Confirmatory Factor Analyses of the Turkish Version of the WISC-IV in Non-clinical Sample: Validation of the Wechsler and CHC Models with Core and Supplemental Subtests

Cihat Çelik

Department of Psychology, Faculty of Economics, Administrative and Social Sciences, Ankara Medipol University, Ankara, Turkey

ABSTRACT

Objective: In this study, the validities of the 4-factor structure and the Cattell-Horn-Carroll (CHC) theory-based models of the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) were investigated by confirmatory factor analyses (CFA) in a Turkish non-clinical sample (n = 793).

Methods: Several models were examined and compared using CFA.

Results: Results revealed that both the 4-factor structure and the CHC-based 5 factor model were supported. However, both Wechsler and CHC bifactor models did not provide the best explanation of WISC-IV's factor structure. Across all models, while the common variance was mostly explained by general intelligence, it was least explained by the group factor in the CHC bifactor model.

Conclusion: Consequently, the factor structure of the WISC-IV Turkish is more suitable for higher-order (indirect hierarchical) models than bifactor models. In addition to the Wechsler 4-factor model, the WISC-IV also measured crystallized ability (Gc), visual processing (Gv), fluid reasoning (Gf), working memory capacity (Gwm), and processing speed (Gs). In particular, either separating Gf and Gv or combining them as the Perceptual Reasoning Index (PRI) provided a meaningful explanation. The Arithmetic subtest had significant cross-loadings. For children in Turkey, this subtest appears a reflection of Gwm and Gc in both Wechsler and the CHC higher-order models.

INTRODUCTION

Intellectual assessment is one of the most frequently used methods in the clinical assessment of children. Many intelligence tests were developed for use in clinical practice, education, research, and cognitive assessment (eg, Wechsler Intelligence Scales [WISC]). Wechsler batteries are the most widely used batteries in the world for the intellectual assessment of children.[1] Given their popularity, Wechsler scales have been adapted and translated for use in several countries based on the evidence providing support for measurement invariance across cultures and between normative and clinical samples.[2-4] Although the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V)[5] has come into use in the United States, the WISC-IV is still widely used by practitioners in Italy, Spain, and Turkey.[6-8] The factor structure of the WISC-IV was examined by the publisher (in the United States) using both exploratory factor analyses (EFA) and confirmatory factor analyses (CFA). These analyses revealed a 4 first-order factors.[9] Independent analyses of the normative data produced comparable factor structure

and measurement invariance across gender[2] and culture[10] for both clinical and non-clinical populations.[11,12] Although in the development of the WISC-IV, it was attempted to reflect conceptualizations of intellectual measurement influenced by the Cattell-Horn-Carroll (CHC) theory which is comprehensive taxonomy of cognitive abilities,[13] but it is not explicitly and completely in line with CHC. Since several studies have demonstrated that both the 4-factor and the CHC-based model were supported,[12,14] it was determined that CHC-based models were more adequate for the data than the 4-factor structure.[1,15,16] Nevertheless, there has been no consensus in the literature that the CHC-based models have better fits than the 4-factor structures of the WISC-IV.[3,17] Likewise, analyses of the United States,[18] Canadian,[19] and German[20] versions of the WISC-V have contested the 5-factor structure.

Studies on CHC-based broad skills have been shown to take different models in consideration. For example, Flanagan and Kaufman[1] suggested that the WISC-IV measured 6 CHC broad abilities (Gf = fluid reasoning,

Gc = crystallized intelligence, Gwm = working memory, Gv = visual processing, Gq = quantitative knowledge, and Gs = processing speed), whereas others[12,21] reported that the WISC-IV measured 5 broad abilities (Gf, Gc, Gwm, Gv, and Gs). Additionally, CHC-based 5 broad abilities were tested in different cultures.[3,14,16] In these researches, the most used WISC-IV CHC-based models included some of the basic Wechsler structures for subtests and associations related to Verbal Comprehension (VC; CHC Gc), Working Memory (WM; CHC Gwm) with/without Arithmetic, and Processing Speed (PS; CHC Gs). However, the WISC-IV Perceptual Reasoning (PR) dimension was split into 2 CHC factors as Block Design and Picture Completion by intending to measure visual processing (Gv), Matrix Reasoning, and Picture Concepts purportedly measuring fluid reasoning (Gf). Therefore, although studies show that the WISC-IV scoring structure is consistent with the CHC theory, it is also seen that there are some important differences.

Following this initial debate on the factorial structure of the WISC-IV, another disagreement is related to the cognitive constructs measured by each subtest. In addition, several WISC-IV subtests (e.g., Arithmetic, Similarities, Picture Concepts, Coding, and Symbol Search) that might measure multiple abilities and could show possible cross-loadings in CFAs have been suggested in the literature.[15,21] In particular, Arithmetic may provide a mixed measurement for fluid and quantitative reasoning, quantitative knowledge, working and short-term memory, VC, and PS.[1,14,15,21] For this reason, in this study, the structure measured by Arithmetic was specifically examined.

Another issue regarding the structure of the WISC-IV is also cultural factors. Even if the WISC-IV has been adapted in several countries, it is also known that culture and language influence intelligence test performance.[22] In some studies, it was revealed that the four-factor structure of the WISC-IV is valid in different cultures.[10,15,23] However, Keith et al.[21] demonstrated that the four-factor solution was less adequate for US children. Most importantly, regarding cognitive constructs measured by each subtest, several findings reported by researchers[14-16] were different than those reported by Keith et al.[21] and they suggested cultural or linguistic specificities. In studies conducted in different cultures, it is seen that different results have emerged in the context of the models used to examine the WISC-IV structure.[3,14,15] The Turkish version of the WISC-IV[7] is an adaptation of the WISC-IV[9] for use in Turkish-speaking children and adolescents ages 6-16 years. There is no study using EFA and CFA methods for examining the factor structure of the WISC-IV in the Turkish normative sample. Considering that some cross-cultural studies have shown similarities and discrepancies of the constructs measured by each of the WISC-IV subtests, interpretation of the WISC-IV subtests in Turkish children still maintains its importance. There is only one study in which the WISC-IV factor structure was examined in clinical and non-clinical samples using the multi-group CFA method in Turkey.[24] Findings of CFA carried out separately for the groups revealed excellent model fit indices for both the correlated first-order and second-order structure of the WISC-IV in both sample groups. However, as a result of multiple group CFA, model fit indices and factor loadings of the clinical sample were found to be better for the correlated 4-factor first-order structure compared to the non-clinical sample.[24]

Another component related to factor analysis is the factor analysis technique used. In studies using CFA, one-factor baseline model, oblique 4-factor model, higher-order factor model, and the bifactor model have been employed. The oblique four-factor model has factors corresponding to the subscales of VC, PR, WM, and PS. The higher-order factor model has first-order factors for VC, PR, WM, and PS, and a single higher-order general factor (g). In this model, the general factor captures common variances of all first-order factors, and the first-order factors capture covariances across subtests comprising the factors.[25] The bifactor model is an orthogonal model, with 5 primary factors. In this model, all subtests load on a general factor, and each subtest loads on its specific factor (VC, PR, WM, or PS). The general factor captures covariance of all subtests, and the VC, PR, WM, and PS-specific factors capture unique covariance of subtests within them after removing covariance captured by the general factor. Thus, specific factors capture their unique variance.[25] Studies comparing these models have reported more support for the bifactor model compared to the four-factor oblique model and the higher-order factor model in the clinical and normative sample.[3,23,26,27] Nakano and Watkins[28] provided support for the higher-order factor model although it differed minimally from the bifactor model. Although the bifactor model was found as a preferred solution compared to the other models, these results need to be replicated in a Turkish non-clinical sample.

The first aim of this study was to test the factor structure of the WISC-IV Turkish core and supplemental subtests by using CFA. This study would provide us to test whether the WISC-IV subtests measured the same constructs in Turkish children. It is believed that the results of this investigation will be instructive for furthering our understanding of the structure of the WISC-IV Turkish variables and for establishing evidence-based interpretive procedures for practitioners and research.

The second purpose of this study was to determine whether CHC theory-based models were also more adequate in the Turkish sample compared to the four-factor structure. For this reason, several models based on the Wechsler four-factor model and the CHC framework were compared. Thus, alternative models were examined to determine whether the CHC-based model provided a better explanation for the WISC-IV subtest scores than the four-factor structure

and to determine more precisely what was the nature of the constructs measured by the subtests.

METHODS

Participants

The sample of the study were children aged from 6 to 16 years (378 males [47.7%], mean age=9.75, SD=2.76 and 415 females [52.3%], mean age=10.16, SD=2.85). Participants were culled from a dataset of children from different schools in Turkey (considering geographical regions and socioeconomic status) who were administered WISC-IV in the WISC-IV Administration Training Program. All WISC-IV administrations were implemented by about 50 psychologists who successfully completed the practical and theoretical exams. Before administration of the test, written informed consent was obtained from parents of children assessed. Standardized administration and scoring procedures were followed by certified psychologists as outlined in the Administration and Scoring Manual. For testing models on a non-clinical sample, children with a Full-Scale IQ (FSIQ) in the range of 80-120 were included in the study. In addition, having any neurological or psychiatric diagnosis and having any problems in the sensory-motor area were determined as exclusion criteria.

Instrument

Wechsler Intelligence Scale for Children-Fourth Edition (Wechsler, 2003)[9]:   The WISC-IV, developed to assess the mental abilities of children within the range of 6-16 age, consists of 10 core and 5 supplemental subtests. In addition to standardized scores for each subtest, 4 indexes (cluster) scores and FSIQ are obtained using 10 core subtest scores.[9] These index scores are as follows: *Verbal Comprehension Index (VCI)* (Subtests: Similarities, Vocabulary, Comprehension, Information, and Word Reasoning), *PRI* (Subtests: Block Design, Picture Concepts, Matrix Reasoning, and Picture Completion), *Working Memory Index Score (WMI)* (Subtests: Digit Span, the Letter-Number Sequencing, and Arithmetic), and *Processing Speed Index (PSI)* (Subtests: Coding, Symbol Search, and Cancellation). Turkish standardization and norm study of the WISC-IV was conducted with a sample comprises 2225 children by taking into account 7 geographical regions, gender, and socioeconomic level to represent each age equally.[7] Mean for index scores and FSIQ is 100 and the standard deviation is 15. For subtest standardized scores, the mean is 10 and the standard deviation is 3.

Procedure and Analyses

All analyses were conducted using 10 core and 5 supplemental subtest scores. Descriptive statistics were performed through the Statistical Package for the Social Sciences (SPSS) version 21.0 (IBM SPSS Corp.; Armonk, NY, USA). CFAs were conducted to evaluate what factor structure WISC-IV displays in Turkish children.

In CFAs, maximum likelihood estimation was employed using AMOS 21. Several indicators of fit were used, such as root mean square error of approximation (RMSEA), standardized root mean residuals (SRMRs), Tucker–Lewis fit index (TLI), and comparative fit index (CFI). SRMRs values less than 0.08 and RMSEA values less than 0.06 indicate a good fit.[29] CFI and TLI suggest a good fit when their values are greater than 0.95. When nested models were compared (ie, one model can be derived from another by placing additional constraints), χ^2 difference ($\Delta\chi^2$) was used to determine whether restrictions in the model resulted in a significant increase in χ^2.[30] To compare non-nested models, Akaike information criterion (AIC) was used; the smaller AIC value suggests a better model. To compare models, ΔCFI >0.01, ΔRMSEA >0.015, and ΔAIC >10 were used.

Several models were specified and examined: (A1) one factor; (A2) four oblique verbal, perceptual, WM, and PS factors; (A3, A4) 2 indirect hierarchical (higher-order) models[3] with four first-order factors (Arithmetic on WMI, Arithmetic on WMI+VCI); and (A5) a direct hierarchical (bifactor) model[23] with four first-order factors. Then alternatively, 6 CHC models (models from B1 to B6) whose cross-loadings were theoretically meaningful were tested. We first tested the model proposed by Keith et al.[21] and Chen et al.[14] This model was also the initial model used in previous analyses conducted on French children.[15,16] In this model (B1), Similarities, Vocabulary, Comprehension, Information, and Word Reasoning scores were placed on Gc. Block Design and Picture Completion scores were placed on Gv, while Matrix Reasoning, Picture Concepts, and Arithmetic scores were placed on Gf. Coding, Symbol Search, and Cancellation scores were placed on Gs, while Digit Span and Letter-Number Sequencing scores were placed on Gwm.

In the last model (B6), the CHC-based bifactor model was tested. The bifactor model hypothesizes that each WISC-IV subtest is influenced simultaneously by 2 orthogonal constructs; these constructs are the general ability factor (g) and the first-order domain-specific group factors (e.g., Gc, Gf, Gv). For this reason, omega (ω) and omega hierarchical (ω_H) for the general factor and omega-hierarchical subscale (ω_{HS}) for the group factors were estimated as model-based reliability.[31] Thus, for the estimation of the latent factor reliability, the program developed by Watkins[32] was used to calculate the coefficient, omega (ω), omega hierarchical (ω_H), and omega-hierarchical subscale (ω_{HS}). The omega program was developed based on the studies of Zinbarg et al.[33,34] and the tutorial prepared by Brunner et al.[35] Omega (ω) estimates the reliability of the latent factor by combining the general and specific factor variance. Omega hierarchical (ω_H),

which Reise[36] termed as the Omega subscale, estimates the reliability of the latent factor with all other latent construct variances removed.[35] It has been suggested that omega coefficients should exceed at least 0.50; however, 0.75 is preferred.[36,37]

RESULTS

Descriptive statistics of the WISC-IV scores are presented in Table 1. With the largest univariate skewness, kurtosis less than 1, and multivariate kurtosis less than 5, results revealed that subtest scores appeared to be relatively normally distributed.

Model fit indices in Table 2 illustrate better results from A1 through A5 models; however, since the one-factor model did not meet the combinatorial criteria (RMSEAs ≥ 0.08 and CFIs < 0.95) and since Coding subtest factor loading on g factor was non-significant, the model was inadequate.

Table 1. Descriptive Statistics of 793 Children Tested on Wechsler Intelligence Scale for Children-Fourth Edition

WISC-IV Scores	M	SD	Skewness	Kurtosis
Compsite				
Verbal Comprehension Index	101.47	12.73	0.12	−0.29
Perceptual Reasoning Index	102.14	12.22	0.04	−0.14
Working Memory Index	99.76	11.21	0.16	−0.20
Processing Speed Index	99.06	12.35	0.10	−0.38
Full Scale IQ	101.04	9.93	−0.15	−0.79
Subtest				
Block Design	9.55	2.84	0.01	−0.21
Similarities	10.15	2.68	−0.13	−0.13
Digit Span	9.72	2.41	0.27	−0.11
Picture Concepts	10.19	2.61	−0.14	−0.02
Coding	9.90	2.60	0.29	−0.02
Vocabulary	10.62	3.17	.02	−.33
Letter-Number Sequencing	10.18	2.24	−0.35	0.31
Matrix Reasoning	11.21	2.59	0.17	−0.32
Comprehension	9.96	2.34	0.05	−0.09
Symbol Search	9.78	2.50	−0.06	0.15
Picture Completion	10.25	2.42	−0.18	−0.18
Cancellation	10.07	2.73	0.21	−0.30
Information	9.00	2.70	0.06	0.06
Arithmetic	9.87	2.41	−0.17	0.05
Word Reasoning	10.50	2.50	0.01	−0.27
Multivariate				−3.59

WISC-IV, Wechsler Intelligence Scale for Children-Fourth Edition.

When looking at all goodness-of-fit indexes for the oblique four-factor model (model A2), it was seen that all fit values were within the ideal range. It is seen that the scoring model fits Turkish data well. In addition, model A3, in which Arithmetic loaded only on WM, yielded a good fit to the data. The model A2 produced a statistically significant better fit than the model A3 ($\Delta\chi^2 = 26.801$, $\Delta df = 2$, $P < .0001$). Since the AIC value of model A3 is lower than the value of model A2, it could not be said that one of the models is better. On the other hand, the model A4 produced a statistically significant better fit than model A3 ($\Delta\chi^2 = 27.818$, $\Delta df = 1$, $P < .0001$) but not model A2 ($\Delta\chi^2 = 1.017$, $\Delta df = 1$, $P > .05$). Additionally, model A4 (which is arithmetically loaded on WMI and VCI) (Figure 1) produced lower AIC than both A2 and A3. Although the bifactor model (model A5) was significantly better than other models (model A2, A3, A4), this model was inadequate because of the Coding subtest factor loading on g was non-significant. Thus, although several models showed acceptable fit among Wechsler models, the model A4 was taken as the optimum model for all WISC-IV subtests since it showed the best fit and the lowest AIC value. Hence, we highlighted model A4 among the Wechsler models.

To determine whether the interpretation of the WISC-IV subtest scores might be improved by applying CHC theory, several alternative CHC models with the hypothesized cross-loadings were tested. In this stage, the first CHC higher-order model (model B1; which is Arithmetic on Gf) was the same model used by Keith et al.[21] and others.[14,15] As shown in Table 2, model B1 fit the data well and it was used as a reference model and compared with previous Wechsler models and other CHC models. To compare the models, the AIC values and χ^2 difference were used. In comparison with the Wechsler models, the difference in the respective AIC values suggests that both model B1 (AIC = 362.763) and model B2 (AIC = 344.822) do not fit the data better than the Wechsler models (Table 2). Unlike some studies,[16,21] based on the difference in the respective AIC values, it can be said that model B1 does not improve model fit compared to the WISC-IV model. However, other studies have presented results consistent with the findings of this study.[3,23,38]

Although model B1 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 provide a better understanding of the constructs measured by the WISC-IV subtests. In the literature, it is seen that the most discussed subtest is Arithmetic. In model B1 and according to Keith and colleagues,[21] Arithmetic loads on the Gf factor. However, Arithmetic has been shown to be related to Gf, Gwm, Gc, and Gs factors.[1,14-16] Even though results revealed that model B1 had a good fit, several alternative cross-loadings were explored to understand the mixed nature of this subtest better. The results of these alternative CHC models (from model B2 to model

Table 2. Comparison of Fit of Models Testing Hypotheses About the Wechsler Intelligence Scale for Children-Fourth Edition Turkish Sample (N = 793)

Models	χ^2	df	χ^2/df	AIC	$\Delta\chi^2$	Δdf	RMSEA	90% CI RMSEA	SRMR	CFI	TLI
Wechsler models											
A1. 1 Factor	682.332	90	7.58	742.332			0.091	[0.085-0.098]	0.079	0.69	0.64
A2. Oblique 4 factors	227.013	84	2.70	329.013			0.046	[0.039-0.054]	0.044	0.92	0.91
A3. Wechsler higher order (4 factor + g)	253.814	86	2.95	321.814	26.801[a]*	2	00.050	[0.043-.057]	0.049	0.91	0.89
A4. Wechsler higher order (Arithmetic on WMI+VCI)	225.996	85	2.66	295.956	1.017[a] / 27.818[b]*	1 / 1	0.046	[0.039-0.053]	0.046	0.93	0.91
A5. Wechsler bifactor	162.125	75	2.16	282.125	64.888[a]* / 63.871[b]* / 91.689[c]*	9 / 10 / 11	0.038	[0.030-0.046]	0.038	0.95	0.94
CHC models											
B1. Arithmetic on Gf	292.763	85	3.44	362.763			0.056	[0.049-0.063]	0.053	0.89	0.86
B2. Arithmetic on Gwm	274.828	85	3.23	344.828	17.935[b]*	0	0.053	[0.046-0.060]	0.052	0.90	0.88
B3. Arithmetic on Gwm+Gf	260.076	84	3.10	332.08	32.68[d]* / 14.752[b]*	1 / 1	0.051	[0.044-0.058]	0.051	0.91	0.88
B4. Arithmetic on Gwm+Gc	219.391	84	2.61	291.391	41.369[b]* / 6.605[e]*	0 / 2	0.045	[0.038-0.052]	0.046	0.93	0.91
B5. Arithmetic on Gwm+Gf+Gc	217.542	83	2.62	291.542	1.849[b]	1	0.045	[0.038-0.052]	0.047	0.93	0.91
B6. CHC bifactor model (Arithmetic on Gwm+Gc)	158.288	76	2.08	250.288	61.103[f]* / 3.837[g]	8 / 1	0.037	[0.029-0.045]	0.040	0.96	0.94

[a]Compare with A2 model; [b]Compare with the previous model; [c]Compare with A3 model; [d]Compare with B1 model; [e]Compare with A4 model; [f]Compare with B4; [g]Compare with A5.
Model A2, A3, A5 = Arithmetic on Working Memory.
*$P < .05$.
CHC, Cattell-Horn-Carroll.

B5) were compared to model B1. These alternative models were tested one by one and then combined to identify and validate the final CHC model. Then, we compared the final CHC model with the model A4 which was the optimum model among the Wechsler models.

We first tested Arithmetic loaded only on Gwm (model B2). Model B2 yielded a good fit to the data and a statistically significant better fit than model B1 ($\Delta\chi^2 = 17.935$, $\Delta df = 2$, $P < .0001$) with lower AIC. Second, we tested whether Arithmetic is loaded on both Gf and Gwm (Model B3). As shown in Table 2, with an AIC of 332.08, model B3 provided better-fitting indices than both model B1 ($\Delta\chi^2 = 32.68$, $\Delta df = 1$, $P < .0001$) and model B2 ($\Delta\chi^2 = 14.752$, $\Delta df = 1$, $P < .0001$). The difference in the respective AIC values suggests that cross-loading of Arithmetic on Gwm and Gf improves the model fit. Third, we tested whether

Arithmetic is loaded on both Gwm and Gc (Model B4) and whether model B4 fits the data well (Figure 2). With an AIC of 291.391, model B4 resulted in a better fitting than both model A4 ($\Delta\chi^2 = 6.605$, $\Delta df = 2$, $P < .0001$) and model B3 ($\Delta\chi^2 = 41.369$, $\Delta df = 0$, $P < .0001$). Lastly, we tested the possibility that Arithmetic measured Gf, Gwm, and Gc (Model B5). As shown in Table 2, model B5 explained the data quite well. However, the χ^2 difference between models B4 and B5 was not significant ($\Delta\chi^2 = 1.849$, $\Delta df = 1$, $P > .05$). Interestingly, consistent with Chen et al.,[14] it can be said that loading Arithmetic to Gf in addition to Gwm and Gc did not make the model stronger. These results suggest that when allowed to load on multiple factors, Arithmetic primarily loads on Gwm, with a salient secondary loading on Gc. These findings were consistent with the previous analyses, which were conducted on the
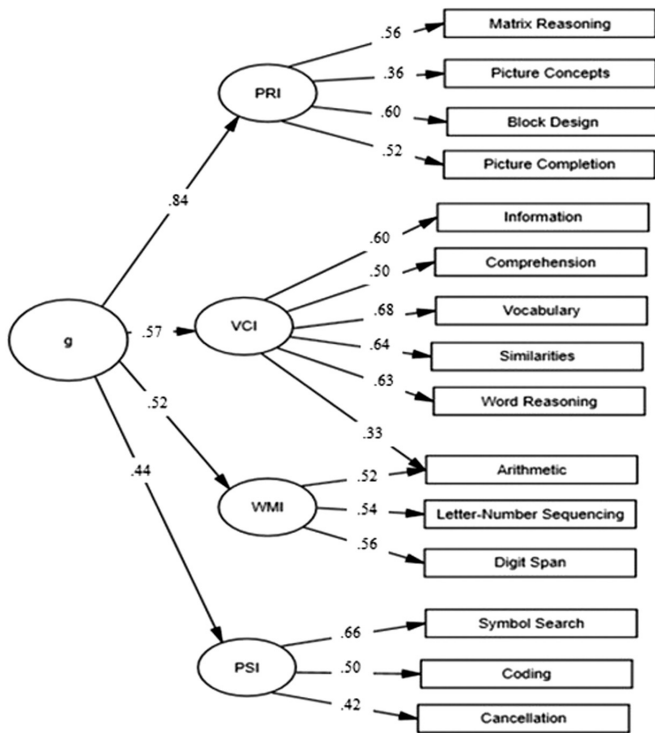
Figure 1. Wechsler indirect hierarchical measurement model, with standardized coefficients, for the Wechsler Intelligence Scale for Children-Fourth Edition (Wechsler, 2003) for 793 children. g, general intelligence; VCI, verbal comprehension factor; PRI, perceptual reasoning factor; WMI, working memory factor; PSI, processing speed factor. Note: All standardized path coefficients are significant (P < .05).

Figure 2. Cattell–Horn–Carroll indirect hierarchical measurement model, with standardized coefficients, for the Wechsler Intelligence Scale for Children-Fourth Edition (Wechsler, 2003) for 793 children. g, general intelligence; Gc, crystallized intelligence; Gf, fluid reasoning; Gv, visual processing; Gwm, working memory; Gs, processing speed. Note: All standardized path coefficients are significant (P < .05).

WISC-IV and demonstrated that the Arithmetic score was a mixed and complex measure.

Therefore, we chose model B4 as the baseline model for the CHC bifactor model. Finally, we tested model B6 in which the Arithmetic score loaded on both Gwm and Gc (Figure 3). As shown in Table 2, model B5 fits the data well. When compared to B4, the difference in the respective AIC values and $\Delta\chi^2$ suggests that model B6 fits the data better than model B4 ($\Delta\chi^2=61.103$, $\Delta df=8$, $P < .0001$) with lower AIC. Nonetheless, when the subtest factor loads of B6 are examined, it is seen that the Comprehension and Coding subtests loaded very weakly on the g factor. In addition, while model B6 was not better fit than model A5 ($\Delta\chi^2 = 3.837$, $\Delta df=1$, $P=.05$), there were no significantly fit statistical differences (ΔCFI >0.01, ΔAIC >10, and ΔRMSEA >0.015) between these models. Thus, although several models showed acceptable fit, model B4 was taken as the optimum model among CHC models. Lastly, when comparing model A4 with model B4, model B4 produced a statistically significant better fit than model A4 ($\Delta\chi^2=6.605$, $\Delta df=2$, $P < .05$) with lower AIC. However, the difference in AIC scores of the 2 models did not exceed the AIC value recommended for significance, and there
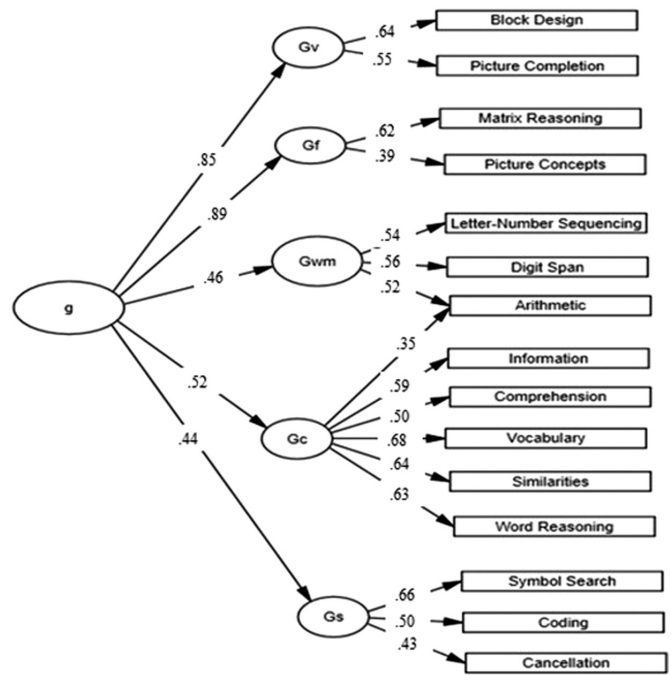
are no significant differences between the other fit index values of both models (ΔCFI >0.01, ΔAIC >10, and ΔRMSEA >0.015). Consequently, when all models are considered together, it can be said that Model A4 and Model B4 are more preferable for the WISC-IV Turkish sample.

Table 3 presents all standardized factor coefficients of the 15 subtests on the general and specific factors in model B6. As indicated, with a low coefficient score of Coding and Comprehension, all subtests showed statistically significant factor coefficients (ranging from 0.10 to 0.65) for the general factor. Significant factor coefficients were also found for all subtests for Specific factors (Gc, Gv, etc.). Coding and Comprehension were particularly poor measures of g, but relatively strong measures of the Gs and Gc factors, respectively. In contrast, Picture Concepts and Picture Completion were relatively strong measures of g, but weak measures of the Gf and Gv factors, respectively. Because of these local misfits, the CHC bifactor model was not adequate and could not be recommended. Nevertheless, Model B6 is meaningful in terms of fit indices, we also present the ECV, ω, ω_H, and ω_{HS} values for the CHC bifactor model in Table 3. As illustrated in Table 3, the ω_H value for the full test (ie, FSIQ) was 0.57 and sufficient for the confident scale interpretation; however, the ω_{HS} values for Gc, Gv, Gf, Gwm, and Gs
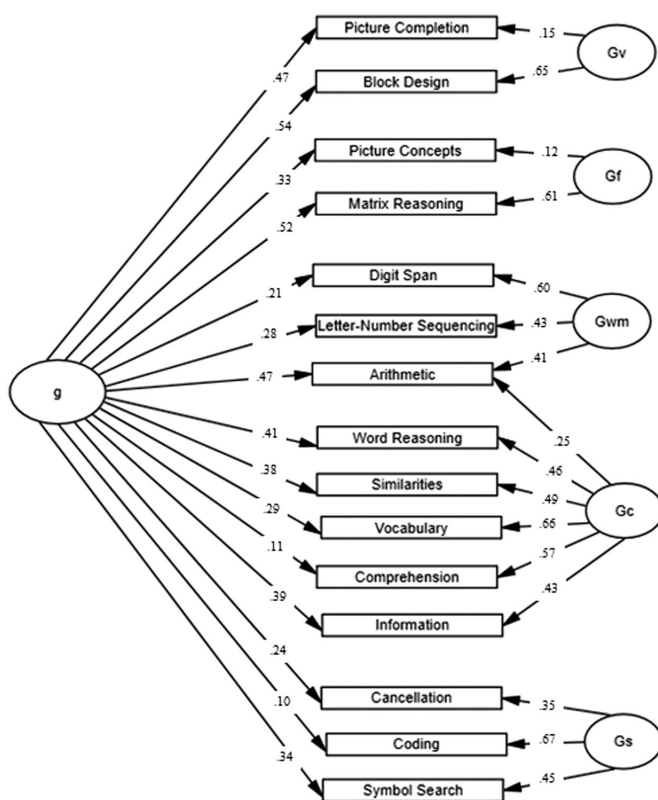
Figure 3. Cattell–Horn–Carroll direct hierarchical measurement (bifactor) model, with standardized coefficients, for the Wechsler Intelligence Scale for Children-Fourth Edition (Wechsler, 2003) for 793 children. g, general intelligence; Gc, crystallized intelligence; Gf, fluid reasoning; Gv, visual processing; Gwm, working memory; Gs, processing speed. Note: All standardized path coefficients are significant ($P < .05$).

subscales were considerably lower (Table 3). Thus, it can be said that unit-weighted scores based on the 5 group factors likely possess too little true score variance for confident clinical interpretation.[36,37]

Finally, the ω_H coefficients for the g factor in CHC bifactor models were high (0.566) and exceeded the 0.50 criterion for confident interpretation.[36,37] Explained common variance was considerably lower for the CHC group factors. The ω_{HS} coefficients for CHC-based group factors were also low and almost all of them were below the suggested minimum criterion of 0.50.[36,37] Consistent with the ECV estimates and ω_{HS} coefficients, these values suggest that "the interpretation of the subscales as precise indicators of unique constructs is extremely limited—very little reliable variance exists beyond that due to the general factor".[36]

DISCUSSION

The first aim of the study was to examine the applicability of some Wechsler and CHC models, for all subtests of the WISC-IV in a Turkish non-clinical sample. As expected,

findings supported almost all models, except the one-factor model. Thus, consistent with other studies, current findings indicate that the one-factor model was not adequate for the WISC-IV Turkish sample.[3,11,39]

Among the Wechsler models, while the Wechsler bifactor model provided a statistically better fit than other models, this model was inadequate because the Coding subtest factor loading on g was non-significant. Unlike with previous studies in non-clinical[3,23] and clinical samples,[38,40] our findings do not support the Wechsler bifactor model in 15 WISC-IV Turkish configurations. Although the Wechsler bifactor model was the best fit among the Wechsler models, Coding does not have predictive power on g; however, it is loaded more significantly by its factor, namely PS. Hence, when considering the Wechsler models, we can suggest that the Wechsler higher-order model in which Arithmetic loaded on WMI and VCI is more useful in the Turkish sample. As consistent with previous studies,[11,14] the results of the current study showed that although it supports the higher-order model of the original structure of WISC-IV, it reveals that the Arithmetic subtest was significantly loaded in both indexes. Similar findings were found in a study conducted in Turkey, and findings of CFA revealed excellent model fit indices for both the correlated first-order and second-order structure of the WISC-IV in clinical and non-clinical samples.[24] Findings related to the factor load of the Arithmetic subtest were not discussed in previous study in Turkey, since supplements subtests were not used. This study shows that the findings reveal important results in terms of demonstrating the structural validity of all subtests of WISC-IV in Turkey.

Regarding the debate about the structure of the WISC-IV, several CHC models were tested. Findings show that the WISC-IV could also be described with CHC 5-factor models. In this study, the basic CHC model (Model B1), in which Arithmetic loaded on the Gf factor, was created in line with the model proposed by Keith and colleagues.[21] Other CHC models were created according to the cross-loadings of the Arithmetic within the scope of the findings in the literature. Model comparisons, the difference in the respective AIC values and $\Delta\chi^2$ test suggest that among the CHC-based models, the CHC bifactor model describing the underlying abilities of all subtests of the WISC-IV is also acceptable. Additionally, consistent with studies,[3,23] these results suggest that the 15 subtests core and supplemental configuration may be well represented by the 5-factor CHC bifactor model. More specifically, consistent with the previous data, our results suggest that the PRI is an indicator of both Gv and Gf.[14-16] For the CHC bifactor model, all subtests showed statistically significant factor coefficients on the g factor and their specific factors. All factor coefficients on the specific factors, except 3 subtests (Picture Concepts, Picture Completion, and Arithmetic), were higher than the general. Sources of

Table 3. Factor Pattern Coefficient and Sources of Variance in Cattell-Horn-Carroll Bifactor Model (Figure 3)

Subtest	General		Gc		Gv		Gf		Gwm		Gs		h^2	u^2	ECV
	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2	b	S^2			
Similarities	0.38	0.147	0.49	0.244									0.391	0.609	0.375
Vocabulary	0.29	0.083	0.66	0.437									0.520	0.480	0.160
Comprehension	0.11	0.013	0.57	0.327									0.340	0.660	0.038
Information	0.39	0.154	0.43	0.186									0.339	0.661	0.453
Word reasoning	0.41	0.167	0.46	0.215									0.383	0.617	0.437
Block design	0.54	0.295			0.65	0.424							0.719	0.281	0.410
Picture completion	0.65	0.219			0.15	0.023							0.242	0.758	0.903
Matrix reasoning	0.52	0.269					0.61	0.373					0.643	0.357	0.419
Picture concepts	0.33	0.108					0.12	0.015					0.122	0.878	0.878
Digit span	0.21	0.043							0.60	0.356			0.400	0.600	0.108
Letter-number sequencing	0.28	0.078							0.43	0.181			0.260	0.740	0.302
Arithmetic	0.47	0.223	0.25	0.062					0.41	0.169			0.392	0.608	0.569
Coding	0.10	0.010									0.67	0.454	0.464	0.536	0.022
Symbol search	0.34	0.116									0.45	0.204	0.321	0.679	0.363
Cancellation	0.24	0.058									0.35	0.122	0.181	0.819	0.322
% Total variance		0.138		0.092		0.028		0.024		0.044		0.049	0.375	0.625	
Common variance (ECV)		0.368		0.245		0.075		0.065		0.118		0.130			
ω		0.817		0.769		0.616		0.504		0.604		0.565			
ω_H/ω_{HS}		0.566		0.508		0.239		0.216		0.417		0.466			

WISC-IV Turkish CHC Factors: Gc, crystallized ability; Gf, fluid reasoning; Gv, visual processing; Gs, processing speed; Gwm, working memory. b, standardized loading of subtest on factor; S^2, variance explained; h^2, communality; u^2, uniqueness; ECV, explained common variance; ω, Omega; ω_H, Omega-hierarchical (general factor); ω_{HS}, Omega-hierarchical subscale (group factors). All values rounded to the nearest hundredth.
*P < .05.
CHC, Cattell-Horn-Carroll.

variance estimates based on the CHC bifactor model show that the greatest proportions of variance are associated with the general factor and that the resulting 4 specific factors account for much smaller proportions of variance.

Examination of model-based reliability coefficients indicated that the g factor had strong ω_H estimates, allowing individual interpretation, in bifactor of CHC configurations ($\omega_H = 0.566$); but, the ω_{HS} estimates for the group factors were low, extremely limited for measuring unique constructs,[35-37] and likely not high enough for individual interpretation.[36,37] However, ω_H values of the CHC bifactor model were found to be lower than the values of all other versions of the WISC-IV (French, Italian, etc.). It is thought that the reason for the lower ω_H value compared to the others may be related to the sample characteristics (which is only a non-clinical sample was included) and the high contribution of the group factors to the explained total variance. For this reason, the CHC bifactor model cannot be considered an optimum model to represent the factor structure of all WISC-IV Turkish subtests. These findings are not consistent with existing data involving normal and clinical samples of children.[27,38,40] Unlike these studies, in this study, it was seen that the subtest factor loadings of the CHC bifactor model were higher in group factors compared to the general factor. For example, whereas Coding and Symbol Search were particularly poor measures of g, they were relatively strong measures of the Gs factor. A similar situation can be said for the Gwm factor where Digit Span and Letter-Number Sequencing were particularly poor measures of g, they were relatively strong measures of group factor. These findings which did not appear in previous studies suggest that although ECV and $\omega_{H/HS}$ values reveal the strength of the general factor, group factors should not be ignored for the Turkish non-clinical sample. Bifactor models may tend to produce better fit indices because they are more general than higher-order or oblique models and "relative model fit does not indicate relative model validity."[41] When considering local fit, the loading of Picture Concepts on the Gf factor was only 0.12, making it a minor influence and leaving only one salient indicator (MR) of that group factor. Likewise, the loading of Picture Completion on the Gv factor was only 0.15, again leaving only one salient indicator (BD) of group factor. Therefore, the one-indicator factors revealed in this study are thought to reduce the power of the CHC-based bifactor model. Thus, despite the good fit indices, the CHC bifactor model cannot be considered more preferable than both the Wechsler model A4 and the CHC model B4.

Following the debate on the factorial structure, another controversy is related to 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. Since it was seen that the most focused point of discussion among the subtests was Arithmetic,[14,16] we tested several models with the cross-loaded Arithmetic subtest. It is observed that while some of them state that Arithmetic loads on the Gf factor (and also on Gwm and Gc),[1,12,21] others suggest that the Arithmetic subtest primarily measures quantitative knowledge (Gq) and secondly Gwm.[9,15] While some[14] reported that it measured Gwm and Gc, others[42] proposed that it required Gq and Gs. The results of this study were consistent with the classification proposed by previous studies,[14,16] and indicated that Arithmetic appeared to measure Gc and Gwm. In addition, when Arithmetic simultaneously cross-loaded on Gc, Gwm, and Gf, the model was fit, but the loading of Arithmetic on Gf was not statistically significant in this model (model B5). Thus, in the final CHC model, Arithmetic simultaneously loaded on both Gwm and Gc and the loading of Arithmetic on Gwm was higher. In other words, for Turkish children, the Arithmetic score measures a mixture of Gwm and Gc. All these different findings indicate that the interpretation of the Arithmetic score is very complex and might never be interpreted in isolation, and cultural differences should be taken into consideration in the evaluation of this subtest.[14,16]

Although the superiority of bifactor versus higher-order models has recently received considerable attention, there is still debate about which model is better theoretically. For example, Murray and Johnson[43] found that fit indices are biased in favor of the bifactor model when there are unmodeled complexities (eg, minor loadings of indicators on multiple factors). Morgan et al.,[44] analyzed simulations of bifactor and higher-order models and confirmed that both models exhibited good model fit regardless of true structure. But when test publishers encourage users to interpret both FSIQ (g) and group factor scores and if test users interpret scores at both levels, then bifactor models appear quite necessary to disclose variance apportions because in the case of the WISC-IV the factor index scores conflate g variance and group factor variance which cannot be disentangled for individuals. Both models would provide a good estimate of general intelligence but "if 'pure' measures of specific abilities are required then bifactor model factor scores should be preferred to those from a higher-order model."[36,41,43] Chen and Zhang[45] noted that bifactor models offer conceptual clarity but "with cross-loaded items on multiple group factor or correlated group factors, the bifactor model loses its major attraction" (p. 335). However, going beyond the theoretical aspects, the findings of this study reveal that the factor structure of the WISC-IV Turkish is more suitable for higher-order (indirect hierarchical) model than bifactor models, consistent with the findings of the previous study[24] with the Turkish sample.

CONCLUSION

As a result of this study, both the Wechsler four-factor models and the CHC-based 5-factor models were supported as meaningful approaches for interpreting the

performance of the WISC-IV in the Turkish non-clinical sample. Nevertheless, we recommend both the CHC higher-order model (model B4) and Wechsler higher-order model (A4) in which the Arithmetic subtest loaded on both Gwm and Gc. Our findings improve understanding of the WISC-IV constructs across cultures. Professionals are encouraged to note not only the similarities but also the discrepancies of the underlying cognitive abilities involved in each WISC-IV score when measuring children across cultures. In particular, the significant loading of the Arithmetic subtest on Gc instead of Gf coincides with the study findings emphasizing cultural differences[14,15]. Furthermore, in addition to Gwm, the cross-loading of the Arithmetic subtest on Gc can be considered as an expected finding. Because, among the items of this subtest, there are some questions that measure the basic four operations/story problem skills, as well as questions involving fractional and speed problems that require academic gain or acquired knowledge. This reveals the importance of crystallized intelligence in acquiring Arithmetic skills. In addition, it should be noted that in the CHC bifactor model, the Arithmetic subtest was a relatively strong measure of g (general intelligence). Taken together, this result appears to indicate that the interpretation of Arithmetic is very complicated and that this subtest might never be interpreted alone.

The present results suggest that after a scale is adapted in a different culture, it is important to reconsider the factor structure of that scale in the new culture by using CFA. Thus, it is thought that it would be more functional to use this scale in the field after the aspects that overlap or differ from its original structure are revealed. Another important finding that stands out in this study is that the higher-order (indirect hierarchical) model is more acceptable compared to bifactor models for WISC-IV. Finally, this study provides practitioners with important information especially in terms of using and interpreting the appropriate subtests and factors of the WISC-IV.

On the other hand, there are some limitations of this study that need to be considered when interpreting the findings and conclusions. First, the sample were intellectually normal Turkish children, which may affect the generalization of the findings to clinically referred or diagnosed children. Thus, further research is needed to elucidate whether the WISC-IV is a valid and reliable measure of intellectual abilities in other Turkish populations. Although the cross-loadings of some subtests were also examined in the other studies, the fact that only the cross-loading of the Arithmetic subtest was examined in this study can be shown as the second limitation. Therefore, further studies including models with cross-loading of the other subtests are needed. Finally, the factor coefficients of the Coding subtest on g factor were not significant in the Wechsler bifactor model; this also reveals that this model is empirically inadequate.

REFERENCES

1. Flanagan D, Kaufman A. *Essentials of WISC-IV Assessment*. NJ: John Wiley & Sons Inc.; 2009.

2. Chen H, Zhu J. Measurement invariance of WISC-IV across normative and clinical samples. *Pers Individ Dif*. 2012;52(2):161-166. [CrossRef]

3. Kush JC, Canivez GL. Construct validity of the WISC-IV Italian editon: a bifactor examination of the standardization sample: Chi niente sa, di niente dubita. *Int J Sch Educ Psychol*. 2021;9(1):73-87.

4. Ogata K. WISC–IV factor structures of Japanese children with borderline, or deficient intellectual abilities: testing measurement invariance compared to simulated norm. *Psychol Sch*. 2019;10(6):767-776. [CrossRef]

5. Wechsler D. *Wechsler Intelligence Scale for Children*, 5th ed. San Antonio, TX: NCS Pearson; 2014.

6. Kush JC, Canivez GL. The higher order structure of the WISC–IV Italian adaptation using hierarchical exploratory factor analytic procedures. *Int J Sch Educ Psychol*. 2019;7(suppl 1):15-28. [CrossRef]

7. Öktem F, Erden G, Gençöz T, Sezgin N, Uluç S. *Wechsler Çocuklar İçin Zekâ Ölçeği-IV (WÇZÖ-IV) Uygulama ve Puanlama El Kitabı Türkçe Sürümü* [Wechsler Intelligence Scale for Children-IV: Application and Manual in Turkish]. İstanbul: Türk Psikologlar Derneği-Pearson Eğitim Çözümleri Tic. Ltd. Şti; 2016.

8. Sotelo-Dynega M, Dixon SG. Cognitive assessment practices: a survey of school psychologists. *Psychol Schs*. 2014;51:1031-1045. [CrossRef]

9. Wechsler D. *Wechsler Intelligence Scale for Children*, 4th ed. San Antonio, TX: Psychological Corporation; 2003.

10. Chen H, Keith TZ, Weiss L, Zhu J, Li Y. Testing for multigroup invariance of second-order WISC-IV structure across China, Hong Kong, Macau, and Taiwan. *Pers Individ Dif*. 2010;49(7):677-682. [CrossRef]

11. Bodin D, Pardini DA, Burns TG, Stevens AB. Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychol*. 2009;15(5):417-424. [CrossRef]

12. Weiss LG, Keith TZ, Zhu J, Chen H. WISC-IV and clinical validation of the four- and five-factor interpretative

approaches. *J Psychoeduc Assess*. 2013;31(2):114-131. [CrossRef]

13. Schneider WJ, McGrew KS. The Cattell-Horn-Carroll model of intelligence. In: Flanagan DP, ed. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. 4th ed. New York: The Guilford Press; 2018:73-163.

14. Chen HY, Keith TZ, Yung-Hwa C, Ben-Sheng C. What does the WISC–IV measure? Validation of the scoring and CHC-based interpretative approaches. *J Res Educ Sci*. 2009;54(3):85-108.

15. Lecerf T, Rossier J, Favez N, Reverte I, Coleaux L. The four- vs. alternative six-factor structure of the French WISC–IV: comparison using confirmatory factor analyses. *Swiss J Psychol*. 2010;69(4):221-232. [CrossRef]

16. Reverte I, Golay P, Favez N, Rossier J, Lecerf T. Structural validity of the Wechsler Intelligence Scale for Children (WISC-IV) in a French-speaking Swiss sample. *Learn Individ Differ*. 2014;29:114-119. [CrossRef]

17. Canivez GL, Kush JC. WAIS–IV and WISC–IV structural validity: alternate methods, alternate results. Commentary on Weiss et al. (2013a) and Weiss et al. (2013b). *J Psychoeduc Assess*. 2013;31(2):157-169.

18. Canivez GL, Watkins MW, Dombrowski S. C. Factor structure of the Wechsler Intelligence Scale for Children–Fifth edition: exploratory factor analyses with the 16 primary and secondary subtests. *Psychol Assess*. 2016;28(8):975-986. [CrossRef]

19. Watkins MW, Dombrowski SC, Canivez GL. Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children-Fifth edition. *Int J Sch Educ Psychol*. 2018;6(4):252-265. [CrossRef]

20. Pauls F, Daseking M, Petermann F. Measurement invariance across gender on the second-order five-factor model of the German Wechsler Intelligence Scale for Children–Fifth edition. *Assessment*. 2020;27(8):1836-1852. [CrossRef]

21. Keith TZ, Fine JG, Taub GE, Reynolds MR, Kranzler JH. Higher-order, multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth edition: what does it measure. *Sch Psych Rev*. 2006;35(1):108-127. [CrossRef]

22. Ortiz S, Ochoa S, Dynda A. Testing with culturally and linguistically diverse populations: moving beyond the verbal-performance dichotomy into evidence-based practice. In: Flanagan AK, ed. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. New York: The Guilford Press; 2012:526-552.

23. McGill RJ, Canivez GL. Confirmatory factor analyses of the WISC–IV Spanish core and supplemental subtests: validation evidence of the Wechsler and CHC models. *Int J Sch Educ Psychol*. 2018;6(4):239-251. [CrossRef]

24. Celik C, Yigit I, Yigit MG, Erden G. Examining the factor structure of the WISC-IV in clinical and non-clinical samples: a multiple-group confirmatory factor analysis. *Dusunen Adam*. 2020;33(3):296-330.

25. Gomez R, Vance A, Watson SD. Structure of the Wechsler Intelligence Scale for Children-Fourth edition in a group of children with ADHD. *Front Psychol*. 2016;7:737. [CrossRef]

26. Canivez GL. Construct validity of the WISC–IV with a referred sample: direct versus indirect hierarchical structures. *Sch Psychol Q*. 2014;29(1):38-51. [CrossRef]

27. Watkins MW, Canivez GL, James T, James K, Good R. Construct validity of the WISC–IV-UK with a large referred Irish sample. *Int J Sch Educ Psychol*. 2013;1(2):102-111. [CrossRef]

28. Nakano S, Watkins MW. Factor structure of the Wechsler Intelligence Scales for Children–Fourth edition among referred Native American students. *Psychol Schs*. 2013;50(10):957-968. [CrossRef]

29. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6(1):1-55. [CrossRef]

30. Loehlin JC. *Latent Variable Models: an Introduction to Factor, Path, and Structural Analysis*, 4th ed. Hillsdale, NJ: Erlbaum; 2004.

31. Gignac GE, Watkins MW. Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behav Res*. 2013;48(5):639-662. [CrossRef]

32. Watkins MW. Omega Computer Software. Computer software [Computer program]. Phoenix, AZ: Ed & Psych Associates; 2013.

33. Zinbarg RE, Revelle W, Yovel I, Li W. Cronbach's α, Revelle's β, and McDonald's ωh: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*. 2005;70(1):123-133. [CrossRef]

34. Zinbarg RE, Yovel I, Revelle W, McDonald RP. Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for ωH. *Appl Psychol Meas*. 2006;30(2):121-144. [CrossRef]

35. Brunner M, Nagy G, Wilhelm O. A tutorial on hierarchically structured constructs. *J Personal*. 2012;80(4):796-846. [CrossRef]

36. Reise SP. The rediscovery of bifactor measurement models. *Multivariate Behav Res*. 2012;47(5):667-696. [CrossRef]

37. Reise SP, Bonifay WE, Haviland MG. Scoring and modeling psychological measures in the presence of multidimensionality. *J Pers Assess*. 2013;95(2):129-140. [CrossRef]

38. Fenollar-Cortés J, López-Pinar C, Watkins MW. Structural validity of the Spanish Wechsler Intelligence Scale for Children-Fourth edition in a large sample of Spanish children with attention-deficit hyperactivity disorder. *Int J Sch Educ Psychol*. 2019;7(suppl 1):2-14. [CrossRef]

39. Devena SE, Gay CE, Watkins MW. Confirmatory factor analysis of the WISC–IV in a hospital referral sample. *J Psychoeduc Assess*. 2013;31(6):591-599. [CrossRef]

40. Canivez GL, Watkins MW, Good R, James K, James T. Construct validity of the Wechsler Intelligence Scale for Children-Fourth UK edition with a referred Irish sample: Wechsler and Cattell–Horn–Carroll model comparisons with 15 subtests. *Br J Educ Psychol*. 2017;87(3):383-407. [CrossRef]

41. Reise SP, Bonifay W, Haviland MG. Bifactor modelling and the evaluation of scale scores. In: Irwing PTB, Hughes DJ, eds. *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*. Chichester: Wiley-Blackwell; 2018:675-707.

42. Phelps L, McGrew KS, Knopik SN, Ford L. The general (g), broad, and narrow CHC stratum characteristics of

the WJ III and WISC–III tests: a confirmatory cross-battery investigation. *Sch Psychol Q.* 2005;20(1):66-88. [CrossRef]

43. Murray AL, Johnson W. The limitations of model fit in comparing bi-factor versus higher order models of human cognitive ability structure. *Intelligence.* 2013;41(5):407-422. [CrossRef]

44. Morgan GB, Hodge KJ, Wells KE, Watkins MW. Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi factor models via Monte Carlo simulations. *J Intell.* 2015;3(1):2-20. [CrossRef]

45. Chen FF, Zhang Z. Bifactor models in psychometric test development. In: Irwing PTB, Hughes DJ, eds. *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development.* Hoboken, NJ: John Wiley & Sons Inc.; 2018:325-345.