RESEARCH ARTICLE

Examining the factor structure of the WISC-IV in clinical and non-clinical samples: a multiple-group confirmatory factor analysis

Cihat Celik[1], Ibrahim Yigit[1], Melike Guzey Yigit[2], Gulsen Erden[2]

[1]Ankara Medipol University, Faculty of Economics, Administrative and Social Sciences, Department of Psychology, Ankara - Turkey
[2]Ankara University, Faculty of Language, History and Geography, Department of Psychology, Ankara - Turkey

ABSTRACT

Objective: The aim of this study is to examine the factor structure and measurement invariance of the Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV), which is used for intellectual assessment in clinical and non-clinical samples, through using the Confirmatory Factor Analysis (CFA) method.

Method: The sample of the study consisted of two groups, of which 415 were clinical (those having attention deficit hyperactivity disorder, specific learning disability, and intellectual disability) and 437 were developmentally normal children without any clinical diagnosis. In order to test first-order (the correlated four-factor model) and higher-order factor structure (model including general intelligence factor) of the WISC-IV, CFA was performed separately for each group. Furthermore, multiple group CFA was also performed to examine measurement invariance across groups.

Results: 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 four-factor first-order structure compared to the non-clinical sample.

Conclusion: The findings suggest that the WISC-IV index and subtest scores measure the same cognitive structure for both groups.

Keywords: Confirmatory factor analysis, intellectual assessment, measurement invariance, structural validity, WISC-IV

INTRODUCTION

Intellectual assessment is one of the methods used in the clinical assessment of children who are referred due to the problems they experience in the academic field or during their placement in educational intervention programs. In such assessments, scores obtained from intelligence tests are often used to distinguish the diagnosis associated with the difficulties experienced by a child (e.g., to distinguish learning difficulties from mental retardation) (1,2) or to identify the strengths or weaknesses in the child's cognitive structure (3-5). In addition to using the information on the mental or cognitive profile of the child in the clinical diagnosis process, it can also provide resources for intervention programs towards

children's difficulties (3,6,7). In this respect, it is very essential that intelligence tests, which have an important place in the decisions regarding the academic life of children during and as well as the following assessment process, have an up-to-date, valid and reliable structure, and these are always among the required qualifications (8-10). Wechsler Intelligence Scales for Children (WISC) are widely used in intelligence assessment in our country as well as abroad. The standardization and norm study of WISC-R, the revised form of WISC, started in 1978 in our country and has been used in practice areas since then (11). Following this version, the third version of WISC (WISC-III) was put into use abroad, but the standardization and norm work of this version was not carried out in our country. With the fourth version, the WISC-IV, the current norms of the scale was established and started to be used in our country. Compared with the WISC-R, significant innovations and changes were made with the WISC-IV in terms of the structure, content and interpretation of the scale (12). Accordingly, the dual intelligence structure (Verbal and Performance Intelligence Quotients) that existed in WISC-R and continued in WISC-III was completely removed and replaced by four new structures: Verbal Comprehension, Perceptual Reasoning, Processing Speed, and Working Memory. Some subtests were removed from the scale (e.g Picture Completion) by adding six new subtests (eg, Matrix Reasoning). Another change is that Information and Arithmetic subtests, which had used as core subtests in previous versions, were moved into supplemental subtests. Thus, with its new structure, the WISC-IV started to be used by enriching the expected contributions in the field from the intelligence assessment. At the same time, research findings showing the validity of this new structure have started to take place in the literature.

One of the most important evidence that tests are accurately measuring what it want to measure is that the test has structure validity. Structure validity refers to the appropriateness of the results and interpretations obtained from test scores to the structure measured by the test (13,14). It is seen that factor analysis studies are the most commonly used methods in evaluating the structure validity. When considering the measurement of intelligence, which is an abstract concept, it is expected to prove that the factors that make up the intelligence tests (e.g. verbal comprehension) measure the structures that is actually intended to be measured. Therefore, both in the development of intelligence tests

and in the subsequent processes, in testing the structure of the scale mainly exploratory factor analysis (EFA) or confirmatory factor analysis (CFA) methods are used (13,15). By evaluating the structure validity both through EFA and CFA, it is revealed whether the factor structure of the test tool is compatible with the relevant theoretical structure (16). While EFA is used to determine the factor structure of the scale, CFA is used to verify a structure that has been theoretically established or tested previously (17). In this context, it is inevitable to examine the structure validity of the developed test or scale, since it will not be possible to understand which structure is actually measured without a clear evidence of structure validity (14). Earlier versions of the Wechsler intelligence scales (especially WISC-R and WISC-III) often tested the structure of the scale using EFA and CFA methods (18-24), whereas in later versions (e.g. the WISC-IV), often highlighted the CFA method (25-31). However, in many countries including ours, it is observed that the WISC-IV continues to be widely used in intelligence assessments (for clinical or research purposes), since there has not yet been a norm study for the WISC-V (32-35). Therefore, in the literature, it is seen that the factor structure of the core subtests of the WISC-IV is generally examined in samples composed of children who are both normally developing and who are referred to clinical evaluation (16,25,30,36-39).

In general, most of the studies conducted with both clinical and non-clinical samples tested the correlated four-factor first-order model, the higher-order factor model, and direct or indirect hierarchical factor models. Among these factor models, the most tested one is the first-order factor model which corresponds to the original correlated four-factor structure of the WISC-IV (Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed). The correlated four-factor model has subtests under each factor that make up that factor, and these four factors are correlated with each other. In second-order models, the "g" factor representing general intelligence is also included in the model and two models can be formed according to the position of the "g" factor. The first of these is the indirect second-order model in which the "g" factor is positioned above the four factors (this is the model tested in this study), the other is the direct second-order model in which the subtests are directly connected to both "g" and four factors. Many studies on the WISC-IV reveal that the correlated four-factor structure is confirmed in both non-clinical (40,41) and clinical samples (30,31,38,42).

The WISC-IV CFA studies carried out with non-clinical samples revealed that usually the norm group data for that country are used (26,30,31,42,43) and the correlated four-factor first-order structure of the WISC-IV is confirmed (24,28,35). On the other hand, in some studies, first and second-order models and five-factor models (these are the models that was formed within the frame of CHC theory, in which the perceptual reasoning cluster is taken in two separate clusters called visual processing and fluid reasoning), and this models were observed to be confirmed (29,36). It is also worth noting that there are a small number of CFA studies in literature using non-clinical samples as different from norm group data (44,45). For example, a study by Reverte et al. (45) conducted with a non-clinical sample reported that the five-factor model had a better fit, although the second-order four-factor model was significant. Similarly, Golay et al. (44) demonstrated in their study that the four-factor model was confirmed, but the direct hierarchical (bifactor) model was stronger than the other models in their study with a non-clinical sample.

On the other hand, when the WISC-IV CFA studies conducted with clinical samples are examined, it is seen that homogeneous or mixed sample groups are used. Accordingly, in studies conducted with children diagnosed with attention deficit hyperactivity disorder (ADHD), different results were obtained in which other models were also tested in addition to correlated four-factor models (46-50). It was reported that this model was confirmed in Taiwanese children with ADHD diagnosis whom the correlated four-factor model was tested (50). In contrast, another study in which second-order models were tested found that the second-order model provided better results in children with ADHD (48). Among the studies in which five-factor and direct hierarchical models were tested, there were studies reporting that the five-factor model had a better fit index in children with ADHD (43), as well as studies reporting that the direct hierarchical model was stronger (40,41). Similarly, in studies examining the WISC-IV factor structure with children only diagnosed with Specific Learning Difficulties (SLD) (37,38,51,52) or mental retardation (39), it is observed that the results differed when the mentioned different models were tested. Accordingly, in addition to studies reporting better direct hierarchical model in children with SLD diagnosis (32,33), there is also a study revealing that there is no difference between significant models (45). In their study in which they tested the indirect hierarchical model and the correlated four-factor model

in children with SLD, Giofrè and Cornoldi (46) reported that both models were confirmed, but the Picture Concepts subtest in the correlated four-factor model was significantly loaded on both verbal comprehension and perceptual reasoning factors. Therefore, they suggested that the "g" factor measured by the WISC-IV was represented theoretically poor in children with SLD and that the general intelligence score might be inadequate to account for the general cognitive functions of these children (46).

Apart from these homogeneous groups mentioned, children in different diagnosis groups (e.g. ADHD, SLD, mental retardation) were mostly observed to be discussed together (16,25,28,31,36,42,53). In these mixed clinical groups, it is observed that many the WISC-IV related models were tested and different results were obtained. In studies where the correlated four-factor model was tested, it was reported that this model was confirmed, but the direct hierarchical model better fit to the data (14,26,47). In the study conducted by Bodin et al. (23) with a mixed neuropsychological sample (ADHD, epilepsy, learning difficulties, traumatic brain injury, cerebral palsy, meningitis encephalitis, etc.) between the ages of 6 and 16, the correlated four-factor model was confirmed, but it was stated that the second-order indirect hierarchical model was more preferable with regard to some values of fit index. In another partially similar study that the factor structure of the WISC-IV was examined in a mixed clinical group of children aged 6-16 with learning difficulties, mental retardation, autism, brain damage and other mental/medical problems, and direct and indirect hierarchical models were confirmed with the correlated four-factor model (31). However, it was reported that the indirect hierarchical model has a better fit index.

When the findings of separately conducted studies on clinical and non-clinical samples are evaluated in general, it is observed that the correlated four-factor model of the WISC-IV fit to the data in both samples. In summary, it was suggested that all other tested models were confirmed but in some studies the direct hierarchical model was fit better (32,33,40,41), while others suggested that the indirect model was more tenable (22,31,46). Similar with both the clinical and non-clinical samples, this table led researchers to multiple-group studies dealing with different groups together in examining the factor structure of the WISC-IV.

This is considered within the scope of measurement invariance, in other words, to examine whether the structure measured by an intelligence test is similar in

different sample groups. Measurement invariance is recognized as an important feature for any clinical tool used to compare individuals in different groups (54,55). Accordingly, the structure of a scale that tries to reveal the differences between groups is predicted to be similar for each group and the results obtained from the measurement tool are expected to have equal psychometric properties (56,57). In other words, it is pointed out that even if the scale is used in different sample groups, it should measure a similar structure. Therefore, failure to prove measurement invariance for a specific test or scale may eliminate the ability of the projected structure to be used in intergroup comparisons (55). Since the WISC-IV is used in clinical practice as part of the diagnostic assessment, the cluster and subtest scores are presumed to measure the same structure in both non-clinical (or normative) and clinical samples. In this respect, it is considered the fact that the WISC-IV measures a similar structure in all groups assessed should be demonstrated. As mentioned before, in previous studies, the factor structure of the WISC-IV was examined separately in clinical and non-clinical sample groups (16,53); However, recent studies showed that these groups were discussed together using the multiple-group CFA method (26,31,42,58).

It is stated that the multiple-group factor analysis approach is methodologically more favorable compared to the single group factor analysis approach, since it compares the factor structure obtained from a particular group with the norm group or various groups (58). For this reason, testing the factor structure of the WISC-IV through multiple group comparisons has come to the fore in recent years. In some of the studies conducted within this context, the factor structure of the WISC-IV was examined by making multiple group factor analysis comparisons in clinical and non-clinical sample groups (52,58,59). Besides, there are also studies in which multiple-group comparisons are made with reference to gender (60) and different cultural norms (26). In all these studies, the correlated four-factor structure of the WISC-IV is supported in terms of measurement invariance; in other words, it was demonstrated that the WISC-IV measures similar structure in both clinical and non-clinical groups.

Becoming widespread in our country in the fields of clinical assessment, education and research, the WISC-IV is one of the most up-to-date intelligence tests (34). Considering the use of this test in many areas and different clinical groups, it is thought that the structural validity and measurement invariance should be tested in our country, regardless of the different sample groups. Although the factor structure of the WISC-R, one of the earlier versions of WISC, was tested with EFA and CFA methods in our country (61,62), it has been realized that there is no study examining both the factor structure of the WISC-IV in different diagnosis groups and the measurement invariance in these diagnostic groups. The main objective of this study is to determine the factor structure of the WISC-IV (correlated four-factor first-order model and second-order model) CFA in clinical samples (children diagnosed with ADHD, SLD, and intellectual disability) and non-clinical samples (children without any clinical diagnosis). It is also aimed to test the measurement invariance of these samples using a multiple-group CFA method. Multiple-group factor analyses are seen to be more favorable than single-group analyses. However, such studies dealing with clinical and non-clinical samples together are in limited numbers. Therefore, the study will contribute to the literature by supporting a limited number of studies. With this study, it is considered that this widely used test will present important findings regarding its validity, especially in clinical diagnostic groups, and these findings will contribute to the practitioners and researchers in the relevant field. Furthermore, given that the WISC-IV was developed in a different culture, has been adapted to our culture and its norms have been developed, it is seen as important to reveal if this scale measures a similar structure in terms of our culture. In other words, in parallel with the studies abroad, it is considered that the findings for the cross-cultural validity of the WISC-IV will be obtained.

METHOD

Sample
The sample of the study consists of two groups as clinical (those with ADHD, SLD and intellectual disability) and non-clinical sample without any clinical diagnosis (those having full scale intelligence quotient [FSIQ] 80 and above). The clinical sample includes 415 children (intellectual disability=230, ADHD=73, learning disability=112), ranging in age from 73 to 203 months (M=131.20, standard deviation [SD]=33.29). In the non-clinical sample, there are 437 children between the ages of 73-202 months (M=130.12, SD=33.65). Of the children in the clinical sample, 165 (39.8%) are girls and 250 (60.2%) are boys. 182 (41.6%) of the children in the non-clinical sample are girls and 255 (58.4%) are boys.

Measures

Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV): The WISC-IV, the fourth version of Wechsler Intelligence Scale (WISC-IV), is an intelligence test carrying out individually to measure the intellectual ability of children between the ages of 6-16 (39). The WISC-IV consists of 10 core and 5 supplemental subtests, and FSIQ are obtained by using 10 core subtests comprising the four index scores. These indices are the Verbal Comprehension Index (VCI; core subtests: Similarities, Vocabulary and Comprehension; supplemental subtests: Information and Word Reasoning); Perceptual Reasoning Index (PRI; core subtests: Block Design, Picture Concepts and Matrix Reasoning; supplemental subtest: Picture Completion); Working Memory Index (WMI; core subtests: Digit Span and Letter-Number Sequencing; supplemental subtest: Arithmetic) and Processing Speed Index (PSI; core subtests: Coding and Symbol Search; supplemental subtest: Cancellation). The average score for all indices and FSIQ is 100, and the standard deviation is 15. For subtest scores, the average score is 10 and the standard deviation is 3. The Turkish standardization and norm study of the WISC-IV was carried out on 2225 children representing each age range in terms of geographical regions, population, gender and socioeconomic level (12,34).

Procedure

The children consisted of the clinical sample of this study were recruited from special education and rehabilitation centers. These children had diagnosed with ADHD, SLD and intellectual disability in pediatric mental health clinics according to DSM-5 diagnostic criteria. The children in the non-clinical sample were recruited through easily accessible samples. Following the WISC-IV administration, children with FSIQ 80 and above and who were indicated not to have any clinical diagnosis by their teachers and families were included in the study. For both samples, the required permission from the institutions and written consent from their families were obtained. Prior to the administration of the WISC-IV to carry out this study, the approval was obtained from the Ethics Committee of Ankara University. All scale administrations were carried out by psychologists or clinical psychologists certified by the Turkish Psychologists Association and having practice competence, following the guidelines in the Turkish Version of the WISC-IV Administration and Scoring Manual (34). Children with any neurological,

psychiatric and/or sensory-motor problems were not included in the non-clinical sample. For the clinical sample, diagnoses other than ADHD, SLD and intellectual disability were used as exclusion criteria in line with the purpose of the study. Furthermore the condition of not having been administered the WISC-IV test within a year was imposed for the two samples included in the study. All the WISC-IV record forms were checked by the authors of this study to identify possible errors, and then data entry was completed.

Statistical Analysis

All the analyses within the scope of the research were carried out via the WISC-IV's 10 core subtest scores, four index scores and the FSIQ. In a standard the WISC-IV administration, it is an indispensable protocol to administer 10 core subtests to calculate index scores and the FSIQ. However, supplemental subtests are administered when the core subtests assessments are invalid or when there is a need to learn more about intellectual abilities of the child. In this study, since not all of the supplemental subtests were administered to all children during the data collection process, they could not be included in the data analysis.

First, descriptive statistics regarding the index scores, FSIQ and subtest scores from the WISC-IV of the children included in the study were calculated using the Statistical Package for the Social Sciences (SPSS-22) program. Secondly, in order to test the correlated four factor first-order model of the WISC-IV and the second-order model in which four factors are linked to the general intelligence factor (the rest of the text will be referred to as the first and second-order models), four CFA were carried out separately. Finally, multiple-group CFA was conducted to test the measurement invariance of the factor structure of the WISC-IV between non-clinical and clinical samples. In all confirmatory factor analyses, the AMOS (Analysis of Moment Structures; Version 21) statistical package program was used and these analyses were carried out through maximum likelihood estimation method and covariance matrices. In CFA studies, the fit index values used to determine to what extent the tested or proposed models was complied with data are as follows: Chi-square value to degree of freedom ratio (χ^2/df) is less than 5; CFI (Comparative Fit Index), GFI (Goodness of Fit Index) and TLI (Tucker Lewis Index) values being .90 and above and RMSEA (Root Mean Square Error of Approximation) and SRMR (Standardized Root Mean Square Residual) values of 0.08 and below (63-65). In addition, BIC (Bayesian

Information Criteria) value was calculated as another fit index criterion. According to this criterion, the model with the lowest BIC value is considered the best-fit model (66).

Multiple-group CFA (measurement invariance) were conducted following the steps suggested by Brown (67). Accordingly, measurement invariance is carried out using models with equality restrictions. These models are configural invariance, metric invariance, factor covariance invariance and error variance invariance. The configural invariance model is a model that has no constraints and both samples are analyzed at the same time. This model is based on the assumption that the tested samples have the same number of factor and factor patterns. Testing the measurement invariance requires this model to have an acceptable fit index value and be compared with other models. Secondly, the metric invariance model is a model with equality constraints regarding factor loadings on samples. In this model, the assumption is that the factor loadings of the subtests loaded on the factors are equal between the tested samples. The next model, the factor covariance invariance model, points to the model, which has an equality constraint for factor covariances on samples. This model is based on the assumption that the relationships between latent factors are equal across samples. Finally, the tested model is the error variance invariance model. In this model, the assumption is that the error variances for the measurable variables are equal on the tested samples (56,67).

RESULTS

Descriptive Statistics
Subtest scores obtained from the WISC-IV, index scores, FSIQ and descriptive statistics are shown in Table 1.

Confirmatory Factor Analysis
Four separate CFA were conducted in both non-clinical and clinical samples to test the first-order and second-order factor structure of the WISC-IV. In these analyses, 10 core subtests of the WISC-IV are used as observable variables; Verbal Comprehension, Perceptual Reasoning, Working Memory, Processing Speed and general intelligence (g factor) were included in the analysis as latent variables.

As shown in Table 2, the analysis results for the non-clinical sample reveal that the first-order fits perfectly with the data (χ^2[df=29, n=437]=68.019, χ^2/df=2.34, CFI=0.95, GFI=0.97, RMSEA=0.06; SRMR=0.04, TLI=0.92, BIC=226.097). Similarly, the second-order model was found to fit perfectly with the data (χ^2[df=31, n=437]=72.940, χ^2/df=2.35, CFI=0.94, GFI=0.97, RMSEA=0.06; SRMR=0.04, TLI=0.92, BIC=218.858). When the factor loadings of both

Table 1: Descriptive statistics

Points, scores	Non-clinical sample (n=437)				Clinical sample (n=415)			
	Min.-Max.	Mean	SD	Variance	Min.-Max.	Mean	SD	Variance
Block design	1-19	9.64	3.18	10.14	1-16	5.54	3.46	11.97
Similarities	3-18	10.31	2.88	8.32	1-17	5.29	3.41	11.63
Digit span	2-19	10.11	2.79	7.79	1-15	5.31	3.08	9.50
Picture concepts	2-18	10.03	2.88	8.31	1-16	5.56	3.98	15.88
Coding	2-18	9.75	2.77	7.70	1-17	5.74	3.48	12.17
Vocabulary	1-19	11.61	3.66	13.41	1-17	4.86	3.70	13.68
Letter-number sequencing	1-17	10.36	2.53	6.41	1-14	5.20	3.59	12.91
Matrix reasoning	4-19	11.10	2.78	7.76	1-19	5.55	3.92	15.42
Comprehension	1-19	10.33	2.82	7.95	1-15	5.27	3.35	11.23
Symbol search	1-17	9.67	2.80	7.84	1-16	5.58	3.54	12.53
VCI	64-148	104.48	15.11	228.37	46-128	71.15	18.64	347.36
PRI	68-149	101.48	14.42	207.93	42-141	71.83	21.97	482.71
WMI	65-147	101.31	12.82	164.28	47-123	72.02	17.51	306.78
PSI	52-141	98.19	14.07	197.93	47-123	74.42	19.12	365.73
FSIQ	80-143	102.35	12.84	164.78	28-131	63.33	23.17	537.12

VCI: Verbal Comprehension Index, PRI: Perceptual Reasoning Index, WMI: Working Memory Index, PSI: Processing Speed Index, FSIQ: Full Scale Intelligence Quotient Score, Min.-Max.: Minimum and maximum values, SD: Standard deviation

models were examined, 10 core subtests were found to be significantly loaded on their own factors and the cluster scores on the general intelligence factor (p<0.001) (See Figures 1 and 2). In this sample, the first-order model and the second-order model were compared with the chi-square difference test and it was found that the two models did not differ from each other in terms of fitting (p>0.05 for $\Delta\chi^2$).

Analysis results for the clinical sample reveal that the first-order model fits perfectly with the data (χ^2[df=29, n=415]=59.079, χ^2/df=2.04, CFI=0.99, GFI=0.97, RMSEA=0.05, SRMR=0.02, TLI=0.99, BIC=215.814; See Table 2). The second-order model was also found to fit perfectly with the data (χ^2[df=31, n=415]=71.959, χ^2/df=2.32, CFI=0.99, GFI=0.97, RMSEA=0.06; SRMR=0.02, TLI=0.98, BIC=216.637; See Table 2). When the factor loadings of both models were examined, it was found that the 10 core subtests loaded significantly (p<0.001) on their own factor and the cluster scores on the general intelligence factor (See Figures 1 and 2). In the clinical sample, the first-order model and the second-order model were compared with the chi-square difference test and the two models were found to be differentiated from each other in terms of fitting. Accordingly, the first-order model has a better fit than the second-order model. Based on this finding, multiple-group CFA was performed in non-clinical and clinical samples using the first-order model.

Multiple-Group Confirmatory Factor Analysis
On clinical and non-clinical samples, multiple-group CFA were conducted to test measurement invariance related to the factor structure of the WISC-IV. The chi-square analyses prior to the multiple-group CFA show that both groups do not differentiate in terms of gender and age (gender: χ^2=0.31, p>0.05; age: χ^2=116.18, p>0.05).

As seen in Table 3, the analysis results show that the configural invariance model fits perfectly with the data (χ^2[df=58]=127.097, χ^2/df=2.19, CFI=0.98, GFI=0.97, RMSEA=0.04). In order to compare the configural invariance model with other models, χ^2 difference test is used (67). Accordingly, it is reported that if the χ^2 difference test is significant, the model with a lower χ^2 value has statistically better fit with the data (63).

In this study, the results of the χ^2 difference test carried out to compare the configural invariance model with other models with equality constraints show that the difference value ($\Delta\chi^2$) for each comparison is significant (See Table 3). These results indicate that the factor structure of the WISC-IV is invariant across the samples used in this study. Since χ^2 values are affected by the sample size (63), it has been suggested that CFI difference (ΔCFI) and RMSEA difference (ΔRMSEA) values can also be used to compare models with equality constraints (68,69). The fact that the CFI difference value is less than 0.010 and the RMSEA difference value is less than 0.015 indicates that the model is significantly invariant (68).

In this study, CFI and RMSEA difference values were found to be exceeded the values stated above when the configural invariance model is compared with other equality constraint models. Evaluating all these findings together, this study shows that the WISC-IV has a valid structure for both samples and that even though the measurement invariance could not be revealed, the factor structure for the clinical sample fits the data better.

DISCUSSION

In this study, the factor structure of the WISC-IV was examined using the multiple-group CFA method. In this context, it was tested whether the four index scores of the WISC-IV and the general intelligence (g) factor measured the similar structure in clinical and non-clinical samples. As stated before, the CFA method was often used in previous studies examining the factor structure of the WISC-IV. In these studies, it is seen

Models	χ^2/df	CFI	GFI	RMSEA	SRMR	TLI	BIC
Non-clinical sample							
Correlated four factor first-order model	2.34	0.95	0.97	0.06	0.04	0.92	226.097
Second-order model	2.35	0.94	0.97	0.06	0.04	0.92	218.858
Clinical sample							
Correlated four-factor first-order model	2.04	0.99	0.97	0.05	0.02	0.99	215.814
Second order model	2.32	0.99	0.97	0.06	0.02	0.98	216.637

Table 2: Fit index values for the tested models

CFI: Comparative Fit Index, GFI: Goodness of Fit Index, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual, TLI: Tucker Lewis Index, BIC: Bayesian Information Criteria

Table 3: Fit index values for multiple-group confirmatory factor analysis

Constrained models	χ^2	df	χ^2/df	$\Delta\chi^2$	Δdf	CFI	GFI	RMSEA	SRMR	TLI	BIC
Model 1											
Configural invariance	127.097	58	2.19	-	-	0.983	0.971	0.037	0.04	0.974	233.863
Model 2											
Metric invariance	280.444	68	4.12	153.347*	10	0.948	0.941	0.061	0.12	0.931	366.678
Model 3											
Factor covariance invariance	364.716	74	4.93	237.619*	16	0.929	0.930	0.068	0.21	0.913	438.631
Model 4											
Errror variance invariance	518.220	84	6.169	391.123*	26	0.894	0.898	0.078	0.23	0.886	571.603

*$p<0.001$, CFI: Comparative Fit Index, GFI: Goodness of Fit Index, RMSEA: Root Mean Square Error of Approximation, SRMR: Standardized Root Mean Square Residual, TLI: Tucker Lewis Index, BIC: Bayesian Information Criteria

that different models were also tested in addition to the model tests related to the four-factor structure of the WISC-IV. In addition to the models tested (structures with two, three and four factors) according to the number of default factors relevant with each other (37,47,58), higher-order/second-order CFA models were also tested (16,31). Generally, in most of these studies, it was revealed that the correlated four-factor structure of the WISC-IV was confirmed in both clinical and non-clinical samples. In this study, it was aimed to test the first and second order factor structure of the WISC-IV comparatively in clinical and non-clinical groups. The results of the research indicate that in four separate analyses, the WISC-IV is a valid measurement tool in the evaluation of children in clinical and non-clinical samples. Furthermore, while the structural validity of the WISC-IV was proven for both samples, it was revealed that the first-order model has a stronger structural validity for the clinical sample.

One of the key points of this research is that the analysis was carried out over the data collected from non-clinical sample which was new, and different from the norm group, and the factor structure of the WISC-IV in was confirmed in this sample. In terms of fit index values, it was revealed that both models created for the non-clinical sample fit perfectly with the data.

In addition, when subtest factor loadings are examined, it is observed that 10 core subtests of the WISC-IV are significantly loaded into their own factors (See Figures 1 and 2). However, in both models, it was revealed that the lowest factor loading were in the subtests in The Working Memory and Processing Speed clusters (symbol search subtest in the second-order model is not included in this scope). It is observed that this finding complied with the findings of studies conducted with norm samples abroad (31,38,60). As a

matter of fact, second-order factor analysis was used in some of these studies and it was generally assessed that low subtest factor loadings were in working memory and processing speed clusters. In addition, these studies show that subtests in verbal comprehension and perceptual reasoning clusters have higher predictive capacity for their clusters. In this study, it was found that only the Picture concepts subtest was loaded at a low level, although it was significant in the perceptual reasoning cluster in both models. It is observed that this finding complied with the findings in EFA results presented in the technical report of the WISC-IV. Accordingly, in the mentioned report, the lowest factor loading in the perceptual reasoning cluster was reported to be the picture concepts (70). This could be due to the fact that the subtest of picture concepts is also correlated with non-verbal abstraction ability. Another reason is that the picture concepts subtest may be under a fifth factor called "fluid reasoning" along with the matrix reasoning subtest under the Cattell-Horn-Carroll (CHC) theory, as discussed in some publications (31,41,71).

Another finding obtained from the non-clinical sample is that there are significant covariances among all clusters in the first-order model. However, the processing speed cluster seems to have a weaker relationship with the other clusters. While the second-order model is examined, rather than the relations between the clusters, the capacity of the clusters to load on the "g" factor is analyzed. Accordingly, it is the processing speed cluster that has the lowest loading on the "g" factor in this model. These findings are generally compatible with the findings of studies showing that processing speed is weak in predicting general intelligence (41,43). When the other inter-cluster covariances in the first-order model are examined, it is found that these relations are in the
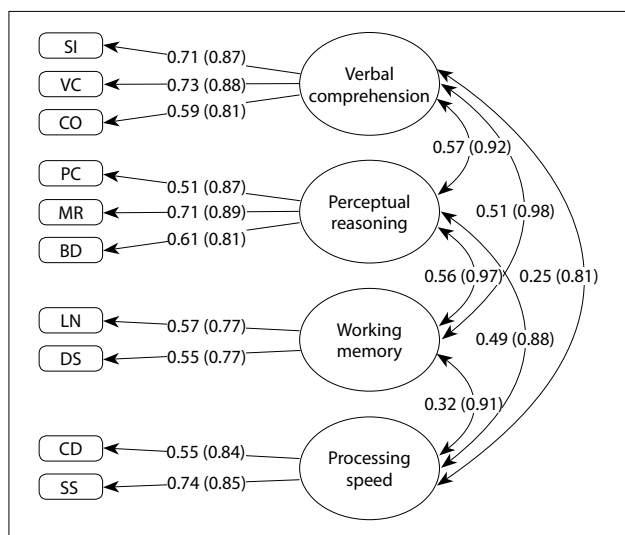
Figure 1. Factor loadings and covariance values for the correlated four-factor first-order model in non-clinical l and clinical sample. Note 1: The parenthesized values are for the clinical sample. Note 2: All values in the figure are standardized and statistically significant (p<0.001).

SI: Similarities, VC: Vocabulary, CO: Comprehension, PC: Picture Concepts, MR: Matrix reasoning, BD: Block design, LN: Letter-number sequencing, DS: Digit span, CD: Coding, SS: Symbol search
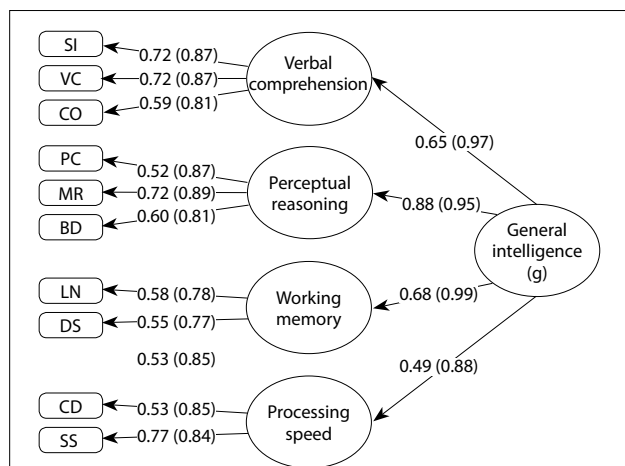

Figure 2. Factor loadings related to the second-order model in non-clinical and clinical sample. Note 1: The parenthesized values are for the clinical sample. Note 2: All values in the figure are standardized and statistically significant (p<0.001).

SI: Similarities, VC: Vocabulary, CO: Comprehension, PC: Picture Concepts, MR: Matrix reasoning, BD: Block design, LN: Letter-number sequencing, DS: Digit span, CD: Coding, SS: Symbol search

expected direction and the strength (37,43,72) and, therefore, the WISC-IV reveals an integrated structure. In addition, no significant difference was found between the two models tested in the non-clinical sample (first-order and second-order models) compared to each other. This finding reveals that both

models have equal measuring capacity for non-clinical sample. As a result, in the assessment of children in the non-clinical sample in our country, it was revealed that the WISC-IV index scores and the "g" factor representing general intelligence strongly measure the structure to be measured.

Another key point of this study is to test the factor structure of the WISC-IV in a sample of children with a clinical diagnosis. The clinical sample consists of different diagnostic groups such as ADHD, SLD and intellectual disability. There are studies abroad examining the factor structure of the WISC-IV in a single diagnostic group (46,51,58), as well as studies in which different clinical diagnosis groups (ADHD, SLD, and intellectual disability, etc.) are discussed together (16,28,31,53). Following this, in this study, as a result of the analyses conducted through a mixed clinical group revealed that both the first-order and the second-order factor structures of the WISC-IV fit perfectly with the data, similar to the findings in the non-clinical sample. Assessing the model with regard to fit indexes, these findings are supported by the findings of previous studies conducted with clinical samples (38,47,50,51). In most studies conducted with clinical groups, first-order and second-order models were tested, and both models were found to be confirmed in clinical samples (14,33,41,47). However, some studies reported that a different second-order model, the direct hierarchical (bifactor) model, is better (42,45,47). In this study, this model could not be tested due to the limitation of the lack of supplemental subtests, so as in other studies, the direct hierarchical model could not be compared with the four-factor and indirect hierarchical models. However, in this study, the first-order model (that is, the four-factor model) tested in the clinical sample was compared with the second-level order with the chi-square difference analysis, and it was revealed that the first-order model was better. In other words, it is observed that the first-order model, in which the factors are corelated and there is no general intelligence factor, generates more useful results in revealing the characteristics of clinical sample.

When the subtest factor loadings in both models in the clinical sample were analyzed, it was found that all of the WISC-IV 10 core subtests were significantly loaded on their own factors. However, as shown in the non-clinical sample, it is found that the lowest factor loadings in the first-order model are in the subtests that fall into The Working Memory and Processing Speed clusters. On the other hand, a similar situation

was found to be not valid for the clinical sample and only the subtest factor loadings of the working memory cluster were partially low at rate. These findings also indicate that the subtests in Verbal Comprehension and Perceptual Reasoning clusters have a higher predictive capacity in the clinical sample. The findings are compatible with the findings of studies on clinical samples abroad (29,50,53). In addition, the strong covariances between the WISC-IV cluster scores in the first-order model of the clinical sample and the significant loading of the four clusters on the "g" in the second-order model reveal that this scale generates consistent results within the clinical sample. All these findings indicate that the index scores obtained from the WISC-IV and FSIQ yield a valid and strong structure in the intellectual assessment of children with SLD, ADHD and intellectual disability, as in the non-clinical sample.

As a result of CFA carried out separately in clinical and non-clinical samples, the first and second order models were confirmed in both sample groups, but it was found that both the fit index values and the subtest factor loadings of the clinical sample were higher than the values in the non-clinical sample group. A similar situation was observed in the values in the covariance matrices in the first-order model, and it was revealed that the covariances between the WISC-IV cluster scores of the clinical sample group were stronger than the non-clinical sample. In the second-order model, it is found that the factor loadings of the clusters loaded on the "g" factor are higher in the clinical sample. Showing that the values obtained from the clinical sample are higher these findings reveal that the WISC-IV can provide better results in terms of clinical sample. However, this finding should not be interpreted as the structure measured by the WISC-IV in the non-clinical sample is weak and it should be considered that this may be related to the structure of models being tested. As mentioned previously, many models have been tested in studies with both clinical and non-clinical samples. For example, in previous studies, the "g" factor that defines general intelligence in direct or indirect second-order factor analyses in both clinical and non-clinical samples is discussed in two different way. In the second-order factor model, the "g" factor is directly correlated with the subtests, whereas in the indirect second-order model, the "g" factor is correlated with the subtests over clusters. Studies testing this model reported that direct second-order model were stronger in both clinical and non-clinical samples (42,45,47). In this context, since the aforementioned direct second-order model and other models were not tested in this study, the finding that the WISC-IV presented better results in clinical sample compared to non-clinical sample should be carefully considered.

Another important finding of this study is related to the results of the multiple-group factor analysis. As a result of multiple-group CFA, in addition to the analyses carried out on individual groups, the measurement invariance of the first-order model of the WISC-IV between clinical and non-clinical samples was discussed. The results was found to be partially consistent with the findings of multiple-group CFA studies (42,52,58). This reveals that the 10 core subtests of the WISC-IV measure the same theoretical structure for both samples and the validity of all first-order factors consisting of subtests. In other words, it was shown that the correlated four-factor first-order model measures a similar structure between clinical and non-clinical samples, but it was found to have a stronger structure in the clinical sample. However, it should be kept in mind that although the factor structure of the WISC-IV for clinical and non-clinical samples has similar characteristics, children in clinical samples may display different score profiles.

Finally, the fact that the first-order model in the multiple-group analysis better fits with the first-order model in the clinical sample compared to the non-clinical sample is consistent with the finding that the first-order model has a better fit than the second-order model in the clinical sample. This finding indicates, on the one hand, that both models have similar measuring capacity in the normal sample, while on the other hand, interpreting the correlated four-factor model in clinical groups such as SLD, ADHD and intellectual disability may produce more favorable results for the child being assessed. This result supports studies reporting that, in clinical samples, particularly in cases where the integrity of the general intelligence is impaired (when the differences between the index scores are very high), interpreting the general intelligence score over the index scores provides more important information about the intellectual capacity of the assessed child (52,73).

Despite its remarkable findings and inferences, this study has some limitations. It is seen that in both non-clinical and clinical samples, the first-order and indirect second-order factor structure of the WISC-IV is confirmed in the literature (30,31,41,42). However, some studies reported that apart from the first-order model, five-factor models and direct second-order models formed within the scope of CHC Theory were

also tested, and these models had better fit index values than the four-factor structure (41,44,45,49,53). In most of these models tested within the scope of CHC Theory, it is observed that the WISC-IV supplemental subtests were also used and five-factor models were formed accordingly (31,41). Within the scope of this study, since supplemental subtests could not be administered to all children, they could not be included in the analysis and therefore neither five-factor models nor direct second-order models could be tested. It is believed that future studies, in which supplemental subtests are also administered and different models formed within the scope of CHC Theory are tested, will reveal remarkable findings. In addition, based on the original factor structure of the WISC-IV, there are two observable subtests under The Working Memory and Processing Speed latent clusters. It should reminded that having fewer than three observable variables under latent factors in CFA studies may lead to obtaining impartial predictive values. In particular, in factor analysis studies including supplemental subtests, it is believed that testing models consisting of more than two factors may eliminate this possible impartial prediction.

The heterogeneous/mixed structure of the clinical sample of this study composed of different clinical diagnosis groups can be considered as another limitation. As mentioned before, studies with homogenous groups are understood to be limited. One of the most important reasons for this may be the difficulties in reaching the sufficient number of sample size to conduct CFA. In addition, researchers may have chosen heterogeneous samples that include many clinical diagnoses to demonstrate that similar structures occur in different groups within the scope of measurement invariance. However, it has been suggested that shared variance applications used in non-clinical or mixed clinical samples may mask the most important cognitive structures such as the distinction between short and long-term memory, and the use of heterogeneous/mixed clinical samples in providing structure validity of measurement tools was criticized (74). As a result, heterogeneous/mixed clinical sample was used in this study, but this sample composed of frequently encountered diagnostic groups in clinics such as SLD, ADHD and intellectual disability. In the sample, the overall intelligence score ranges of children in the intellectual disability diagnosis group are observed to be wide (28-69). However, they were diagnosed with mild and moderate intellectual disability through assessment of general

intelligence score together with four indices and subtest scores. In this respect, it is considered that it would be more appropriate to interpret the findings of the study on groups with SLD, ADHD and mental retardation, rather than generalizing them to all clinical groups.

Contribution Categories		Author Initials
Category 1	Concept/Design	C.C., I.Y., M.G.Y., G.E.
	Data acquisition	G.E., C.C., I.Y.
	Data analysis/Interpretation	I.Y., M.G.Y., C.C.
Category 2	Drafting manuscript	C.C., I.Y., M.G.Y.
	Critical revision of manuscript	G.E., C.C.
Category 3	Final approval and accountability	C.C., I.Y., M.G.Y., G.E.
Other	Technical or material support	C.C., G.E., M.G.Y.
	Supervision	G.E.

Ethics Committee Approval: Ethics committee approval of the study was obtained from Ankara University Ethics Committee Department (Date: 16/02/2017, Number: 49).

Informed Consent: Written informed consent was obtained from the patients.

Peer-review: Externally peer-reviewed.

Conflict of Interest: No declared conflict of interest was reported by the author.

Financial Disclosure: No declared financial support was reported by the author.

REFERENCES

1. Canivez G. Psychometric versus actuarial interpretation of intelligence and related aptitude batteries: In Saklofske DH, Schwean VL, Reynolds CR (editors). The Oxford handbook of child psychological assessment. New York, NY: Oxford University Press, 2013, 84-112.

2. Spruill J, Harrison PL. Assessment of mental retardation/intellectual disability with the WISC-IV: In Prifitera A, Saklofske DH, Weiss LG (editors). WISC-IV clinical assessment and intervention. San Diego, CA: Academic Press, 2008, 273-298.

3. Çelik C. Evaluation of cognitive functions in specific learning disability and effectiveness of intervention methods. PhD Thesis, Ankara University Institute of Social Sciences, Ankara, 2019.

4. Hale J, Alfonso V, Berninger V, Bracken B, Christo C, Clark E, et al. Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. Learn Disabil Q 2010; 33:223-236.

5. Ortiz SO. CHC Theory of intelligence. In Goldstein S, Princiotta D, Naglieri JA, editors. Handbook of intelligence. New York, NY: Springer, 2015, 209-227.

6. Braden JP, Kratochwill TR. Treatment utility of assessment: Myths and realities. School Psych Rev Taylor & Francis 1997; 26:475-485.

7. Schwean VL, McCrimmon A. Attention-deficit/hyperactivity disorder: Using the WISC-IV to inform intervention planning: In Prifitera A, Saklofske DH, Weiss LG (editors). WISC-IV clinical assessment and intervention. San Diego, CA: Academic Press, 2008, 193-215.

8. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). Standards for educational and psychological testing. Amer Educational Research Assn, 1999.

9. Goodwin LD, Leech NL. The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. Meas Eval Couns Dev 2003; 36:181-191.

10. Linn RL. The standards for educational and psychological testing: Guidance in test development: In Lane S, Raymond MR, Haladyna TM (editors). Handbook of test development. New York, NY: Routledge, 2011, 41-52.

11. Savasir I, Sahin N. Wechsler Intelligence scale for children (WISC-R) manual. Ankara: Turkish Psychological Association Publications, 1995.

12. Uluc S, Oktem F, Erden G, Gencoz T, Sezgin N. Wechsler Intelligence Scale for Children-IV: A new era for Turkey in evaluation of intelligence in the clinical context. Turkish Psychological Articles 2011; 14: 49-57. (Turkish)

13. Kane M. Content-related validity evidence in test development: In Haladyna TM, Downing SM (editors). Handbook of test development. New Jersey: Lawrence Erlbaum Associates, Inc, 2006, 131-153.

14. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol 1995; 50:741-749.

15. Keith TZ, Reynolds MR. Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests: In Flanagan DP, McDonough EM (editors). Contemporary intellectual assessment: Theories, tests, and issues. The Guilford Press, 2018, 853-900.

16. Devena SE, Gay CE, Watkins MW. Confirmatory factor analysis of the WISC-IV in a hospital referral sample. J Psychoeduc Assess 2013; 31:591-599.

17. O'Brien K. Factor analysis: An overview in the field of measurement. Physiother Canada 2007; 59:142-155.

18. Anderson T, Wallace ED Jr. Confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Revised with normal and psychiatric adolescents. J Res Adolesc 1995; 5:319-332.

19. Kush JC, Watkins MW, Ward TJ, Ward SB, Canivez GL, Worrell FC. Construct validity of the WISC-III for White and Black students from the WISC-III standardization sample and for Black students referred for psychological evaluation. School Psych Rev 2001; 30:70-88.

20. Lee LMP, Lam YR. Confirmatory factor analyses of the Wechsler Intelligence Scale for Children-Revised and the Hong Kong-Wechsler Intelligence Scale for Children. Educ Psychol Meas 1988; 48:895-903.

21. Naglieri JA. Factor structure of the WISC-R for children identified as learning disabled. Psychol Rep 1981; 49:891-895.

22. Ramanaiah NV, O'Donnell JP, Ribich F. Multiple-group factor analysis of the Wechsler Intelligence Scale for Children. J Clin Psychol Wiley Online Library 1976; 32:829-831.

23. Silverstein AB. Alternative multiple-group solutions for the WISC and the WISC-R. J Clin Psychol 1982; 38:166-168.

24. Tingstrom DH, Pfeiffer SI. WISC-R factor structure in a referred pediatric population. J Clin Psychol 1988; 44:799-802.

25. Bodin D, Pardini DA, Burns TG, Stevens AB. Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. Child Neuropsychol 2009; 15:417-424.

26. 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; 49:677-682.

27. MacLean H, McKenzie K, Kidd G, Murray AL, Schwannauer M. Measurement invariance in the assessment of people with an intellectual disability. Res Dev Disabil 2011; 32:1081-1085.

28. Richerson LP, Watkins MW, Beaujean AA. Longitudinal invariance of the Wechsler Intelligence Scale for Children–Fourth Edition in a referral sample. J Psychoeduc Assess 2014; 32:597-609.

29. Watkins MW. Structure of the Wechsler Intelligence Scale for Children--Fourth Edition among a national sample of referred students. Psychol Assess 2010; 22:782-787.

30. Watkins MW, Wilson SM, Kotz KM, Carbone MC, Babula T. Factor structure of the Wechsler Intelligence Scale for Children–Fourth Edition among referred students. Educ Psychol Meas 2006; 66:975-983.

31. 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; 31:114-131.

32. Wechsler D. Wechsler intelligence scale for children–Fifth Edition (WISC-V). San Antonio, TX: Psychological Corporation, 2014.

33. Kush JC, Canivez GL. Construct validity of the WISC–IV Italian edition: A bifactor examination of the standardization sample: Chi niente sa, di niente dubita. Int J Sch Educ Psychol 2019;1-15.

34. Oktem F, Erden G, Gencoz T, Sezgin N, Uluc S. Wechsler Intelligence Scale for Children-IV (WÇZÖ-IV) application and scoring manual Turkish version. Ankara: Turkish Psychologists Association Publications-Pearson Assessments, 2016.

35. Sotelo-Dynega M, Dixon SG. Cognitive assessment practices: A survey of school psychologists. Psychol Sch Wiley Online Library 2014; 51:1031-1045.

36. Nakano S, Watkins MW. Factor structure of the wechsler intelligence scales for children–fourth edition among referred native american students. Psychol Sch 2013; 50:957-968.

37. Styck KM, Watkins MW. Structural validity of the WISC-IV for students with learning disabilities. J Learn Disabil 2016; 49:216-224.

38. Watkins MW, Canivez GL, James T, James K, Good R. Construct validity of the WISC–IVUK with a large referred Irish sample. Int J Sch Educ Psychol 2013; 1:102-111.

39. Wechsler D. Wechsler intelligence scale for children–Fourth Edition (WISC-IV). San Antonio, TX: Psychological Corporation, 2003.

40. Chen H-Y, Keith TZ, Yung-Hwa C, Ben-Sheng C. What does the WISC-IV measure? Validation of the scoring and CHC-based interpretative approaches. Jiaoyu Kexue Yanjiu Qikan 2009; 54:85-108.

41. Keith TZ, Fine JG, Taub GE, Reynolds MR, Kranzler JH. Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure. School Psych Rev 2006; 35:108-127.

42. Chen H, Zhu J. Measurement invariance of WISC-IV across normative and clinical samples. Pers Individ Dif 2012; 52:161-166.

43. Lecerf T, Rossier J, Favez N, Reverte I, Coleaux L. The four-vs. alternative six-factor structure of the French WISC-IV. Swiss J Psychol 2010; 69:221-232.

44. Golay P, Reverte I, Rossier J, Favez N, Lecerf T. Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. Psychol Assess 2013; 25:496-508.

45. 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.

46. 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; 7:2-14.

47. 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.

48. Styck KM, Watkins MW. Structural validity of the WISC-IV for students with ADHD. J Atten Disord 2017; 21:921-928.

49. Thaler NS, Barchard KA, Parke E, Jones WP, Etcoff LM, Allen DN. Factor Structure of the Wechsler Intelligence Scale for Children: Fourth Edition in Children with ADHD. J Atten Disord 2015; 19:1013-1021.

50. Yang P, Cheng CP, Chang CL, Liu TL, Hsu HY, Yen CF. Wechsler Intelligence Scale for Children 4th edition-Chinese version index scores in Taiwanese children with attention-deficit/hyperactivity disorder. Psychiatry Clin Neurosci 2013; 67:83-91.

51. 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; 87:383-407.

52. Giofrè D, Cornoldi C. The structure of intelligence in children with specific learning disabilities is different as compared to typically development children. Intelligence 2015; 52:36-43.

53. Canivez GL. Construct validity of the WISC-IV with a referred sample: direct versus indirect hierarchical structures. Sch Psychol Q 2014; 29:38-51.

54. Millsap RE, Kwok OM. Evaluating the impact of partial factorial invariance on selection in two populations. Psychol Methods 2004; 9:93-115.

55. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organ Res 2000; 3:4-70.

56. Chen FF, Sousa KH, West SG. Teacher's corner: Testing measurement invariance of second-order factor models. Struct Equ Model 2005; 12:471-492.

57. Steinmetz H, Schmidt P, Tina-Booh A, Wieczorek S, Schwartz SH. Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. Qual Quant 2009; 43:599-616.

58. Ogata K. WISC-IV factor structures of Japanese children with borderline, or deficient intellectual abilities: testing measurement invariance compared to simulated norm. Psychology 2019; 10:767-776.

59. Gomez R, Vance A, Watson S. Bifactor model of WISC-IV: Applicability and measurement invariance in low and normal IQ groups. Psychol Assess 2017; 29:902-912.

60. Chen H, Zhu J. Factor invariance between genders of the Wechsler Intelligence Scale for Children–Fourth edition. Pers Individ Dif 2008; 45:260-266.

61. Celik C, Yigit I, Erden G. Confirmatory factor analysis of the revised form of the Wechsler Intelligence Scale for Children: A sample of children with normal mental development. Turkish Psychological Articles 2015; 18:21-29. (Turkish)

62. Celik C, Yigit I, Erden G, Vural MA. The evaluation of the factor structure of Wechsler Intelligence Scale for Children-Revised Form in children with intellectual disability. Turk J Child Adolesc Ment Health 2016; 23:17-28. (Turkish)

63. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model a Multidiscip J 1999; 6:1-55.

64. Kelloway EK. Using LISREL for structural equation modeling: A researcher's guide. California: Sage, 1998.

65. Kline RB. Principles and practice of structural equation modeling. New York, NY: The Guilford Press, 2015.

66. Schwarz G. Estimating the dimension of a model. Ann Stat 1978; 6:461-464.

67. Brown TA. Confirmatory factor analysis for applied research. Second ed., New York, NY: The Guilford Press, 2015.

68. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Model J 2007; 14:464-504.

69. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model J 2002; 9:233-255.

70. Williams PE, Weiss LG, Rolfhus EL. WISC-IV technical report# 2: Psychometric properties. https://images.pearsonclinical.com/images/pdf/wisciv/WISCIVTechReport2.pdf.

71. 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; 7:15-28.

72. 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; 6:239-251.

73. Flanagan DP, Kaufman S. Essentials of assessment with WISC-IV. New York, NY: Wiley, 2004.

74. Delis DC, Jacobson M, Bondi MW, Hamilton JM, Salmon DP. The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: lessons from memory assessment. J Int Neuropsychol Soc 2003; 9:936-946.