the interpretation of the PRI is a mixture of Gf and Gv. Remember that while the WISC-IV has incorporated subtests to measure Gf (Matrix Reasoning, Picture Concepts), no Gf index is calculated.

The influence of the CHC theory on test development and test interpretation has mainly increased during the past 10 years (McGrew, 2009). The CHC theory is a hierarchical model and based on the Cattell-Horn *Gf-Gc* model (Horn & Noll, 1997) and on the Carroll *three-stratum level theory* (Carroll, 1993). The CHC theory is the best empirically supported psychometric model of intelligence (i.e., individual differences in factor analytic studies, changes in abilities across the lifespan; Taub, Floyd, Keith, & McGrew, 2008; Taub, McGrew, & Witta, 2004), in which cognitive functioning is subdivided into one general factor (i.e., Spearman's *g* factor), 10 broad abilities, and about 70 narrow abilities (Floyd, Bergeron, McCormack, Anderson, & Hargrove-Owens, 2005; Schrank, 2005). The ten broad abilities are: fluid intelligence (Gf), crystallized intelligence (Gc), visual processing (Gv), short-term memory (Gsm), processing speed (Gs), long-term retrieval (Glr), auditory processing (Ga), quantitative knowledge (Gq), decision / reaction time (Gt), and reading and writing ability (Grw-R). Recently, McGrew (2005) proposed adding four broad factors (tactile abilities, kinesthetic abilities, olfactory abilities, and psychomotor abilities) and separating speed factors into three speed abilities.

The significance of the development of the CHC theory is not only theoretical but has also two main applied consequences. First, more recent batteries, like the KABC-II or the Woodcock-Johnson III (not available for French-speaking sample), are based on the CHC theory (Kaufman, Kaufman, Kaufman-Singer, & Kaufman, 2005; Schrank, 2005). Second, the CHC model is also used as the foundation for developing, administering, and interpreting tests of intelligence and cognitive abilities, which is called the "*cross-battery approach*" (Flanagan & Ortiz, 2001; Flanagan, Ortiz, & Alfonso, 2007; Floyd et al., 2005). This method suggests that clinicians should cross batteries to measure a greater breadth of Gf-Gc factors. The main goal of this approach is to narrow the gap between theory and practice by mapping each subtest from each intelligence battery onto the CHC-Narrow Ability Classification (CHC-NAC; Alfonso et al., 2005; Flanagan & McGrew, 1997; McGrew, 1997). In other words, at the first step, subtests can be interpreted at the narrow abilities level, which represent the lowest level on the hierarchy of abstraction. At the second step, different narrow-abilities tests can be combined to form clusters of higher level construct; this is the broad cognitive abilities level. This work of classification has been done with the North American or with the Canadian Wechsler Scales, but also with all other main batteries, like the CAS, the DAS, the KABC, etc. (McGrew, 1997; Phelps, McGrew, Knopik, & Ford, 2005) and has applied implications at an individual level. To our knowledge, no CHC classification has been done on the French Wechsler scales.

The analyses of the strengths and weaknesses among these broad or narrow abilities is crucial for understanding the nature of a child's difficulties and are related to academic domains. It has been shown that CHC abilities are related to math, writing, and more generally to specific learning disabilities (Floyd, Evans, & McGrew, 2003; Floyd, Keith, Taub, & McGrew, 2007). More specifically, it has been demonstrated that lexical knowledge (VL), language development (LD), working memory (MW), and perceptual speed (P) are related to reading achievement and to writing achievement (Flanagan, 2000; Flanagan & Mascolo, 2005). Gf seems to be related to math achievement. It is also important to mention that Flanagan (2000) showed that the CHC model is better than the standard four-factor solution because it explained 25% more variance in reading achievement than the Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), and freedom from distractibility interpretation system. It means that CHC-NAC would allow clinicians to make more adequate predictions than the standard four-factor solution proposed in the French technical manual.

In sum, studies have validated the CHC model as a framework for making differential diagnoses and for guiding test selection. In addition, the narrow and the broad abilities levels are clinically useful because they provide information for between-individual and within-individual variability. Regarding the Wechsler Scales, and particularly the subtests of the WISC-IV, hypotheses have been proposed concerning the broad and narrow abilities measured by each subtest (Flanagan & Kaufman, 2004; Flanagan & Mascolo, 2005; Flanagan & McGrew, 1998; McGrew, 1997; Phelps et al., 2005). However, the French technical manual does not provide data for a model, which contains the Gf, Gc, Gv, Gsm, and Gs factors; this is the main purpose of this study. In line with Keith et al. (2006) and to Flanagan and Kaufman (2004), we assumed that a CHC-based model would provide better fit than the four-factor solution reported in the technical manual.

The Present Investigation

The main purpose of this study is to answer two questions raised regarding the structure of the French WISC-IV and the constructs measured by each subtest and therefore to improve the validity of the interpretation of the French WISC-IV by applying the structure of CHC theory. In order to determine the factorial structure and the construct-relevant variance, confirmatory factor analysis (CFA) will be conducted on the subtests of the French WISC-IV. More precisely, several CHC models will be compared with an oblique four-factor solution, as reported in the French technical manual, and mainly with a higher-order model with four first-order factors and one second-order factor.

The first issue concerns the debate about the structure of the WISC-IV. Flanagan and Kaufman (2004) suggested that the WISC-IV measures six broad abilities, while Keith and colleagues (2006) reported evidence that it measures 5 broad abilities; both solutions differ from the "standard" interpretation of the current version (with 4 factors). These authors suggested that the structure of the WISC-IV is not a good explanation of the constructs measured by the test. Concerning the French version, Grégoire (2006) assumed that the WISC-IV measures 5 broad abilities. Following these authors, one of the most important interests in this investigation is the comparison of competing theoretical models. We will use CFA techniques to analyze the structure of the French intercorrelation matrix of the WISC-IV and to compare the current structure of the WISC-IV to models derived from the CHC theory. In other words, we will test whether the data support the four-factor solution of the French WISC-IV or whether they support a CHC model (Keith, 1997, 2005; Keith et al., 2006).

Following this first debate on the factorial structure, a second controversy concerns the constructs measured by each subtest. Although several studies were conducted on the Wechsler Scales and particularly on the WISC, there are still some questions about the constructs underlying the different subtests proposed by the WISC-IV. The second objective of this study was to determine what construct(s) is (are) measured by each subtest and to test several hypotheses. For instance, Arithmetic was described as a fluid intelligence (Gf) and a short-term memory (Gsm) measure by Keith and colleagues (2006), as a fluid intelligence (Gf) and a quantitative knowledge (Gq) measure by Flanagan and Kaufman (2004); as a quantitative knowledge and a short-term memory measure by the Psychological Corporation; and finally as a Gsm and a Gc measure by Grégoire (2006). Similarly, expert consensus classified Similarities and Word Reasoning as measures of fluid intelligence (Gf; see also Flanagan & Kaufman, 2004), while Keith and colleagues and Grégoire classified them as measures of crystallized intelligence (Gc). Furthermore, Matrix Reasoning is classically defined as a fluid intelligence measure (Grégoire, 2006), but it has been suggested that this task requires visual processing (Carroll, 1993). Moreover, some authors have suggested that this type of task also requires working memory (Salthouse, 1993). Some debates concern the subtest Symbol Search. Keith and colleagues showed that this subtest measures Gs and Gv, while most authors classify it only as a Gs measure.

Data, Analyses Reported in the WISC-IV Technical Manual

The French WISC-IV technical manual reported the intercorrelations between the subtests (Wechsler, 2005, p. 45) as well as results of a series of exploratory and confirmatory factor analyses. These findings were used to support the current factorial structure of the WISC-IV with four factors: VCI, PRI, WMI, and PSI. Three others competing models were tested and rejected: a model with one general factor, a model with two factors (1: VCI+WMI; 2: PRI+PSI), and a model with three factors (1: VCI+ Arithmetic; 2: PRI; 3: WMI+PSI).

Moreover, the French WISC-IV manual reported only little information about the *goodness-of-fit indexes*, and factor loadings were not presented. Most importantly, no model based on the CHC theory was tested. For instance, no CFA was conducted with Arithmetic on a single factor (i.e., Gq) or on both Gsm and Gc as suggested by Grégoire (2006). It should be mentioned that although the publishers seemed to accept the existence of a general factor of intelligence, no models included this *g* factor.

Following CFA conducted on the WISC-IV, the main purpose of this study was to determine whether the current factorial structure of the WISC-IV is really the most adequate or whether models based on the CHC theory would more adequately describe the factorial structure of the French WISC-IV. This study also allowed us to determine whether the WISC-IV subtests measure the same constructs in American and French children (Keith et al., 2006). In sum, we addressed two questions, one concerning the structure underlying the WISC-IV, comparing the usual four factors and a structure based on the CHC theory, and a second question concerning the constructs measured by each subtest of the French WISC-IV.

Method

Participants

The sample used for this study was the French WISC-IV standardization sample described in the French WISC-IV technical manual (ages 6:0 through 16:11 years) and included 1103 participants. This entire sample included 11 separate age levels. As described in the French technical manual, the sample was stratified according to socioeconomic status and demographic region; age and sex were controlled. The CFAs were conducted on the basis of the subtest intercorrelation matrix reported in the WISC-IV manual, which is based on standardized scores (Table 5.1, p. 45). Indeed, standardized scores were computed for each age group separately ($M = 10$; $SD = 3$).

Instrument

The WISC-IV is an individually administered intelligence test for children (aged 6 to 16:11 years). The WISC-IV has 10 core subtests (*M* = 10; *SD* = 3) that allow the calculation of the four indices (VCI, PRI, PSI, and WMI). The Full Scale IQ is based on the sum of these 10 core subtests (for a brief description of the subtests, see Appendix Table A). The FSIQ and the four indices are all based on a mean of 100 and standard deviation of 15. The WISC-IV also has five supplemental subtests (*M* = 10; *SD* = 3).

Analyses

As mentioned above, the main purpose of the present investigation was to test the factorial structure of the French WISC-IV. CFAs were conducted on the intercorrelation matrix to compare the current factorial structure of the WISC-IV to models based on the CHC theory. The CFA analyses are presented in the results section. This question was also used to determine the constructs measured by each subtest. Amos 17 (with SPSS) was used to run CFA models. The intercorrelation matrix and the standard deviations reported in the WISC-IV manual were used as input; remember that correlations were computed on the basis of standardized scores (*M* = 10; *SD* = 3). All WISC-IV subtests were used in the present investigation (core and supplemental).

Analyses will provide fit statistics, which indicate the adequacy of the “measurement model” and the “structural model.” We considered multiple indicators of fit to assess the alternative models (e.g., Hu & Bentler, 1999). We used the chi-square (χ^2) statistic (which will be used with the degrees of freedom to assess the probability that the model is accurate). However, χ^2 is related to sample size (with a large sample, all χ^2 could be significant), so we used the root mean square error of approximation (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) as primary fit indices (Byrne, 2001). These scores were supplemented by the Tucker-Lewis fit index (TLI; also called non-normed fit index, or NNFI), which is relatively unrelated to sample size, and which was used to evaluate model fit. In addition, the comparison fit index (CFI) was used. Values for these two scores larger than .95 indicated excellent fit. According to current rules, a low χ^2 suggests that the model fits the data relatively adequately; SRMRs below 0.08 and RMSEAs below 0.06 indicate a good fit of the model to the data (a value of .08 suggests an adequate fit), particularly if TLI values are greater than 0.95 (Hu & Bentler, 1999). We also used the Akaike information criterion (AIC) to compare models that are not nested: The smaller AIC suggests the better model.

Results

Does the WISC-IV Measure VCI, PRI, WMI, and PSI?

In the first step, the current structure with four correlated factors (VCI, PRI, WMI, and PSI) was tested. As shown in Table 1 (Model 1), the model provided a good fit to the data. Although CFAs were conducted on the intercorrelation matrix instead of the empirical data, results were relatively similar to those reported in the French WISC-IV manual (pp. 52-53; RMSEA = .04; TLI = .957; AGFI = .954). This first step provided evidence that our results replicated those presented in the French WISC-IV technical manual.

INSERT TABLE 1 ABOUT HERE

Because we assumed that all subtests measure a general intelligence factor (Watkins, 2006; Watkins, Wilson, Kotz, Carbone, & Babula, 2006) and because our goal was to compare our findings with those reported by Keith and colleagues (2006), a hierarchical model was tested in the second step, with four factors and one general factor (Table 1, Model 2). The model provided a good fit to the intercorrelation matrix. We observed that WMI had a loading of .88 on the *g* factor, while PRI had a loading of .87, and VCI had a loading of .84. The PSI factor had a lower loading (.52). The difference in respective AIC values suggests that the addition of the *g* factor does not improve the model fit. Nevertheless, this second model was used as a "WISC-IV" reference model and compared with the subsequent CHC models that were tested.

Does the WISC-IV Measure CHC Abilities?

As mentioned, some authors have tested the CHC structure in the WISC-IV, and some controversy has remained on the constructs measured by each subtest. The first "basic" CHC model tested was exactly the same model used by Keith and colleagues (2006). It should be noted that this basic CHC model is relatively consistent with the CHC classification made by Flanagan and Kaufman (2004). In this model, Similarities, Vocabulary, Comprehension, Information, and Word Reasoning were placed on the Gc factor. Block Design and Picture Completion were placed on the Gv factor, while Picture Concepts, Matrix Reasoning, and Arithmetic were placed on the Gf factor. Digit Span and Letter-Number were placed on the Gsm factor; and finally Coding, Symbol Search, and Cancellation were placed on the Gs factor. As shown in Table 1 (Model 3), this basic CHC model fit the data well, AIC was 365.89. This basic CHC model was used as a reference model and compared with the following CHC models. Note that, unlike Keith and colleagues, the difference in respective AIC values suggests that the basic CHC model does not improve the model fit in comparison with the WISC-IV model. Indeed, AIC was 365.89 for the basic CHC model and 355.18 for the WISC-IV higher-order model (Remember that the smaller AIC suggests the better model). This is a main discrepancy with Keith's findings, who observed a superiority of their initial CHC model over the WISC-IV model. With the French version of the WISC-IV, the basic CHC model does not provide a better explanation of the constructs measured by the subtests than the current factorial structure. Nevertheless, we observed, in line with Keith and colleagues, that Gf had a loading of 1.00 on the *g* factor. The Gs factor had a lower loading (.52). Although this basic CHC model does not support the hypothesis that the CHC based-model provides a better description of the WISC-IV subtests, several alternative CHC models were tested to improve the understanding of the constructs measured by the French WISC-IV subtests. The results of these CHC models (from Model 4 to Model 19) will be compared to those reported by Keith and colleagues (Model 3 in our study). Thus, the main purpose was to find a model that is both meaningful and statistically well-fitting for the French WISC-IV.

Arithmetic Measures Gf and/or Gsm and/or Gc and/or Gs?

In the basic CHC model and according to Keith and colleagues (2006), Arithmetic loaded on the Gf factor (because it is assumed that Arithmetic measures the narrow ability Quantitative Reasoning). However, it has been suggested that Arithmetic should require Gf and Gq for older children or should require Gsm and Gf for younger children (Flanagan & Kaufman, 2004) or Gq and Gsm (Psychological Corporation) or Gsm and Gc (Grégoire, 2006), or finally Gq and Gs (Phelps et al., 2005).

First, we tested the possibility that Arithmetic loads on both Gf and Gsm (Model 4) as assumed by Flanagan and Kaufman (2004). As shown in Table 1, this model provided better fitting indices than the WISC-IV higher-order model or the basic CHC higher-order model, with an AIC of 337.56. The difference in respective AIC values suggests that the addition of a loading from Arithmetic on Gsm does improve the model fit in comparison to the WISC-IV model. Second, we tested the possibility that Arithmetic loads on both Gsm and Gc (Model 5) as assumed by Grégoire (2006). As shown in Table 1, this CHC model fits the data well. This model in which Arithmetic loaded on both Gsm and Gc resulted in a better fitting model than the model in which Arithmetic loaded only on Gf (i.e., the basic CHC model). Although not reported here, models in which Arithmetic loaded only on Gsm or only on Gc were not better than the model in which Arithmetic loaded only on Gf (AIC was 372 when Arithmetic loaded only on Gsm; AIC was 439 when Arithmetic loaded only on Gc). Third, we tested the possibility that Arithmetic should measure Gf, Gsm, and Gc (Model 6). As shown in Table 1, this CHC model fit the data very well. Arithmetic had a modest loading on Gsm (.365) and on Gc (.284) and a lower loading on Gf (.165). This result was consistent with Grégoire's (2006) hypothesis, but does not support Keith and colleagues' (2006) results. Fourth, given the low loading on Gf, we tested the possibility that Arithmetic loads on both Gf and on a separate factor Gq as assumed by Flanagan and Kaufman. Although not reported here, the loading for Arithmetic on Gq exceeded 1.0, while the loading on Gf was statistically nonsignificant. This suggests that Arithmetic does not measure a mix of Gf (RQ) and Gq (A3) and was a measure of Gq rather than a measure of Gf. We then tested the possibility that Arithmetic loads on both Gq and Gsm (Model 7) as assumed by the *Psychological Corporation*. This model fit the data well. Subsequently, we tested the possibility that Arithmetic loads on both Gq and Gs as assumed by Phelps and colleagues (2005). Although not reported here, the loading for Arithmetic on Gs was nonsignificant (-.01). Arithmetic does not measure processing speed. Finally, we also tested the possibility that Arithmetic loaded only on quantitative knowledge ability (Model 8). As shown in Table 1, this CHC model fits the data only slightly better than the basic CHC model.

Taken together, these results suggest that Arithmetic should measure Gq, Gsm, and Gc. Thus, we tested this possibility (Model 9). As shown in Table 1, this CHC model fits the data well. The difference in respective AIC values suggests that this model is better than the WISC-IV higher-order model. This last model makes sense because Arithmetic has a high loading on Gq (.727) and modest loadings on Gsm (.355) and Gc (.264). Thus, in the final CHC model, Arithmetic loaded on Gq, Gsm, and Gc.

Word Reasoning and Similarities Measure Gc and/or Gf and/or Verbal Gf?

Expert consensus classified the Similarities and Word Reasoning subtests as measures of crystallized intelligence (Gc) and fluid intelligence (Gf), while Keith and colleagues (2006) and Grégoire (2006) classified them as measures of Gc. The results showed that Word Reasoning (Model 10) and Similarities (Model 11) have very low and not statistically significant loadings on Gf (.001 and -.001, respectively) and that fit indices were not better than the basic CHC higher-order model. This result was consistent with Grégoire and with Phelps and colleagues (2005). Neither Word Reasoning nor Similarities measure fluid intelligence. Furthermore, we tested the possibility that these two tasks measure a separate verbal fluid intelligence factor (Model 12). As shown in Table 1, goodness-of-fit statistics suggested that the hypothesized model did not represent an adequate fit to the data. Therefore, in the final CHC model, we concluded that Word Reasoning and Similarities measure only Gc ability as assumed by Grégoire and Keith and colleagues.

Block Design Measures Gv, and/or Gf, and/or Gs, and or Gsm?

Frequently, Block Design is considered to measure fluid intelligence (Kaufman, 1994), while some authors have considered it a unique measure of visualization (Burton et al., 2001; Carroll, 1993; Grégoire, 2006; Keith, 2005; McArdle, Hamagami, Meredith, & Bradway, 2000). Thus, we tested the possibility that Block Design loads on Gv and Gf (Model 13). As shown in Table 1, goodness-of-fit statistics were similar to those obtained with the basic CHC model although Block Design had a nonsignificant loading on Gf (.02). In accordance with modification indices proposed in previous models and exploratory factor analysis, we tested the possibility that Block Design measures both Gv and Gs (Model 14). As shown in Table 1, allowing Block Design to load on Gv and Gs resulted in a significant improvement in fit in comparison to the basic CHC model. In the final CHC model, Block Design loaded on Gv and Gs.

Picture Completion Measures Gv and/or Gc?

Picture Completion primarily measures Gv and particularly flexibility of closure (CF), but some data and some experts have suggested that it also measures crystallized intelligence (Gc; Grégoire, 2006; Phelps et al., 2005). As shown in Table 1 (Model 15), a model in which Picture Completion loads on both Gv and Gc provided an improvement in fit compared to the basic CHC model (AIC = 335.96). Note that the model in which Picture Completion was placed only on Gc did not fit better than the model in which Picture Completion loaded only on Gv (AIC = 385.20). Thus, Picture Completion appears to measure Gc (.35) and Gv (.27). In the final CHC model, Picture Completion loaded on Gc and Gv.

Coding Measures Gs and/or Gsm?

It has been suggested that children who are able to learn and memorize the stimuli show better performance in the Coding subtest. Thus, we tested the possibility that Coding loads both on Gs and on Gsm. Although not reported here, the loading for Coding on Gsm was statistically nonsignificant (-.055). Thus, Coding measures only Gs.

Picture Concepts Measures Gf, and/or Gc, and/or Gv?

It has been suggested that Picture Concepts, a new subtest, which was created to measure mainly Gf, also requires Gc ability and/or Gv. Thus, we tested the possibility that Picture Concepts loads on both Gf and Gc (Model 16). As shown in Table 1, this model did not provide an improvement in fit compared to the basic model. Moreover, the loading for Picture Concepts on Gc was statistically nonsignificant (-.03). A second model in which Picture Concepts loaded on Gv and Gf resulted in an improvement in fit (Model 17; AIC = 358.90). In the final CHC model, Picture Concepts loaded on Gv and Gf.

Matrix Reasoning Measures Gf and/or Gv?

Matrix Reasoning is classically defined as a fluid intelligence measure (Grégoire, 2006), but it has been suggested that this type of task requires visual processing (Carroll, 1993) and/or working memory (Salthouse, 1993). Therefore, we tested the possibility that Matrix Reasoning loads on Gf and/or on Gsm and/or on Gv. As shown in Table 1, allowing Matrix Reasoning to load on Gf and Gv (Model 18) improves model fit (AIC = 353.60). Second, we tested the possibility that Matrix Reasoning measures Gf and Gsm. Although not reported here, the loading for Matrix Reasoning on Gsm was statistically nonsignificant (-.04). Thus, in the final CHC model, Matrix Reasoning loaded on Gv and Gf.

Symbol Search Measures Gs and/or Gv?

Last, we tested the possibility that Symbol Search loads both on Gs and Gv (Model 19). As reported in Table 1, allowing Symbol Search to load on both Gs and Gv improved the model fit in comparison to the basic CHC model (AIC = 363.83). In the final CHC model, Symbol Search loaded on Gs and Gv.

INSERT FIGURE 1 ABOUT HERE

The Final CHC Model

The "individual" models tested before were combined to identify and validate the final CHC model. In this model, Similarities, Vocabulary, Comprehension, Information, Word Reasoning, Arithmetic and Picture Completion were placed on Gc ability. Picture Concepts and Matrix Reasoning were placed on Gf ability. Block Design, Picture Completion, Pictures Concepts, Matrix Reasoning, and Symbol Search were placed on Gv ability. Digit Span, Letter-Number, and Arithmetic were placed on Gsm ability. Coding, Symbol Search, Cancellation, and Block Design were placed on Gs ability. Finally, Arithmetic was placed on Gq. Although not reported here, three loadings of this model were statistically nonsignificant (or marginally significant) and thus removed (Gv – Symbol Search, $p = .257$; Gc – Arithmetic, $p = .096$; Gv – Picture Concepts, $p = .064$). The model was then reestimated and showed that all loadings were statistically significant. However, because the factor loading of Gs on Block Design was lower than .20, it was also removed. After deleting the loadings that were statistically nonsignificant and the loading that was very low, all remaining loadings were statistically significant and higher than .20. For this final CHC model (Model 20), the goodness-of-fit indicated a very good fit of the model to the data. The difference in respective AIC values suggests that this CHC model is better than the

WISC-IV model (Model 2). The results are presented in Figure 1. As concerns the second-order loadings, results showed that Gq, Gf, and Gc have the highest loading on *g*, and that Gv and Gs have lower loadings.

Discussion

The exploratory and confirmatory factor analyses presented in the French technical manual of the WISC-IV support a four-factor solution with Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index, and Processing Speed Index. Although this new factorial structure is more attuned to theory, it is not really based on the dominant model: the *Cattell-Horn-Carroll* (CHC) theory of cognitive ability, which suggests that the WISC-IV has an alternative factorial structure. Thus, some questions remain about the structure of the WISC-IV and about the constructs measured by each subtest.

Concerning the debate about the structure of the WISC-IV, the results of the present investigation show that the French WISC-IV could be described with 6 factors: fluid intelligence (Gf), crystallized intelligence (Gc), visual processing (Gv), quantitative knowledge (Gq), processing speed (Gs), and short-term memory (Gsm). Most importantly, the difference in the respective AIC values suggests that a model based on the CHC theory describes the underlying abilities of the subtests of the French WISC-IV better than the current structure (i.e., comparison between Model 2 and Model 20). This finding is consistent with the data presented by Keith and colleagues (2006) and by Flanagan and Kaufman (2004) on a North-American sample. Thus, the "standard" structure of the WISC-IV is not a good explanation of the constructs measured by the test. More specifically, our results suggest that the Perceptual Reasoning index is an indicator of both visual processing (Gv) and fluid intelligence (Gf).

Following this first debate on the factorial structure, a second controversy concerns the constructs measured by each subtest. As mentioned above, there are still some questions about the constructs underlying the different subtests proposed by the WISC-IV. The second objective of this study was to determine what construct(s) is (are) measured by each subtest and to test several hypotheses. Results from this investigation provide some information about the constructs measured by each subtest and about the interpretation of the subtests of the French WISC-IV (Table 2). Most importantly, the results suggest some discrepancy between the North American and the French data.

As concerns Similarities and Word Reasoning, the results indicate that both subtests measure crystallized intelligence, but not fluid intelligence. This result is consistent with Keith and colleagues (2006) and Grégoire (2006), who classified both subtests as measures of crystallized intelligence (Gc). This finding is clinically relevant for the interpretation of these two subtests and showed that they do not provide verbal fluid measures. We recommend that clinicians group Similarities and Word Reasoning with Information, Comprehension, and Vocabulary, and interpret them as a measure of crystallized intelligence.

The results show that Block Design measures visual processing (Burton et al., 2001; Carroll, 1993; Grégoire, 2006; Keith, 2005; McArdle et al., 2000) and is not a measure of Gf. Thus, when the three Perceptual Reasoning subtests are inconsistent, we recommend that clinicians, in a first step, interpret Block Design as a measure of Gv and Matrix Reasoning and Pictures Concepts as measures of fluid intelligence. In a second step, we recommend grouping the Block Design subtest with the three processing speed subtests: Coding, Symbol Search, and Cancellation. Indeed, our results showed that Block Design had a statistically significant loading on Gs, although it was removed because the loading was lower than .20. Thus, this result suggests that Block Design may be more consistent with the processing speed subtests for some examinees.

As concerns Picture Completion, although designed to measure Gv, it primarily measures Gc (Flanagan McGrew, & Ortiz, 2000; Grégoire, 2006; Kaufman, 1994) and to a lesser degree Gv in the French WISC-IV. It should be noted that our data suggested, in contrast to Keith and colleagues (2006), that Picture Completion primarily measures Gc but not Gv. Thus, we recommend, in a first step, grouping Picture Completion with the Gc subtests (Similarities, Vocabulary, etc.). In a second step, if Picture Completion is inconsistent with the Gc subtests, we recommend grouping Picture Completion with Block Design and interpreting them as measures of Gv. Consequently, visual processing is underestimated in the WISC-IV because only one subtest, Block Design, is really an adequate measure of this broad ability.

The Matrix Reasoning subtest mainly measures fluid intelligence (Grégoire, 2006), but also had a statistically significant loading on Gv. This result is consistent with Carroll (1993), who suggested that some fluid intelligence tests require visual processing abilities. Similarly, it has been suggested that Picture Concepts, designed to measure Gf, also requires Gv and/or Gc abilities (Grégoire, 2006). Contrary to the Matrix Reasoning subtest, the results indicate that Picture Concepts only measures fluid intelligence. The loading of Picture Concepts on Gv was marginally significant and therefore removed. Thus, we recommend, in a first step, grouping Matrix Reasoning and Picture Concepts and interpreting them as measures of fluid intelligence. If Picture Concepts and Matrix Reasoning are inconsistent, we recommend grouping Matrix Reasoning with Block Design and interpreting them as measures of Gv ability. Nevertheless, given the loadings, Picture Concepts could be more consistent with Block Design than with Matrix Reasoning. Thus, for some examinees Picture Concepts could be interpreted as a measure of Gv.

Symbol Search appears to measure only Processing Speed (Gs). In contrast to Keith and colleagues (2006), the loading of Symbol Search on Gv was statistically nonsignificant. Similarly, our results indicate that Coding only measures Gs and is not an indicator of Gsm. Thus, we recommend grouping Symbol Search, Coding, and Cancellation and interpreting them as measures of Gs. It means that the processing speed index of the current interpretation is an adequate measure of this ability.

According to the *Psychological Corporation*, the Arithmetic subtest measures quantitative knowledge (Gq) and short-term memory (Gsm). In contrast, Keith and colleagues (2006) suggested that Arithmetic loaded on the Gf factor (and also on Gsm and on Gc), while Flanagan and Kaufman (2004) assumed that Arithmetic should require Gf and Gsm for younger children. Grégoire (2006) suggested that it measures Gsm and Gc and finally, Phelps and colleagues (2005) proposed that it requires Gq and Gs. The results from this investigation were consistent with the classification proposed by the *Psychological Corporation* and indicate that Arithmetic appears to measure quantitative knowledge (Gq) and short-term memory (Gsm); the loading of Arithmetic on Gc was statistically nonsignificant. Overall, the data suggest that Arithmetic in the French WISC-IV does not measure fluid intelligence (Gf). Indeed, in the models in which Arithmetic simultaneously loaded on both Gq and Gf (model Gf + Gq; or model Gf + Gq + Gc; or model Gf + Gq + Gc), the loading of Arithmetic on Gf was statistically nonsignificant and the loading of Arithmetic on Gq was higher. Furthermore, in the model in which Arithmetic loaded on Gf, Gsm, and Gc, results indicate that Arithmetic appears to measure Gsm and then Gc rather than Gf. The loading on Gf was lower than .20. Contrary to Keith and colleagues, Arithmetic should not be considered with Matrix Reasoning and Pictures Concepts because Arithmetic does not measure Gf. Thus, Arithmetic may assess Gq and Gsm. These findings have repercussions on the interpretation of the French WISC-IV subtest of Arithmetic. In the first step, we recommend interpreting this subtest as a measure of quantitative knowledge (Gq). However, the assessment of this ability is underestimated in the WISC-IV because only Arithmetic appears to measure it. Thus, in the second step and in a systematic way, Arithmetic should be considered a measure of Gsm, along with Digit Span and Letter-Number Sequencing. Finally, if the short-term memory subtests are inconsistent, Arithmetic should be compared with Gc subtests because some examinees may rely on verbal ability to solve Arithmetic problems (remember that the loading was marginally significant, $p = .096$). It should be noted that in contrast to Keith and colleagues, the French subtest Arithmetic is not the strongest measure of *g* (.67, Table 2). Taken together, this result appears to indicate that interpretation of Arithmetic is very complicated and that this subtest might never be interpreted alone. Nevertheless, we recommend administering Arithmetic systematically.

INSERT TABLE 2 ABOUT HERE

The comparison of our data with those reported by Keith and colleagues (2006) supports the existence of cultural differences between North American and French children. Therefore, the CHC narrow ability classification of the WISC-IV subtests based on North-American studies is not completely adequate for the French WISC-IV subtests. Consequently, more data are needed to determine the underlying processes of the French subtests.

Conclusion

Although more related to the contemporary model of cognitive abilities, the WISC-IV is not explicitly based on the CHC theory. To gain clinical validity from the WISC-IV, we suggest that practitioners interpret the results of the subtests according to the CHC narrow ability classification (CHC-NAC). To our knowledge, the present investigation is the first one conducted on the French data. Thus, we recommend interpreting the results of the French WISC-IV according to the CHC classification, in addition to the standard indices, because some subtests assess two or more abilities. Indeed, it is important to remember that CFAs and goodness-of-fit indices suggest whether the model is plausible, but they do not determine which model is the best one. Following this recommendation, the main priority of interpretation is the question about the “homogeneity” of each construct, and only homogenous broad or narrow abilities might be interpreted. As mentioned above, Picture Completion should first be considered a measure of Gc, along with the other verbal subtests. If these subtests are inconsistent, Picture Completion should be compared next with Block Design and considered to be a measure of Gv if these two subtests are consistent. As concerns Arithmetic, this subtest should be considered first as a measure of quantitative knowledge. However, this ability is underestimated in the WISC-IV because only Arithmetic is an appropriate measure of it. Arithmetic should be considered secondly as a measure of Gsm, along with Digit Span and Letter-Number Sequence. Finally, if the tests are inconsistent, Arithmetic should be compared with Gc subtests. Contrary to Keith and colleagues (2006), Arithmetic should not be considered with Matrix Reasoning and Pictures Concepts, because Arithmetic does not measure Gf in the French WISC-IV. Processing speed (Gs) corresponds to the standard processing speed index, with Coding, Symbol Search, and Cancellation. We also recommend that Block Design be considered as a measure of Gs along with the other processing speed subtests because our data show that it loads on Gs for some examinees (the loading was statistically significant but lower than .20). The main discrepancy between the standard interpretation and the CHC interpretation concerns the Perceptual Reasoning index. Indeed, this index is a mixture of visual processing (Gv) and fluid intelligence (Gf). Therefore, some inconsistencies could be observed between Matrix Reasoning and Block Design, for instance, for some examinees. Matrix Reasoning should first be considered a measure of Gf, along with Pictures Concepts. If the tests are inconsistent, Matrix Reasoning should be compared with Block Design and then Picture Completion and considered a measure of Gv. Finally, in contrast to Keith and Colleagues, our data show that the Working Memory index is not a mixture of Gsm and Gf, but a mixture of Gq and Gsm. To test the different hypotheses for one child, we recommend using the "*Cross-battery*" procedure (Flanagan, 2000; Flanagan et al., 2007) in which standard scores (M =

100, *SD* = 15) are computed. Because some abilities are underestimated (Gq, Gv), subtests from other batteries might be administered. For instance, the subtest Triangles from the KABC-II could be administered to assess Gv adequately. Finally, because several subtests assess two or more abilities, the clinician may consider several hypotheses to understand the cognitive functioning of a particular child. The clinician has to work as Sherlock Holmes to find, for a particular child, the consistent and homogenous “subtests grouping”. Indeed, the more we focus on the configuration of a particular child (*person-oriented* approach), the more powerful our hypotheses will be about his/her functioning.

References

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (2nd ed., pp. 185-202). New York: The Guilford Press.

Burton, D. B., Sepehri, A., Hecht, F., Vandenbrock, A., Ryan, J. J., & Drabman, R. (2001). A confirmatory factor analysis of the WISC-III in a clinical sample with cross-validation in the standardization sample. *Child Neuropsychology, 7*, 104-116.

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.

Flanagan, D. P. (2000). Wechsler-based CHC cross-battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler Test Scores. *School Psychology Quarterly, 15*, 295-329.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, New Jersey: John Wiley, & Sons, Inc.

Flanagan, D. P., & Mascolo, J. T. (2005). Psychoeducational assessment and learning disability diagnosis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (2nd ed., pp. 521-544). New York: The Guilford Press.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 314-325). New York: The Guilford Press.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf-Gc theory: Joint confirmatory factor analysis of the WJ-R and KAIT in a non-white sample. *Journal of School Psychology, 36*, 151-182.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf - Gc theory. A contemporary approach to interpretation*. Needham Heigths, MA: Allyn and Bacon.

Flanagan, D. P., & Ortiz, S. (2001). *Essentials of cross-battery assessment*. New York: Wiley & Sons, Inc.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment. Second edition*. Hoboken, New Jersey: Wiley & Sons, Inc.

Floyd, R. G., Bergeron, R., McCormack, A. C., Anderson, J. L., & Hargrove-Owens, G. L. (2005). Are Cattell-Horn-Carroll broad ability composite scores exchangeable across batteries? *School Psychology Review, 34*, 329-357.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*, 155-171.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll Cognitive Abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22*, 200-233.

Grégoire, J. (2006). *L'examen clinique de l'intelligence de l'enfant. Fondements et pratiques du WISC-IV (The clinical examination of the intelligence of the child. Foundations and practice of the WISC-IV)*. Sprimont: Mardaga.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf - Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 53-91). New York: The Guilford Press.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, *Structural Equation Modeling, 6*(1), 1-55.

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

Kaufman, J. C., Kaufman, A. S., Kaufman-Singer, J., & Kaufman, N. L. (2005). The Kaufman Assessment Battery for children - Second Edition and the Kaufman Adolescent and Adult test. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (2nd ed., pp. 344-370). New York: The Guilford Press.

Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373-402). New York: The Guilford Press.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 581-614). New York: The Guilford Press.

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

McArdle, J. J., Hamagami, F., Meredith, W., & Bradway, K. P. (2000). Modeling the dynamics hypotheses of Gf-Gc theory using longitudinal life-span data. *Learning and Individual Differences, 12*, 53-79.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 151-179). New York: The Guilford Press.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (2nd ed., pp. 136-181). New York: The Guilford Press.

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.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The general (g), broad, and narrow CHC stratum characteristics of the WJ III and WISC-III tests: A confirmatory cross-battery investigation. *School Psychology Quarterly, 20*, 66-88.

Prifitera, A., Saklofske, D. H., & Weiss, L. G. (2005). *WISC-IV. Clinical use and interpretation*. San Diego, USE: Elsevier, Inc.

Prifitera, A., Weiss, L. G., Saklofske, D. H., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV. Clinical use and interpretation* (pp. 3-32). San Diego, USA: Elsevier, Inc.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology, 84*, 171-199.

Schrank, F. A. (2005). Woodcock-Johnson III tests of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (2nd ed., pp. 371-401). New York: The Guilford Press.

Taub, G. E., Floyd, R. G., Keith, T. Z., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*, 187-198.

Taub, G. E., McGrew, K. S., & Witta, E. L. (2004). A confirmatory analysis of the factor structure and cross-age invariance of the Wechsler Adult Intelligence Scale—Third edition. *Psychological Assessment, 16*, 85-89.

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

Watkins, M. W., 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.

Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children - Third Edition (WISC-III)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2005). *Manuel de l'Echelle d'Intelligence de Wechsler pour Enfants - 4e édition (Manual for the Wechsler Intelligence Scale for Children - Fourth Edition)*. Paris : Editions du Centre de Psychologie Appliquée.

Zhu, J., & Weiss, L. (2005). The Wechsler scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (pp. 297-324). New York: The Guilford Press.

Table 1

Comparison of Fit of Models About the WISC-IV Structure

Models	χ^2	df	AIC	Dev. WISC[1]	Dev CHC[2]	RMSEA	SRMR	TLI	CFI	AGFI
1. WISC-IV (4 factors)	276.86	84	348.87			.046	.0316	.957	.965	.954
2. WISC-IV model (4 factors + *g*)	287.18	86	355.18		B	.046	.0331	.956	.964	.954
3. Basic CHC model	297.89	86	365.89	W		.047	.0338	.954	.962	.949
4. Arithmetic on Gf + Gsm	265.55	84	337.56	B	B	.044	.0321	.959	.967	.955
5. Arithmetic on Gsm + Gc	244.43	84	316.43	B	B	.042	.0303	.964	.971	.958
6. Arithmetic on Gf + Gsm + Gc	237.75	83	311.75	B	B	.041	.0301	.965	.972	.959
7. Arithmetic on Gq + Gsm	256.82	84	328.82	B	B	.043	.0315	.961	.969	.956
8. Arithmetic on Gq	285.72	85	355.72	W	B	.046	.0334	.956	.964	.951
9. Arithmetic on Gq + Gsm + Gc	239.32	83	313.32	B	B	.041	.0302	.965	.972	.958
10. Word Reasoning on Gf + Gc	297.89	85	367.89	W	W	.048	.0338	.953	.962	.949
11. Similarities on Gf + Gc	297.89	85	367.89	W	W	.048	.0338	.953	.962	.949
12. Verbal Gf (SI + WR)	462.94	84	534.94	W	W	.064	.0468	.915	.932	.917
13. Block Design on Gv + Gf	297.89	86	365.89	W	=	.047	.0338	.954	.962	.949
14. Block Design on Gv + Gs	281.70	85	351.70	B	B	.046	.0321	.956	.965	.951
15. Picture Completion on Gv + Gc	265.96	85	335.96	B	B	.044	.0310	.960	.968	.954
16. Picture Concepts on Gf + Gc	297.67	85	367.66	W	W	.048	.0337	.953	.962	.949
17. Picture Concepts on Gf + Gv	288.90	85	358.90	W	B	.047	.0333	.955	.963	.951
18. Matrix Reasoning on Gf + Gv	283.60	85	353.60	B	B	.046	.0337	.956	.964	.952
19. Symbol Search on Gs + Gv	293.38	85	363.38	W	B	.047	.0336	.954	.963	.949

20. Final CHC model	215.85	82	291.85	B	B	.038	.0281	.969	.976	.962

[1] AIC value compared to Model 2; [2] AIC value compared to Model 3; B means better than Models 2 or 3; W means Worse than Models 2 or 3

Table 2

Loading on the g Factor, and Primary and Secondary Ability Measures by Each Subtest of the French WISC-IV

	Loading on the g factor			
	Keith et al.	Lecerf et al.	Primary	Secondary
Arithmetic	.79	.67	Gq	Gsm (Gc ?)
Matrix Reasoning	.69	.53	Gf	Gv
Picture Concepts	.59	.46	Gf	(Gv ?)
Coding	.45	.31	Gs	
Symbol Search	.53	.39	Gs	
Digit Span	.55	.45	Gsm	
Number Letter	.63	.53	Gsm	
Block Design	.66	.49	Gv	(Gs ?)
Picture Completion	.59	.51	Gc	Gv
Similarities	.71	.78	Gc	
Vocabulary	.75	.80	Gc	
Information	.71	.76	Gc	
Word Reasoning	.63	.67	Gc	
Comprehension	.64	.63	Gc	
Cancellation	.27	.25	Gs	

Note. Gc = crystallized intelligence, Gf = fluid intelligence; Gv = visual processing; Gs = processing speed; Gq = quantitative knowledge; Gsm = short-term memory

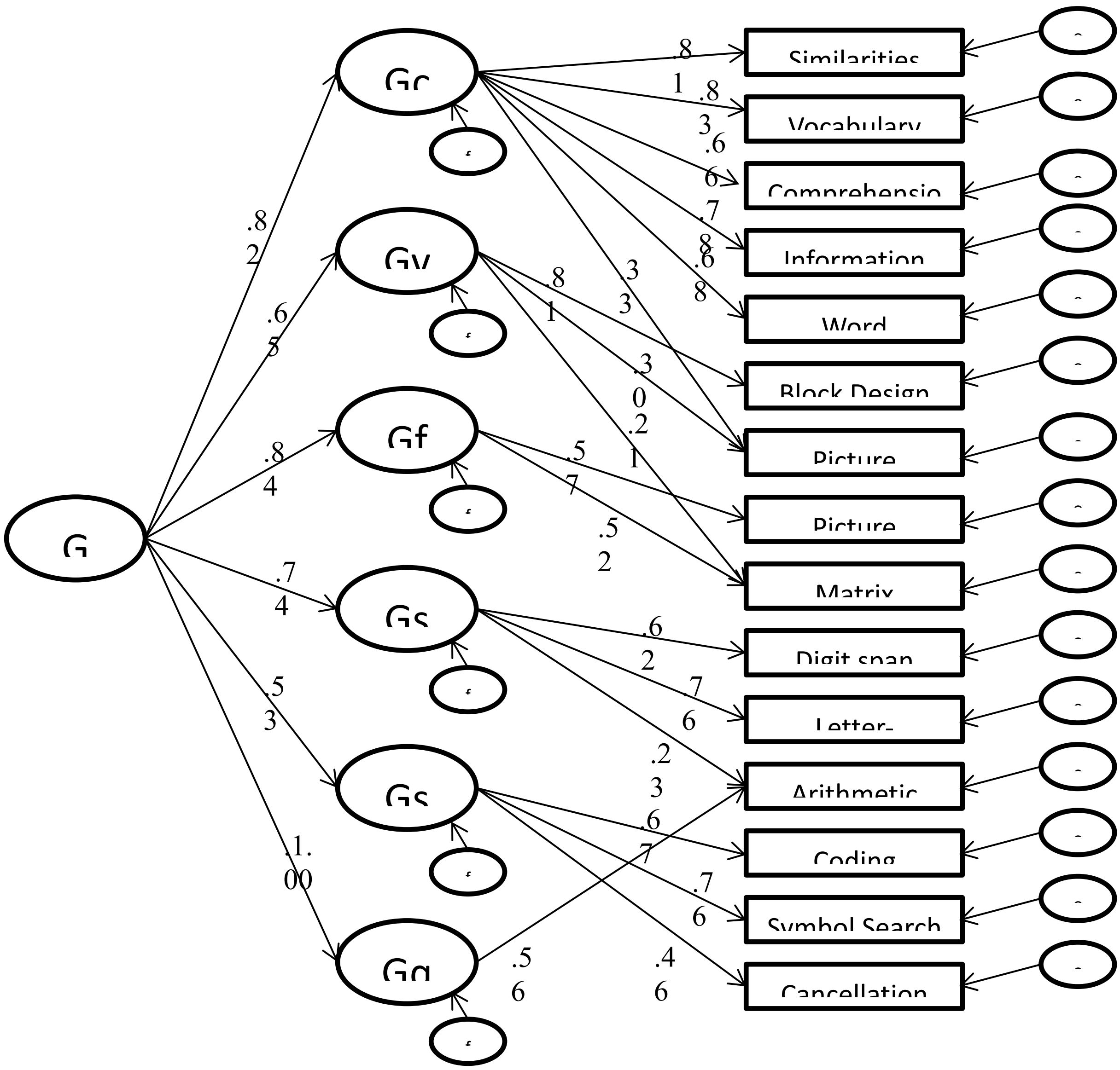

Figure 1. The final CHC model of the French WISC-IV

Table A: *WISC-IV subtest definitions*

Subtest	Description
Block Design	Children are required to replicate printed geometric patterns with red-and-white blocks within a time limit. Block Design has 14 items.
Similarities	This subtest requires stating why two objects or two concepts are alike. Similarities has 23 items.
Digit Span	Children are required to repeat digits verbatim in the first part (Digit-Span forward) and to repeat digits in reverse order in a second part (Digit-Span backward). Series range in length from 2 to 9. For each sequence length, there are 2 trials.
Picture Concepts	This subtest requires selecting one picture from among two or three rows of pictures to form a concept (i.e., a group with a common feature). Picture Concepts has 28 items.
Coding	In the Coding subtest, children are required to copy symbols that are matched with geometric patterns or numbers within 120-s time limit.
Vocabulary	Children are required to provide definitions for words or are required to give the names of the objects. Vocabulary has 36 items.
Letter-Number Sequencing	Children are required to recall numbers in an ascending order and then letters in an alphabetical order; letters and numbers are presented in random order. Series range in length from 2 to 8. For each sequence length, there are 3 trials.
Matrix Reasoning	Children are required to choose one of five responses to complete the missing part of a colored matrix or visual patterns. Matrix Reasoning has 35 items.
Comprehension	The Comprehension subtest requires explaining a series of situations, activities, etc. (general principles, social situations). Comprehension has 21 items.
Symbol Search	In the Symbol Search subtest, children are required to scan and to determine whether a symbol was present or absent in the array within a 120-s time limit.
Picture Completion	The Picture Completion subtest requires identification of the most important missing part of a picture within a 20-s time limit. Picture Completion has 38 items.
Cancellation	In the Cancellation subtest, children are required to scan a random and a non-random arrangement of pictures and to mark pictures of animals within a 45-s time limit.
Information	In the Information subtest, children are required to answer questions concerning general knowledge (geographical facts, calendar information, historical figures, etc.). Information has 33 items.
Arithmetic	Arithmetic subtest requires solving simple and complex arithmetic problems within a time limit. Arithmetic has 34 items.
Word Reasoning	In the Word Reasoning subtest, children are required to identify a concept that was described by a series of clues (1 to 3 clues). Word Reasoning has 24 items.