Structural validity of the Wechsler Intelligence Scale for Children (WISC-IV) in a French-speaking Swiss sample

Isabelle Reverte\textsuperscript{a,\ast}, Philippe Golay\textsuperscript{a,b}, Nicolas Favez\textsuperscript{a,b}, Jérôme Rossier\textsuperscript{c}, Thierry Lecerf\textsuperscript{a,b}

\textsuperscript{a} Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland
\textsuperscript{b} Distance Learning University, Switzerland
\textsuperscript{c} Institute of Psychology, University of Lausanne, Switzerland

\textbf{A R T I C L E   I N F O}

Article history:
Received 9 January 2013
Received in revised form 13 August 2013
Accepted 12 October 2013

Keywords:
Confirmatory factor analysis
WISC-IV
CHC theory

\textbf{A B S T R A C T}

The Wechsler Intelligence Scale for Children—fourth edition (i.e. WISC-IV) recognizes a four-factor scoring structure in addition to the Full Scale IQ (FSIQ) score: Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Working Memory (WMI), and Processing Speed (PSI) indices. However, several authors suggested that models based on the Cattell–Horn–Carroll (CHC) theory with 5 or 6 factors provided a better fit to the data than does the current four-factor solution. By comparing the current four-factor structure to CHC-based models, this research aimed to investigate the factorial structure and the constructs underlying the WISC-IV subtest scores with French-speaking Swiss children (N = 249). To deal with this goal, confirmatory factor analyses (CFAs) were conducted. Results showed that a CHC-based model with five factors better fitted the French-Swiss data than did the current WISC-IV scoring structure. All together, these results support the hypothesis of the appropriateness of the CHC model with French-speaking children.

\ast 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).
\ast\ast Corresponding author at: FSEP — Psychology, University of Geneva, 40, Bd. du Pont d’Arve, CH-1205 Geneva, Switzerland. Tel.: +41 22 379 92 80.
E-mail addresses: isabelle.reverte@unige.ch (I. Reverte), thierry.lecerf@unige.ch (T. Lecerf).

1. Introduction

Since 1949, the interpretation of the WISC scores has changed from a two-factor (WISC; Wechsler, 1949) to a four-factor solution (WISC-IV; Wechsler, 2005). The last version, the WISC-IV, distinguishes four indices: Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Processing Speed (PSI) and Working Memory (WMI), in addition to the Full Scale IQ (FSIQ) score. This last version represents a significant revision and is more psychometrically and theoretically grounded (Grégoire, 2009). However, the WISC-IV still presents some important limitations concerning the relation between the interpretation grid and contemporary theory. Indeed, the factorial structure of the WISC-IV is not completely aligned with the Cattell–Horn–Carroll (CHC) model of intelligence measurement, which is the consensual psychometric-based model of human cognitive abilities (Alfonso, Flanagan, & Radwan, 2005). Several studies conducted on the WISC-IV subtest scores demonstrated that CHC-based models were more adequate to the data than the four-factor scoring structure (Flanagan & Kaufman, 2004; Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Lecerf, Rossier, Favez, Reverte, & Coleaux, 2010). It should be noted that similar analyses were recently conducted on the latest version of the Wechsler Adult Intelligence Scale (i.e. WAIS-IV), and results also suggested that a CHC-based model provided a better description of the WAIS-IV subtest scores than did the four-factor structure (Benson, Hulac, & Kranzler, 2010). Taken together, these findings reveal that the interpretation of the WISC-IV and the WAIS-IV subtest scores may be improved by applying the CHC structure. The main purpose of the present investigation was to determine whether CHC theory-based models were also more adequate than the four-factor scoring structure with French-speaking Swiss children.

1.1. The CHC model

The CHC theory of cognitive abilities represents the integration of the Cattell–Horn \textit{Gf–Gc} theory (Horn & Cattell, 1967; Horn & Noll, 1997) and Carroll’s three-stratum level theory (Carroll, 1993). In the CHC model, cognitive functioning is subdivided into one general factor (Spearman’s \textit{g} factor), 16 broad abilities, and about 100 narrow abilities (Newton & McGrew, 2010). Some of the most studied broad abilities include Fluid reasoning (\textit{Gf}), Comprehension–Knowledge (\textit{Gc}), Visual processing (\textit{Gv}), Short-term memory (\textit{Gsm}), Processing speed (\textit{Gs}) and Quantitative knowledge (\textit{Gq}). Because the CHC theory is currently the most empirically supported psychometric model of cognitive abilities, it had a significant influence on the measurement of intelligence. For instance, the developers of the Kaufman Assessment Battery for children, or the Woodcock–Johnson Tests of cognitive abilities, applied the CHC theory to the recent revisions (Kaufman, Kaufman, Kaufman-Singer, & Kaufman, 2005; Schrank, 2005).

As mentioned before, although the WISC-IV and the WAIS-IV subtest scores measure several dimensions central to the CHC theory, the
developers of these scales did not refer to the CHC terminology. So, the Wechsler factor index scores have been titled Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed, instead of Comprehension–Knowledge (Gc), Short-term memory (Gsm), Processing speed (Gs), etc. Nevertheless, it has been shown that the WISC-IV subtest scores could be interpreted according to the CHC model. In particular, Flanagan and Kaufman (2004) suggested that the subtest scores of the WISC-IV measure six CHC factors (Gf, Gc, Gsm, Gv, Gq, and Gs), while Keith et al. (2006) showed that the subtest scores measure mainly five factors (Gf, Gc, Gsm, Gv, and Gs). Thus, although the WISC-IV scoring structure is consistent with CHC theory, several important differences exist. For instance, the interpretation of the Perceptual Reasoning index (PRI) is not congruent with the CHC model, because this latter theory distinguishes Gf and Gv. In other words, PRI could be considered as a mixture of Fluid reasoning and Visual processing. In the present study, we investigated the factorial structure of the WISC-IV with French-speaking Swiss children by comparing the current four-factor structure to CHC-based models.

Following this first debate on the factorial structure of the WISC-IV, a second controversy concerns the cognitive constructs measured by each subtest score. Let us recall that Switzerland is a country with three linguistic regions: one German part, one Italian part, and one French part. Because there is no specific standardization data available for the WISC-IV in the French Speaking part of Switzerland, the adaptation of the WISC-IV developed in France is used instead. Following APA’s Standards for Psychological and Educational Testing, guidelines from the International Test Commission and French/Swiss national’s recommendations, it is needed to justify the new use of a psychological instrument if it is used in a way that has not been validated (i.e., with French speaking Swiss children instead of French children). Because this information has been sadly lacking to this date, evidence and theory bearing on the intended use and interpretation should be provided and is indeed one objective of the present study.

Even if France and French-speaking Switzerland are very close geographically, there is indeed no guarantee that cultural and linguistic backgrounds are strictly equivalent. It has also been demonstrated that culture and language loadings influence intelligence test performance (Ortiz, Ochoa, & Dynda, 2012). Thus, the question of the validity of the WISC-IV subtest score interpretation remains prominent with French-Swiss children, because some cross-cultural studies have shown similarities and discrepancies of the constructs measured by each WISC-IV subtest score. For instance, Chen, Keith, Chen, and Chang (2009) investigated the structural validity of the WISC-IV with Taiwanese children. These authors demonstrated that the four-factor scoring structure and the CHC-based model were both supported in each country. In contrast, Keith et al. (2006) demonstrated that the four-factor scoring solution was less adequate with U.S. children. Most importantly, regarding the cognitive constructs measured by each subtest score, several findings reported by Chen et al. (2009) were different than those reported by Keith et al. (2006), suggesting cultural or linguistic specificities. For instance, for Taiwanese children, the Picture Completion score may be considered as a measure of Visual processing (Gv), while it may be considered as a measure of both Visual processing and Comprehension–Knowledge (Gc) for U.S. children. Lecerf et al. (2010) conducted CFA on the French standardization sample data to determine whether the WISC-IV subtest scores measured VCI, PRI, WMf and PSI, or whether the WISC-IV could be better described with a CHC-based model. Results suggested that a CHC theory-based model with 6 factors (Gf, Gc, Gsm, Gv, Gq, and Gs) better fitted the French data. Regarding the cognitive constructs measured by each subtest score, several findings reported by Lecerf et al. (2010) were different from those reported by Keith et al. (2006) or from those reported by Chen et al. (2009). For instance, for U.S. children, Arithmetic score might be considered as a measure of Fluid reasoning (Gf), while it might be considered as a measure of Quantitative knowledge (Gq), Short-term memory (Gsm) and Comprehension–Knowledge (Gc) for French children.

By extension, prior to any interpretations of the performance of French-speaking Swiss children on the WISC-IV subtest scores, it should be demonstrated that subtest scores have comparable meaning for French and Swiss-French children. To clarify this issue, several methodological approaches are appropriate. Because the comparison between the standard four-factor model and a CHC-based model was the prime focus of this study, confirmatory factor analyses (CFAs) were conducted. More precisely, in a first step, the current four-factor scoring model was compared with several alternative models with hypothesized cross loadings (models A1 to A4).

The basic four-factor model was based on the scoring structure reported in the French WISC-IV technical manual. Because a higher-order factor conceptualization of intelligence is often postulated, and in order to compare our findings with those reported in previous studies (Chen et al., 2009; Keith et al., 2006; Lecerf et al., 2010), we tested higher-order models which included four lower-order factors (VCI, PRI, WMf, and PSI) and one higher-order general factor (g).

In a second step, several CHC theory-based models with hypothesized cross loadings were tested and compared (models B1–B7). The initial CHC model was the one reported by Keith et al. (2006). We specified loadings and cross-loadings for subtest scores in a priori based on the technical manual of the WISC-IV (Wechsler, 2005), and the hypotheses formulated by Flanagan and Kaufman (2004), by Keith et al. (2006), by Chen et al. (2009) and finally by Lecerf et al. (2010). Thus, alternative competing models were tested (1) to determine whether a CHC-based model provided a better explanation of the WISC-IV subtest scores than the four-factor structure; and (2) to determine more precisely, what was the nature of the constructs measured by the subtest scores.

2. Method

2.1. Participants

A total of 249 children aged from 8 to 12 years (124 males, mean age = 10.16, SD = 1.12 and 125 females, mean age = 10.26, SD = 1.17) took part in this study. The children came from different schools of the Canton of Geneva (Switzerland). The sample was stratified according to socioeconomic status of the parents. The objective was to create a sample that was representative of the Geneva children population. Additionally, all children spoke and understood French well and were in the school grade appropriate to their chronological age. The administration of the WISC-IV was done during school hours in 2 sessions of approximately 45 min by trained psychologists.

2.2. Instrument

The WISC-IV has 10 core subtests and 5 supplemental subtests. First, the Verbal Comprehension index (VCI) is measured by 3 core subtest scores: Similarities, Vocabulary, and Comprehension, and 2 supplemental subtest scores, Information, and Word Reasoning. Second, the Perceptual Reasoning index (PRI) is measured by 3 core subtest scores: Block Design, Matrix Reasoning, and Picture Concepts, and 1 supplemental subtest score, Picture Completion. Third, the Working Memory index (WMf) is measured by 2 core subtest scores: Digit Span and Letter–Number Sequencing, and 1 supplemental score, Arithmetic. Fourth, the Processing Speed index (PSI) is measured by 2 core subtest scores: Coding, and Symbol Search, and 1 supplemental subtest score, Cancelation.

2.3. Analyses

Confirmatory factor analyses (CFAs) with maximum likelihood estimation were conducted using AMOS 19. Several indicators of fit were used such as the root mean square error of approximation
(RMSEA), the standardized root mean residuals (SRMRs), the Tucker–Lewis fit index (TLI), and the Comparison fit index (CFI). SRMR values less than .08 and RMSEA less than .06 indicate a good fit (Hu & Bentler, 1999). TLI and CFI suggest good fit when the values are larger than .95. When nested models were compared (i.e. one model can be derived from another by placing additional constraints), the $\chi^2$ difference (likelihood ratio test, $\Delta\chi^2$) was used to determine whether restrictions in the model resulted in a significant increase in $\chi^2$ (Loehlin, 2004). To compare non-nested models, we used the Akaike Information Criterion (AIC); the smaller AIC value suggests the better model. The secondary purpose of this study was to determine the constructs measured by each subtest score. Thus, competing four-factor models (models A1 to A4) and CHC theory-based models (models B1 to B7) with theoretically meaningful cross-loadings were tested. Each time we encountered a non-significant factor loading, it was deleted and the model was re-estimated. When there were less than 3 observed variables per latent variable (locally under-identified models), necessary restrictions were added by constraining the factor loadings to equality. When negative residual variance terms were encountered, they were fixed to 0 (Phelps, McGrew, Knopik, & Ford, 2005).

As mentioned before, the initial four-factor model was based on the standard WISC-IV scoring structure. As concerns alternative CHC models, we initially tested the model proposed by Keith et al. (2006), and by Chen et al. (2009), which was also the initial model used in our previous analyses conducted on the French children (Lecerf et al., 2010). In this model, Similarities, Vocabulary, Comprehension, Information and Word Reasoning scores were placed on Gc. Block Design and Picture Completion scores were placed on Gv, while Coding, Symbol Search and Cancellation scores were placed on Gs. Finally, Matrix Reasoning, Picture Concepts, and Arithmetic scores were placed on Gf, while Digit Span and Letter–Number Sequencing scores were placed on Gsm.

3. Results

3.1. Does the WISC-IV measure the four-factor structure with French-speaking Swiss sample?

As shown in Table 1, the first higher-order model, in which Arithmetic load on Gf did not improve the model fit (model A2, not nested, AIC = 208.75). A model in which Arithmetic score was allowed to load on both WMI and PRI (model A3) however provided a superior fit when compared with previous models ($\Delta\chi^2 = 12.07, \Delta df = 1, p < .05$). Finally, Arithmetic scores with cross-loadings to WMI and VCI (model A4) yielded a better fit value in comparison to model A3 ($\Delta\chi^2 = 11.97, \Delta df = 1, p < .05$). Because change in $\Delta\chi^2$ and AIC values were very similar for models A3 and A4, the Arithmetic score can be interpreted as a measure of WMI and PRI on the one hand or as a measure of WMI and VCI on the other hand. Overall, these results indicated that the interpretation of the Arithmetic score is complex and performance on this subtest might be explained by more than one factor.

3.2. Does the WISC-IV measure the CHC factors with French-speaking Swiss sample?

In this second step, the first and basic CHC higher-order model (model B1) was the same model used by Keith et al. (2006), Chen et al. (2009) and Lecerf et al. (2010). This basic higher-order CHC model was used as a baseline and compared with the following CHC models. Because model B1 was not nested within model A1, we used the AIC values to compare models. In comparison with the WISC-IV models, the difference in respective AIC values suggests that the basic CHC model fitted the data better than did the basic four-correlated factor model (AIC = 203.50; model A1: AIC = 208.17).

Because the objective was to determine whether the interpretation of the WISC-IV subtest scores might be improved by applying CHC theory, several alternative CHC models with hypothesized cross-loadings were tested. These alternative models were tested one at a time and then combined to identify and validate a final CHC model. First, Flanagan and Kaufman (2004) suggested that Arithmetic score should require Gf and Gq for older children and Gsm and Gf for younger children. Other studies showed that this subtest score loaded on Gq and Gsm (Lecerf et al., 2010; Psychological Corporation), or on Gsm and Gc (Grégoire, 2009), or finally on Gq and Gs (Phelps et al., 2005). It should be noted that in contrast to the CFA conducted on the French WISC-IV (Lecerf et al., 2010), we did not test models with 6 factors, because Arithmetic was the only subtest score that appears to measure Quantitative knowledge (Gq). Thus, we must keep in mind that Arithmetic score may also be influenced by Gq. First, we tested a model in which Arithmetic score loaded on Gf, Gsm, and Gc. Although not reported here, while the $\Delta\chi^2$ difference was statistically significant ($\Delta\chi^2 = 15.04, \Delta df = 3, p < .05$), the factor loadings of the Arithmetic subtest score on Gf and on Gc were not significant (.173 and .172, respectively). Second, we tested a model in which Arithmetic score loaded on both Gsm and Gc. The AIC value suggested that allowing Arithmetic score to load on both Gsm and Gc improved the model fit in comparison to the basic CHC model (model B2, AIC = 193.04). Third, we tested a model in which Arithmetic score loaded on both Gf and Gsm. As reported in Table 1, the $\Delta\chi^2$ was statistically significant (model B3). Taken together, results indicated that the Arithmetic score could be interpreted as a measure of Gsm and Gc on the one hand or as a measure of Gsm and Gf on the other hand, because both models were plausible. These findings were congruent with previous

<table>
<thead>
<tr>
<th>Models</th>
<th>$\chi^2$</th>
<th>df</th>
<th>AIC</th>
<th>$\Delta\chi^2$</th>
<th>$\Delta df$</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>TLI</th>
<th>CFI</th>
<th>AGFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1. WISC-IV (4 factors + g)</td>
<td>140.17</td>
<td>86</td>
<td>208.17</td>
<td>.050</td>
<td>.071</td>
<td>.938</td>
<td>.949</td>
<td>.903</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2. 4 factors + g, Arithmetic on PRI</td>
<td>142.75</td>
<td>87</td>
<td>208.75</td>
<td>2.58*</td>
<td>1</td>
<td>.051</td>
<td>.051</td>
<td>.937</td>
<td>.948</td>
<td>.901</td>
</tr>
<tr>
<td>A3. 4 factors + g, Arithmetic on WMI + PRI</td>
<td>130.68</td>
<td>86</td>
<td>198.68</td>
<td>12.07*</td>
<td>1</td>
<td>.046</td>
<td>.038</td>
<td>.949</td>
<td>.958</td>
<td>.910</td>
</tr>
<tr>
<td>A4. 4 factors + g, Arithmetic on WMI + VCI</td>
<td>128.22</td>
<td>85</td>
<td>198.22</td>
<td>2.46*</td>
<td>1</td>
<td>.045</td>
<td>.033</td>
<td>.950</td>
<td>.959</td>
<td>.910</td>
</tr>
<tr>
<td>B1. Basic CHC model</td>
<td>139.50</td>
<td>88</td>
<td>203.50</td>
<td>0.67*</td>
<td>2</td>
<td>.049</td>
<td>.056</td>
<td>.942</td>
<td>.952</td>
<td>.907</td>
</tr>
<tr>
<td>A2. Arithmetic on Gm + Gc</td>
<td>127.04</td>
<td>87</td>
<td>193.04</td>
<td>12.46*</td>
<td>1</td>
<td>.042</td>
<td>.052</td>
<td>.955</td>
<td>.963</td>
<td>.914</td>
</tr>
<tr>
<td>A3. Arithmetic on Gf + Gm</td>
<td>125.54</td>
<td>86</td>
<td>193.54</td>
<td>13.96*</td>
<td>2</td>
<td>.043</td>
<td>.051</td>
<td>.955</td>
<td>.963</td>
<td>.915</td>
</tr>
<tr>
<td>A4. Block Design on Gv + Gs</td>
<td>128.76</td>
<td>87</td>
<td>194.76</td>
<td>10.74*</td>
<td>1</td>
<td>.044</td>
<td>.051</td>
<td>.953</td>
<td>.961</td>
<td>.911</td>
</tr>
<tr>
<td>A5. Matrix Reasoning on Gf + Gv</td>
<td>130.09</td>
<td>86</td>
<td>198.09</td>
<td>9.41*</td>
<td>2</td>
<td>.045</td>
<td>.054</td>
<td>.950</td>
<td>.959</td>
<td>.910</td>
</tr>
<tr>
<td>B7. Final CHC model</td>
<td>106.12</td>
<td>82</td>
<td>182.12</td>
<td>33.38*</td>
<td>6</td>
<td>.034</td>
<td>.044</td>
<td>.971</td>
<td>.977</td>
<td>.923</td>
</tr>
</tbody>
</table>

* Compared with model A1.
b Compared with previous model.
c Compared with model B1.
*p < .05.
analyses conducted on the WISC-IV that demonstrated that the Arithmetic score was a mixed and complex measure.

Word Reasoning and Similarities subtest scores mainly involve Gc (Keith et al., 2006; Grégoire, 2009; Lecerf et al., 2010). However, it has been suggested that these subtest scores may also require inductive reasoning that is Gf. Thus, we tested models in which Similarities and Word Reasoning scores were allowed to load both on Gc and on Gf. Although not reported here, the factor loadings of Similarities and Word Reasoning scores on Gf were not statistically significant. Then, we tested the possibility that these two tasks loaded on a separate verbal fluid intelligence factor. The results indicated that this model was worse than the initial CHC model (AIC = 233.67). Therefore, in the final CHC model, Word Reasoning and Similarities scores loaded only on Gc.

Block Design subtest score mainly requires Visual processing (Lecerf et al., 2010). However, some authors suggested that Block Design score involved also Fluid reasoning (Gf; Kaufman, 1994). Therefore, we tested a model in which Block Design score was allowed to load on both Gf and Gc. The goodness-of-fit statistics suggested that this model did represent an adequate fit to the data; however and most importantly, results indicated that the loading of the Block Design score on Gf was not statistically significant. According to previous analyses (Lecerf et al., 2010), we also tested the possibility that Block Design score measured Gv and Processing speed (model B4). As reported in Table 1, model B4 was significantly better than model B1 (Δχ² = 10.74, Δdf = 1; p < .05); thus, Block Design loaded on Gv and Gs in the final CHC model.

Some authors suggested that Picture Completion score involved not only Gv but also Gc (Grégoire, 2009; Lecerf et al., 2010; Phelps et al., 2005). However, a model in which Picture Completion score loaded on both Gv and Gc did not improve model fit, and most importantly the factor loading on Gc was not significant. Thus, Picture Completion score loaded only on Gv in the final model. Picture Concept score is considered to be the foremost measure of Gf. Some authors have suggested that this subtest score may also require Gf and/or Gsm and/or Gv ability. We first tested a model in which Picture Concept score loaded on Gf, Gc and Gv simultaneously. The χ² difference with model B1 did not result in a statistically significant increase in model fit, and most importantly all factor loadings were not significant (Δχ² = 4.87, Δdf = 2, p > .05). It should be noted that a model in which Picture Concept score loaded on both Gf and Gv also did not statistically improve model fit (Δχ² = 4.82, Δdf = 2, p > .05), and the Gv factor loading was not significant. In sum, in the final CHC model, Picture Concepts was considered to only substantially load on Gf.

Matrix Reasoning score is classically defined as a measure of Gf. However, it has been suggested that this type of task requires also visual processing (Gv, Carroll, 1993) and/or Working Memory (Salthouse, 1992). We tested the possibility that Matrix Reasoning score loaded on both Gf and Gv. As reported in Table 1 (model B5), the Δχ² is statistically significant. Thus, in the final CHC model, Matrix Reasoning score loaded on both Gv and Gf.

Symbol Search score is classically defined as a measure of Gs, but it has been suggested that this type of task may also require Visual processing (Gv). Thus, we tested this possibility (model B6). Results indicated that allowing Symbol Search score to load on Gv and Gs did statistically improve the model fit (Δχ² = 11.24, Δdf = 2, p < .05). Thus, in the final CHC model, Symbol score loaded both on Gv and on Gs.

3.3. Final CHC model

The final step was to re-estimate jointly the “individual” models tested before. Four subtest scores were found to have significant cross-loadings in the former individual CHC-based models: Block Design on Gv and Gs, Matrix Reasoning score on Gf and Gv, Arithmetic score on Gsm and Gc, and finally Symbol Search score on Gs and Gv. The final model (model B7) was estimated and all loadings showed to be statistically significant. The fit statistics suggest that this final CHC model provided excellent fit to the data and was the best model overall (AIC = 182.12). As concerns the second-order loadings on the g factor, results showed that Gv, Gsm and Gs had the lower loadings, respectively 0.67, 0.62 and 0.27 (Fig. 1).

4. Discussion

The purpose of this study was (1) to determine whether a CHC theory-based model provided a better explanation of the WISC-IV subtest scores than did the WISC-IV scoring structure; (2) and to assess the nature of the constructs measured by the WISC-IV subtest scores. For the French-speaking Swiss children, we found strong support for a CHC-based model with four cross-loadings: Block Design score loaded both on Gv and on Gs, Arithmetic score loaded both on Gsm and on Gc (or on Gsm and Gf), Symbol Search loaded both on Gs and on Gv, and Matrix Reasoning loaded both on Gf and on Gv. Therefore, we argue that with Swiss-French children the interpretation of the WISC-IV subtests might be improved by applying the CHC framework. To deal with this goal, normative tables for the five CHC composite scores have been developed on the basis of the French data and could be used for interpreting the subtest scores according to the CHC theory for the French WISC-IV (Lecerf, Golay, Reverte, Seni, et al., 2012) and the French WAIS-IV (Lecerf, Golay, & Reverte, 2012). Our findings were also consistent with previous data that showed that a CHC-based model in which Gf and Gv were distinct was closer to the data than models including a single Perceptual Reasoning factor (PRH) (Keith et al., 2006; Lecerf et al., 2010). In sum, concerning the first purpose of this study, results indicated that a CHC theory-based model provided a better explanation of the WISC-IV subtest scores than did the four-factor scoring structure.

Results of our work are summarized in Table 2, in which main and secondary abilities measured by each WISC-IV subtest scores are presented. First, it deserves noticing that some subtest scores measured exactly the same constructs across cultures and studies. As found by Keith et al. (2006), Chen et al. (2009), or Lecerf et al. (2010), Information, Vocabulary, Comprehension, and Word Reasoning scores provided good measures of Gc, while Digit Span and Letter–Number Sequencing scores also provided excellent measures of Gsm (Table 2). In addition, Cancelation and Coding scores provided good measures of Gs and Picture Concepts provided a measure of Gf. However, several findings of the present study were different from those reported by Keith et al. (2006) for U.S. children, and/or from the results reported by Chen et al. (2009) for Taiwanese children, and/or from those reported by Lecerf et al. (2010) for French children.

Concerning Similarities score, it should be noted that our results showed that this subtest score might be considered as a measure of Gc. This result is congruent with the data reported by Keith et al. (2006), Grégoire (2009), and Lecerf et al. (2010), but is not consistent with the results found by Chen et al. (2009). These authors indicated that Similarities score measured Gc and Gf. In contrast, for U.S., French, and Swiss-French children, Similarities measured only Gc, but not Gf.

Regarding Picture Completion, our results suggested that this subtest score is only a measure of Gv. This finding is not completely congruent with Keith et al. (2006) and with Lecerf et al. (2010), who found that Picture Completion score loaded on both Visual Processing (Gv) and Comprehension–Knowledge (Gc). Surprisingly, this finding was consistent with the result reported by Chen et al. (2009), who also found that Picture Completion score loaded only on Gv. Thus, the result of the current study was quite different from the results reported by Lecerf et al. (2010) for French children. Let us recall that while the model tested with the French children sample included 6 factors (Lecerf et al., 2010), the model tested in the present study included only 5 factors. This can help explain the discrepancy between French and Swiss-French children. However, when a model with five CHC factors was tested with the French standardization sample, results
indicated that Picture Completion still loaded significantly on both Gv and Gc. We cannot speculate that there might be a cultural difference between French and Swiss-French children and a similarity between Asian and Swiss-French children. Thus, we rather think that there might be sampling effect between French and Swiss-French children. More precisely, while analyses were conducted on a nationally representative sample of French children aged from 6 to 16 years old, our sample was limited to children aged from 8 to 12 years old. We hypothesized that older children from 12 to 16 years also used verbal processing in order to solve the Picture Completion items, while younger children relied only on Visual processing. However, further research is needed to explain this discrepancy, because psychometric

![Fig. 1. The final CHC model of the French WISC-IV in a Swiss French speaking sample.](image)

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Primary and secondary ability measures by each subtest of the WISC-IV.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Keith et al.</td>
</tr>
<tr>
<td></td>
<td>Primary</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>Gf</td>
</tr>
<tr>
<td>Matrix Reasoning</td>
<td>Gf</td>
</tr>
<tr>
<td>Picture Concepts</td>
<td>Gf</td>
</tr>
<tr>
<td>Coding</td>
<td>Gs</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>Gs</td>
</tr>
<tr>
<td>Digit Span</td>
<td>Gsm</td>
</tr>
<tr>
<td>Number Letter</td>
<td>Gs</td>
</tr>
<tr>
<td>Block Design</td>
<td>Gv</td>
</tr>
<tr>
<td>Picture Completion</td>
<td>Gv</td>
</tr>
<tr>
<td>Similarities</td>
<td>Gc</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Gc</td>
</tr>
<tr>
<td>Information</td>
<td>Gc</td>
</tr>
<tr>
<td>Word Reasoning</td>
<td>Gc</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Gc</td>
</tr>
<tr>
<td>Cancellation</td>
<td>Gs</td>
</tr>
</tbody>
</table>

Gc = Crystallized intelligence; Gf = Fluid intelligence; Gv = Visual processing; Gs = Processing speed; Gsm = Short-term-memory; Gq = Quantitative knowledge.
properties for Picture Completion score in the present study were similar to those reported with the French standardization sample. According to Keith et al. (2006), Chen et al. (2009), and Lecerf et al. (2010), results of our work indicated that the Block Design score is a measure of Visual processing (Gv) but is not a measure of fluid intelligence (Gf). In addition, consistent with the data reported by Lecerf et al. (2010) for the French children standardization sample, result of the current work indicated that this subtest score measured also Processing speed. In sum, this finding was quite similar from the results reported by Lecerf et al. (2010) and suggested that Processing speed is engaged for French and Swiss-French children in Block Design score for some children, but not for U.S. and Asian children.

According to Grégoire (2009) and Chen et al. (2009), the Matrix Reasoning subtest score measured only Fluid reasoning (Gf). In contrast, Keith et al. (2006) and Lecerf et al. (2010) suggested that this subtest score measures mainly Gf but to a lesser degree Gv. Results of our work were consistent with this latter hypothesis, because Matrix Reasoning score showed significant cross-loadings on Gf and Gv. This finding was quite similar from the results reported by Lecerf et al. (2010) and suggested that Visual processing is engaged in Matrix Reasoning score for French, Swiss-French and U.S. children. According to Keith et al. (2006) and to Chen et al. (2009), the present study indicated that Symbol Search score measured mainly Processing speed (Gs) and to a lesser degree Visual processing (Gv). This finding was quite different from the results reported by Lecerf et al. (2010) for French children. Indeed, they showed that Symbol Search score measured only Processing speed (Gs). Again, we cannot speculate that there might be a cultural difference between French and Swiss-French children and a similarity between Asian, U.S. and Swiss-French children.

Finally, the Arithmetic score seems more complex for Asian, French and Swiss-French children than for U.S. children. Concerning the discrepancy between French and Swiss-French children, the model tested with the French children showed that Arithmetic score measured mainly Gf (Lecerf et al., 2010). When a model with only five CHC factors was tested on the same data, and constrained to load on one factor, Arithmetic showed the highest loading on Working Memory (Gsm). When cross-loadings were allowed, Arithmetic score showed significant cross-loadings on Gsm (.368), on Gc (.275) and on Gf (.172). In other words, for French children, Arithmetic score measures a mixture of Short-term memory (Gsm), Comprehension–Knowledge (Gc), and Fluid reasoning (Gf). In contrast, Arithmetic score only showed a significant loading on Gsm, with no significant Gc and Gf cross-loadings with the Swiss-French sample. All together, the different results indicated that the interpretation of Arithmetic score is always complex and this study is no exception in this regard. In conclusion, this subtest score might never be interpreted in isolation.

5. Conclusion and limitations

The results of the present investigation showed that the French WISC-IV administered to French-Swiss children could better be described with 5 factors. This finding was consistent with analyses conducted on the WAIS-IV, which also demonstrated that a CHC theory-based model was more adequate (Benson et al., 2010). Nevertheless, some limitations of the present study are evident. First, this study was conducted on a rather small sample. Although we consider the sample size to be sufficient for conducting confirmatory factor analysis, this can help explain some discrepancies between French and Swiss-French children, like the difference for Symbol Search score for instance. Regarding the difference between French and Swiss-French children, we hypothesize that there might be an effect of sample. Indeed and as mentioned before, our sample was limited to children aged from 8 to 12 years old while the French sample included children from 6 to 16. However, further research is needed to clarify this discrepancy.

These results are also partly limited because the WISC-IV subtest scores were used in isolation. We believe that the analysis of WISC-IV subtest scores would be substantially improved by the addition of additional anchor measures from other intelligence tests. A third limitation concerns the fact that we did not use a cross-validation approach. Therefore, further study with a new validation sample could be needed.

Finally, because the prime focus of this study was the comparison between the four-factor model with a CHC-based model, we cannot the answer question regarding factor pattern, factor metric, factor variance, etc., invariance between both groups. Results of this paper constitute the mandatory preliminary steps to investigate factorial invariance between different groups. Correct CHC 5-factor model should be established first and invariance analysis performed next. To deal with this goal, we are currently conducting multi-group measurement and structural invariance tests.

References


