Journal of Educational, Health and Community Psychology Vol 12, No 4, 2023 E-ISSN 2460-8467

Locating the Invariant Faxtor's Cognitive Ability Test

Aryo Bimo Adjie
Padjadjaran University
Indonesia
aryo18001@mail.unpad.ac.id
https://orcid.org/0000-0003-2930-3382

Fatiya Nurhadi Nasyiah
Indonesia University of Education
Indonesia
fatiyanurhadi@upi.edu

Abstract

Measurement invariance with respect to groups is an essential aspect of the fair use of scores of intelligence tests and other psychological measurements. It is widely believed that equal factor loadings are sufficient to establish measurement invariance in confirmatory factor analysis. This study investigated the invariance of the Faxtor Cognitive Ability Test (FCAT) between genders and age groups. A second-order 4-factor model was tested on a nationally-representative sample of 3,850 aged 11 to 53 years. The results demonstrated full strict invariance between genders and configural invariance between age groups. The FCAT subtests demonstrate the same underlying theoretical latent constructs, the same strength of relationships among factors and subtests, the same validity of each first-order factor, and the same communalities, regardless of the gender, thus supporting the same interpretive approach and meaningful comparisons of the FCAT between male and female. The findings also showed variations across age groups, non-invariance, and evidence that age influences the latent variable differences score on the FCAT.

Keywords: *Measurement Invariance, Gender, Age, Faxtor Cognitive Ability Test*

Received 10 September 2023/Accepted 27 November 2023

Introduction

One issue that emerges in measurement is whether the results of a measurement instrument administered to different groups truly measure the same attribute (Horn & McArdle, 1992) and whether different individual characteristics respond similarly (Vandenberg & Lance, 2000). For instance, this study focuses on cognitive measurement, whether cognitive measurement functions equally at different ages or in different genders. In cognitive measurement, the measurement tool should truly measure the latent ability without any other factor that affects groups of test takers differently as this relates to the fairness and validity of the test (Wicherts, 2016; Wicherts & Dolan, 2010). A measurement tool must be able to measure the same thing in different people (Blankson & McArdle, 2015). This is often known as invariance.

The concept of invariance sounds similar to the concept of bias because both are related to the fairness of measurement in each group, but they are quite different. Bias leads to the presence of undesirable factors (it can be constructing bias, method bias, instrument bias, administration bias, or item bias) that impact the difference between estimated parameters and true parameters (Van de Vijver & Poortinga, 2005). Meanwhile, measurement invariance leads to scores that are equally comparable between groups by examining whether each item (on the measured construct) functions equally in each group by analyzing its psychometric properties, specifically based on factor loadings and intercepts (Meredith, 1993).

Horn & McArdle (1992) explained that measurement invariance is evidence that while an instrument is administered to different groups of individuals, it measures the same thing and functions equally. Therefore, the interpretation of measurement results based on differences in characteristics or age becomes clear and precise (Blankson & McArdle, 2013; Horn & McArdle, 1992). Without this measurement invariance, differences in scores between individuals cannot be simply interpreted. Measurement invariance is obtained by comparing measurements between the groups to examine whether the items on the measuring instrument contribute equally to the latent construct being measured (Meredith, 1993).

The instrument being explained in this study is the Faxtor Cognitive Ability Test (FCAT) (Yudiana & Putra, 2022). FCAT is one of the cognitive measurement tools in Indonesia developed based on Cattel-Horn-Carrol Theory (CHC). This theory explains cognitive abilities with the term General Cognitive Abilities (GCA) which consists of broad and narrow abilities that are more specific (Schneider & McGrew, 2012). FCAT measured GCA through 4 broad abilities known as Knowledge Comprehension (Gc), Fluid Reasoning (Gf), Visual Processing (Gv), and Processing Speed (Gs). GCA measurement using FCAT is reliable and has evidence of construct validity, including based on the internal structure with the confirmatory factor analysis (CFA) method, intercorrelation method between subtest scores, broad, and GCA score, also correlation with other variables (Yudiana & Putra, 2022). Thus, it can be ascertained that measurement through the subtest on the FCAT represents GCA within the CHC theoretical framework. GCA measured by FCAT is based on age norms of 9 age groups (11 – 53 years).

FCAT norms are developed based on the mean of each age groups, instead of based on the factor loading of each broad ability. This drives the need for measurement invariance testing to ensure that the cognitive measurement parameters on FCAT are invariant in each group so that the interpretation of scores becomes equivalent for each group (Wicherts, 2016). Therefore, this study tested measurement invariance in each age groups and gender in the measurement of GCA using FCAT.

Four levels of invariance measurement are conducted in this study, namely configural, metric, scalar, and strict invariance. Each level has restrictions that must be accomplished (Byrne & Stewart, 2006; Chen et al., 2005; Meredith, 1993). These constraints (explained below) relate to factor structure, factor loadings, intercepts, and measurement errors (Putnick & Bornstein, 2016). These four levels of measurement invariance are analyzed using multigroup confirmatory factor analytic (MCFA) as a widely used method (Hertzog & Schaie, 1986; Milfont & Fischer, 2010).

Method

Participants

A total of 3,850 participants were obtained from almost all-geographical areas in Indonesia during research from 2019 to 2021. Specifically, the majority of the participants (n=3248, 84%) were from Java, followed by participants from Sumatera (n=396, 10%), Sulawesi (n=84, 2%), Bali-Nusa Tenggara (n=58, 2%), Kalimantan (n=33, 1%), and Maluku-Papua (n=31, 1%). As many as 2153 participants (56%) were female, while the rest 1697 (44%) were male. Before commencing the study II, informed consent was gained from all the participants, either autonomously or through their legal representatives. The participants ranged from 11 to 58 years old (M=19.14, SD=1.155) and divided into 9 age groups. Participants had a variety of educational backgrounds, ranging from participants at the elementary to postgraduate level. The data were obtained through online psychological assessment of Faxtor Indonesia.

Table 1
Participants Characteristics

Participants		N (%)
Gender	Male	1.697 (44%)
	Female	2.153 (56%)
Age (years)	Group 1 (11 - 12 years)	400 (10%)
	Group 2 (13 – 14 years)	300 (8%)
	Group 3 (15 – 16 years)	500 (13%)
	Group 4 (17 – 20 years)	600 (16%)
	Group 5 (21 – 24 years)	600 (16%)
	Group 6 (25 – 28 years)	350 (9%)
	Group 7 (29 – 33 years)	600 (16%)
	Group 8 (34 – 39 years)	300 (8%)
	Group 9 (≥ 40 years)	200 (5%)
Education Levels	Elementary School	698 (18%)
	Middle High School	624 (16%)
	High School	1.134 (29%)
	Diploma	153 (4%)
	Bachelor	1.166 (30%)
	Master	75 (2%)
Geographical areas	Jawa	3.248 (84%)
	Bali & Nusa Tenggara	58 (2%)
	Kalimantan	33 (1%)
	Papua	31 (1%)
	Sulawesi	84 (2%)
	Sumatera	396 (10%)

Instrumentation

Faxtor's Cognitive Ability Test (FCAT). The FCAT is a computer-based assessment designed to measure an individual's cognitive abilities based on Cattell-Horn-Carroll (CHC) theory (Schneider & McGrew, 2012). The test measures four broad cognitive abilities — Fluid Reasoning (Gf), an ability to analyze through examination of information pattern in solving problems; General Crystalized (Gc), an ability to comprehend, reason, and process verbal information and a description of one's verbal knowledge; General Visualization (Gv), an ability to observe incisively, choose relevant information, organize and examine information from multiple point of views to solve problems; and General Speed (Gs), an ability to control attention quickly and accurately, and conduct simple cognitive tasks skillfully (Carroll, 1993; Kyllonen & Kell, 2017; Schneider & McGrew, 2012). All the subtests are administered using multiple choice format ranging from four to ten options. Details about subtests are explained in table 2.

Table 2
Test specification of FCAT

Subtest	Broad/Narrow Ability	Number of Item	Test Specification
Non-Verbal Reasoning (NVR)	Gf/Induction	30	The examinee is presented with a certain figural pattern of related stimuli and required to select one figure of several stimuli that would complete or continue the pattern.
Number Sequence (NS)	Gf/Quantitative reasoning	20	The examinee is presented with an incomplete series of related numbers and required to select those that best complete the series.
Verbal Logic (VL)	Gc/Lexical knowledge	30	The examinee is presented with a word and required to select the best definition that related to them.
Lexical Knowledge (LXK)	Gc/Lexical knowledge	24	The examinee is presented with a word and required to select the best word which is a synonym or antonym.
Associative Memory (ASM)	Gc/Language Development	30	The examinee is presented with a word and required to select the best additional word which is the synonym, compound, or metaphor word that relate to them.
Visualization (VS)	Gv/Visualization	25	The examinee is presented with a two-dimensional figure with a cut shape and required to select one that does not fit the full shape or standing model.
Spatial Reasoning (SPT)	Gv/Visualization	20	The examinee is presented with a cube and required to select one cube net to form the standing model.
Coding Test (CDT)	Gs/Perceptual speed-search	60	The examinee is presented with a figure and is required to rapidly view rows of stimuli that are similar in each row within a specified time limit. The task is to find one stimuli which is identical with the figure shown.

Subtest	Broad/Narrow Ability	Number of Item	Test Specification
Perceptual Speed Search (PSS)	Gs/Perceptual speed-compare	100	The examinee is presented with two patterns and required to rapidly view rows of stimuli that are similar in each row within a specified time limit. The task is to find one stimuli which is identical with either one or two patterns shown.

Analysis

Four levels of nested models were tested to investigate the degree of invariance in FCAT. The initial and weakest level was configural invariance. It assumed the same number of factors and the same overall factor pattern across groups. The second level was first-order factor loading invariance, also called metric (or weak factorial) invariance. Loadings of subtests on factors were constrained so that factor loadings were equal across groups. When the factor loadings are equal, scales of latent variables are the same for both groups and the unit of measurement is identical. That is, for each unit change in latent variable, scores on subtests change by the same amount in both groups. The third level was intercept invariance, also known as scalar (strong factorial) invariance. In this level of invariance, any group differences in subtest means are a result of true mean differences in latent factors. Subtests have the same intercepts across groups given the same latent means for an underlying factor. To examine whether "all group differences on the measured variables are captured by, and attributable to, group differences on the common factors" (Widaman & Reise, 1997), we tested invariance of residuals, also called strict factorial invariance. These residuals are a combination of subtest-specific unique variance and random measurement errors. The fifth level was second-order factor loading invariance. This level assumed first-order latent factors show the same amount of change in each group for the same amount of increase in GCA.

Multiple insides of model fit were used to evaluate and compare the various models in this study (Bentler & Bonett, 1980; Hu & Bentler, 1998, 1999; Kline, 2005; Marsh et al., 1988). Single models were evaluated using the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). An RMSEA less than 0.05 corresponded to a good fit, and 0.08 was considered an acceptable fit (McDonald & Ho, 2002). For completeness, we included the 90% confidence interval for RMSEA. SRMR values less than 0.08 were considered acceptable (Hu & Bentler, 1999). A value of 0.95 served as the cutoff for acceptable fit on all indices ranging from zero to 1, with 1 indicating a perfect fit (Bentler & Bonett, 1980; Li-tze Hu & Bentler, 1998; Li-tze Hu & Bentler, 1999; Kline, 2005; Marsh et al., 1988). Change in chi-square ($\Delta\chi 2$) was used to evaluate competing, nested models (Bentler & Bonett, 1980). The Akaike Information Criterion (AIC) and sample size adjusted Bayesian Information Criterion (aBIC) were also used for comparisons of non nested models (Boomsma, 2003; Loehlin, 2004), with smaller values indicating a better fit. Comparatively, aBIC has a greater reward for parsimony than does the AIC.

To determine evidence of invariance, since there is little consensus concerning the most appropriate criterion (Byrne & Stewart, 2006), two perspectives were evaluated for invariance analyses: (a) the traditional perspective based on $\Delta\chi 2$ and (b) the practical perspective based on differences in comparative fix index (ΔCFI). When evaluating the traditional perspectives, given the large sample and the number of comparisons being made, we used a more-strict definition of statistical significance of $\Delta\chi 2$ ($p < .001$). Comparatively, the $\Delta\chi^2$ test is known to be sensitive to sample size and moderate discrepancies from normality (Chen et al., 2005; Kline, 2005; West et al., 1995). Therefore, Cheung and Rensvold (2002) recommended ΔCFI as superior to $\Delta\chi^2$ for tests of invariance because it is independent of both model complexity and sample size and because it is not correlated with the overall fit measures. “A value of ΔCFI smaller than or equal to –.01 indicates that the null hypothesis of invariance should not be rejected” (Cheung & Rensvold, 2002). An absolute ΔCFI value higher than .01 (i.e., |ΔCFI| N .01) was proposed as an indicator of a meaningful fall in fit. Given the large sample sizes, large modeled variables, and the number of comparisons being made in this study, we decided to evaluate the invariance by $\Delta\chi 2$ and ΔCFI jointly to secure meaningfulness and prevent any unnecessary oversensitivity. The criterion for rejecting the null hypothesis of invariance was set as a p value of less than .001 for the $\Delta\chi 2$ test and an absolute ΔCFI value higher than .01. Analyses were conducted using R studio program with lavaan packages (Team, 2022; Rosseel, 2012).

Result

Study 1: Age groups

Table 1 lists all steps in the invariance analyses. The baseline model (figure 1) fit was first checked for each sample. The model fits each datum well, suggesting that the following invariance verification was meaningful. Variance–covariance matrices were constrained to be equal across groups (Model 1). This constrained model fits the data well (CFI = 0.984; RMSEA = 0.041), suggesting fairly invariant FCAT subtest covariance patterns in general. Because equality of variance–covariance matrices between age groups is supported, the FCAT generally measures the same constructs between ages. Because any factor structure is derived from these variance–covariance matrices, this result revealed that the FCAT factor structure in every age groups should be similar.

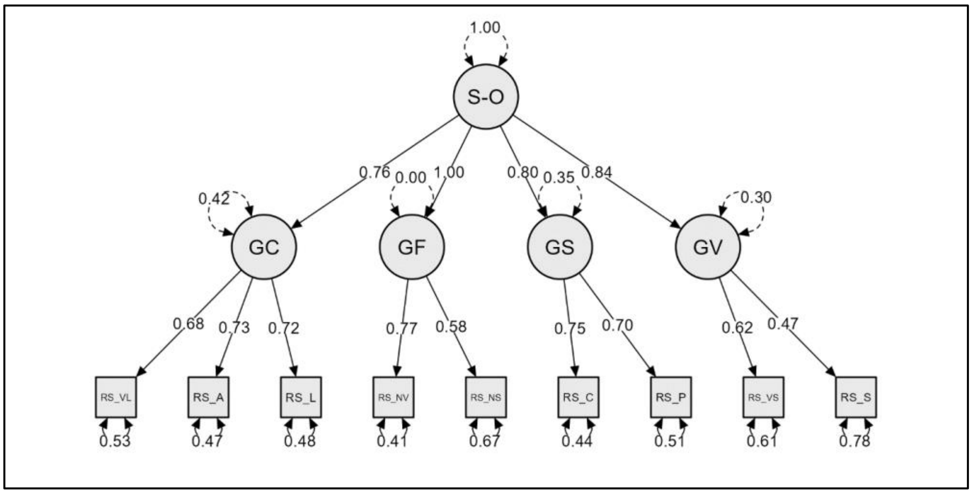

Figure 1. Baseline model of FCAT (Model 1 in Table x and x)

When testing nested models, first, the configural model (Model 2) provided an acceptable fit to the data. Every age group shared the same FCAT first- and second-order five-factor patterns and the corresponding subtests loaded on the same factors. With the factor pattern established, we imposed cross-group constraints on the first-order factor loadings (Model 3). There was deterioration of fit with these constraints by both $\Delta\chi2$ and ΔCFI, implying that the subtests were not measure the same latent factors in both groups. Next, we constrained the subtest intercepts to be equal (Model 4). To identify this model properly, we fixed the means of the first-order factors in every group to zero. Thus, the factor means for every group represent the mean differences. The addition of subtest intercepts constraints reduced the fit according to $\Delta\chi2$ and ΔCFI. The ΔCFI value was > 0.01, implying that the subtest intercepts are not the same in age groups. Next, when the subtest residuals were constrained to be equal across groups (Model 5), there was deterioration of fit with these constraints.

Table 3
Multi-sample goodness of fit based on age groups

Model	χ^2	df	CFI	RMSEA	RMSEA 90% CI	SRMR	AIC	aBIC	Model	ΔCFI	$\Delta\chi^2$	Δdf	p
Phase I : Baseline model fit for each group													
Group 1 (n=400)	48.85	23	0.941	0.053	0.032 - 0.074	0.044	19617.97	19705.78					
Group 2 (n=300)	18.54	23	1.000	0.000	0.000 - 0.000	0.027	15459.33	15540.82					
Group 3 (n=500)	44.60	23	0.973	0.043	0.024 - 0.062	0.039	26142.09	26234.81					
Group 4 (n=600)	39.08	23	0.983	0.034	0.014 - 0.052	0.031	30681.38	30778.11					
Group 5 (n=600)	22.47	23	0.985	0.042	0.016 - 0.066	0.028	21916.67	21991.42					
Group 6 (n=350)	19.00	23	1.000	0.000	0.000 - 0.000	0.028	17926.31	18011.18					
Group 7 (n=700)	65.30	23	0.957	0.055	0.040 - 0.071	0.037	30767.36	30864.09					
Group 8 (n=300)	50.25	23	0.942	0.063	0.039 - 0.087	0.042	15507.97	15589.45					
Group 9 (n=200)	29.45	23	0.987	0.037	0.000 - 0.073	0.035	10226.72	10299.29					
Phase II : Measurement Invariance across groups													
Model 1 Baseline model	279.15	23	0.974	0.041	0.035 - 0.047	0.029	200296.60	200434.27	-	-	-	-	-
Model 2 Configural Invariance	346.22	207	0.978	0.040	0.033 - 0.047	0.031	197111.70	198857.00	-	-	-	-	-
Model 3 Metric Invariance	601.19	271	0.948	0.053	0.032 - 0.074	0.059	197238.60	198583.60	2 vs 3	0.030	254.98	64	0.000
Model 4 Scalar Invariance	807.05	303	0.921	0.062	0.036 - 0.088	0.066	197380.50	198525.30	3 vs 4	0.027	205.86	32	0.000
Model 5 Strict factorial invariance	1362.61	375	0.845	0.078	0.071 - 0.085	0.098	197792.10	198486.50	4 vs 5	0.076	555.56	72	0.000

Study 2: Gender

When testing nested models, first, the configural model (Model 2) provided an acceptable fit to the data. Males and females shared the same FCAT first- and second-order five-factor patterns and the corresponding subtests loaded on the same factors as shown in table 4. With the factor pattern established, we imposed cross-group constraints on the first-order factor loadings (Model 3). There was no deterioration of fit with these constraints by both $\Delta\chi2$ and ΔCFI, implying that the subtests measure the same latent factors in both groups. Next, we constrained the subtest intercepts to be equal (Model 4). To identify this model properly, we fixed the means of the first-order factors in the male group to zero, but freed those in the female group. Thus, the factor means for the female group represent the mean differences. The addition of subtest intercepts constraints not reduced the fit according to $\Delta\chi2$ and ΔCFI. The ΔCFI value was 0, implying that the subtest intercepts are the same in both groups. Next, when the subtest residuals were constrained to be equal across groups (Model 5), there was no deterioration of fit with these constraints.

Most importantly, these estimates were found invariant between genders. Table 5 lists factor loadings for each of the FCAT subtests. For all age groups and genders, the subtests with top Gc, Gf, Gv, and Gs loadings are: ASM, NVR, VS and CDT. Then for all age groups and genders, the broad abilities General Fluid (Gf) has consistently with top GCA loadings.

Table 4

Multi-sample goodness of fit based on genders

Model	χ^2	df	CFI	RMSEA	RMSEA 90% CI	SRMR	AIC	aBIC	Model	ΔCFI	$\Delta\chi^2$	Δdf	p
Phase I : Baseline model fit for each group													
Male (n = 2153)	117.83	23	0.981	0.044	0.036 - 0.052	0.026	112103.35	112228.19					
Female (n = 1697)	53.243	23	0.993	0.030	0.020 - 0.040	0.017	87972.87	88103.34					
Phase II : Measurement Invariance across groups													
Model 1 Baseline model	279.15	23	0.974	0.041	0.035 - 0.047	0.029	200296.60	200434.27	-	-	-	-	-
Model 2 Configural Invariance	208.63	46	0.983	0.043	0.039 - 0.047	0.023	200145.80	200533.60	-	-	-	-	-
Model 3 Metric Invariance	228.45	54	0.982	0.041	0.034 - 0.048	0.029	200149.60	200487.40	2 vs 3	0.001	19.826	8	0.000
Model 4 Scalar Invariance	272.39	58	0.977	0.044	0.036 - 0.052	0.031	200185.50	200498.30	3 vs 4	0.004	43.940	4	0.000
Model 5 Strict factorial invariance	294.50	67	0.976	0.042	0.035 - 0.047	0.033	200189.70	200446.10	4 vs 5	0.001	22.109	9	0.000

Table 5

Factor loadings FCAT's subtest and broad abilities

Observe Variables	Latent Variables	Factor Loadings											Mean
		Genders		Age groups									
		Male	Female	1	2	3	4	5	6	7	8	9	
VL	Gc	0.67	0.70	0.54	0.48	0.58	0.69	0.67	0.60	0.61	0.66	0.77	0.63
ASM	Gc	0.73	0.72	0.70	0.67	0.73	0.56	0.70	0.69	0.68	0.62	0.67	0.68
LXK	Gc	0.72	0.74	0.53	0.58	0.58	0.61	0.69	0.70	0.61	0.68	0.66	0.65
NVR	Gf	0.78	0.76	0.66	0.64	0.69	0.70	0.66	0.78	0.66	0.69	0.77	0.71
NS	Gf	0.56	0.59	0.40	0.49	0.52	0.52	0.48	0.51	0.50	0.49	0.54	0.51
VS	Gv	0.61	0.63	0.81	0.60	0.75	0.68	0.67	0.87	0.61	0.53	0.92	0.70
SPT	Gv	0.48	0.46	0.53	0.68	0.70	0.70	0.67	0.50	0.59	0.60	0.53	0.59
CDT	Gs	0.74	0.76	0.62	0.66	0.68	0.53	0.52	0.58	0.66	0.51	0.55	0.62
PSS	Gs	0.69	0.71	0.29	0.48	0.31	0.52	0.55	0.53	0.45	0.53	0.50	0.51
Gc	GCA	0.71	0.83	0.61	0.70	0.63	0.75	0.74	0.74	0.71	0.77	0.87	0.73
Gf	GCA	1.00	1.00	1.00	0.95	0.93	1.00	1.00	1.00	1.00	0.95	1.00	0.98
Gv	GCA	0.86	0.83	0.49	0.88	0.73	0.73	0.85	0.74	0.94	0.87	0.80	0.79
Gs	GCA	0.83	0.80	0.88	0.77	0.78	0.83	0.87	0.90	0.81	0.78	0.93	0.83

Discussion

Measurement invariance provides an indication of the extent to which we can say that we are measuring the same thing across different groups (Blankson & McArdle, 2015). Study 1 and Study 2 aims to examine the measurement of invariance FCAT based on age groups and genders. The result of study 1 show that the configural model across age groups may be more plausible than the stronger forms of invariance (metric, scalar, and strict) based on the more stringent ΔCFI tests. Hence, when conducting analyses across age groups, researchers might not be loss of invariance information. These results also indicate that comparisons across age groups cannot be made of the variances and covariances among the latent variables, in consequences any such comparisons of the latent means or observed means, covariances, and variances should be done cautiously (Bontempo et al., 2012; Gregorich, 2006; McArdle et al., 2007; Widaman et al., 2010; Widaman & Reise, 1997). More result indicates that each age groups had a different factor loading based on AIC comparison with configural model has a better fit than another models.

With regard to gender invariance, the results differ for study 2. The result show that the FCAT fit for strictness of the test based on ΔCFI tests. FCAT models fit across genders when the mean, intercepts, and residual were constrained. This result indicates that FCAT's subtests generally demonstrate the same underlying theoretical latent constructs across genders the, the same strength of relationships among factors and subtests, the same validity of each first-order factor, and the same communalities. Invariant results provide evidence that FCAT index scores and subtests have the same meaning for both genders, FCAT results for males and females can be interpreted in the same way, and that meaningful comparisons between genders can be made.

Major findings are show in table 5 that the Fluid Reasoning (Gf) factor had a standardized loading of 0.98 on the second-order GCA factor. In the literature, there are considerable reports suggesting that fluid reasoning factors often show GCA loadings approaching or even reaching unity (Bickley et al., 1995; Gustafsson, 1984; Keith et al., 2006). Once again, fluid reasoning is demonstrated to be the cornerstone of human cognition. Among all FCAT's subtests, NVR had the highest average factor loading (0.71), followed by Visualization (0.70), and Associative Memory (0.68).

We recommend that future validity evidence be accumulated continuously. Invariant meaning for FCAT in other subpopulations (e.g., clinical groups, or cultures) be explored, and studies based on clinical performance or diagnostic differentiation be conducted to provide more evidence of validity and increase our understanding of how the FCAT functions in various relevant groups.

Conclusion

Because of the complexity of the model and the strictness of the test, we concluded that the FCAT exhibits acceptable levels of invariance among four factors between the male and female groups. Differences in subtest scores on the FCAT are generally caused by latent constructs, and the score test is not be influenced by the gender status. The findings showed variations across age groups, non-invariance, and evidence that age influences the latent variable differences score on the FCAT.

Acknowledgment

The study was funded and supported by Faxtor Indonesia in 2023 as a research project.

Conflict of Interest

All authors declare that they have no conflicts of interest.

References

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. doi: 10.1037/0033-2909.88.3.588

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence*, *20*(3), 309–328. doi: 10.1016/0160-2896(95)90013-6

Blankson, A. N., & McArdle, J. J. (2015a). Measurement invariance of cognitive abilities across ethnicity, gender, and time among older americans. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *70*(3), 386–397. doi: 10.1093/geronb/gbt106

Blankson, A. N., & McArdle, J. J. (2015b). Measurement invariance of cognitive abilities across ethnicity, gender, and time among older Americans. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *70*(3), 386–397. doi:10.1093/geronb/gbt106

Bontempo, D. E., Grouzet, F. M. E., & Hofer, S. M. (2012). Measurement issues in the analysis of within-person change. In *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences.* (pp. 97–142). Routledge/Taylor & Francis Group.

Boomsma, A. (2003). Structural equation modeling: Foundations and extensions (advanced quantitative techniques in the social sciences series. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(2), 323–331. doi: 10.1207/S15328007SEM1002_10

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 287–321. doi: 10.1207/s15328007sem1302_7

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

9 Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 471–492. doi: 10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. doi: 10.1207/S15328007SEM0902_5

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups?. *Medical Care*, *44*(3), S78–S94. doi: 10.1097/01.mlr.0000245454.12228.8f

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203. doi: 10.1016/0160-2896(84)90008-4

Hertzog, C., & Schaie, K. W. (1986). Stability and change in adult intelligence: Analysis of longitudinal covariance structures. *Psychology and Aging* 1(2), 159-171. doi: 10.1037/0882-7974.1.2.159

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117–144. doi: 10.1080/03610739208253916

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. doi: 10.1037/1082-989X.3.4.424

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

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children--Fourth Edition: What does it measure. *School Psychology Review*, *35*(1), 108–127. doi: 10.1080/02796015.2006.12088005

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. Guilford publications.

Kyllonen, P., & Kell, H. (2017). What is fluid intelligence? Can it be improved?. *Cognitive abilities and educational outcomes: A Festschrift in honour of Jan-Eric Gustafsson*, 15-37. doi: 10.1007/978-3-319-43473-5_2

Loehlin, J. C. (2004). *Latent Variable Models*. Psychology Press. doi: 10.4324/9781410609823

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391–410. doi: 10.1037/0033-2909.103.3.391

McArdle, J. J., Fisher, G. G., & Kadlec, K. M. (2007). Latent variable analyses of age trends of cognition in the Health and Retirement Study, 1992-2004. *Psychology and Aging*, *22*(3), 525–545. doi: 10.1037/0882-7974.22.3.525

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64–82. doi: 10.1037/1082-989X.7.1.64

Meredith, W. (1993a). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 134-148. doi: 10.1007/BF02294825

Meredith, W. (1993b). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. doi: 10.1007/BF02294825

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research* 3(1), 111-130. doi: 10.21500/20112084.857

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. In *Developmental Review* 41, 71-90. doi: 10.1016/j.dr.2016.06.004

Rosseel, Y. (2012). Lavaan : An *R* package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. doi: 10.18637/jss.v048.i02

Schneider, W. J., & McGrew, K. S. (2012a). The Cattell-Horn-Carroll model of intelligence. The Guilford Press.

Schneider, W. J., & McGrew, K. S. (2012b). The Cattell-Horn-Carroll model of intelligence. The Guilford Press.

Team, R. D. C. (2010). R: A language and environment for statistical computing. *(No Title)*.

Van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment* 1–38. doi: 10.4324/9781410611758

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. In *Organizational Research Methods*, 3(1), 4-70. doi: 10.1177/109442810031002

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. *Sage Publications*

Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, *30*(7), 1006-1016. doi: 10.1080/13854046.2016.1205136

Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, *30*(7), 1006-1016. doi: 10.1080/13854046.2016.1205136

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice*, *29*(3), 39-47. doi: 10.1111/j.1745-3992.2010.00182.x

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18. doi: 10.1111/j.1750-8606.2009.00110.x

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research.* 281-324. doi: 10.1037/10222-009

Yudiana, W., & Putra, K. A. (2022). *Faxtor Cognitive Ability Test Technical Manual Guideline* (Safira Putri Salsabila, Fathiyya Nur Rahmani, Aryo Bimo Adjie, & Anjar Kartaputra, Eds.). Faxtor.