Validity and Reliability Study of IST 2000R Indonesian Version

Fadillah[1], Medianta Tarigan[2]
Bandung Institute of Technology, Indonesia[1]
Department of Psychology, Indonesian University of Education, Indonesia[2]
fadillah@itb.ac.id

Abstract

In addition to the limited use of the Intelligence Structure Test 2000 Revised (IST 2000R), the adaptation process of the IST 2000R into the Indonesian version is thought to influence this test. Therefore, this study aims to conduct a review of the psychometric properties of the IST 2000R Indonesian version. The research method used is quantitative research and 919 participants, aged 17-30 years, M = 380; F=539 were involved as the sample. The data analysis technique used consisted of the construct validity test using Confirmatory Factor Analysis (CFA) and estimation of internal consistency reliability Cronbach Alpha. Analysis showed that fit results after modification indices with a p-value is 0.05675, RMSEA index is 0.023 (≤0.05), GFI 0.99, CFI 0.99. and NFI 0.98 (≥ 0.9). The results of Composite Reliability (CR) on IST 2000R show that the reliability of sub-tests that measure numerical and abstract intelligence is good, while the reliability of verbal sub-test items is low shown by CR is 0.44.

Keywords: Confirmatory factor analysis, Intelligence Structure Test 2000 revised, reliability, validity

Received 18 June 2022/Accepted 30 November 2022 ©Author all rights reserved

Introduction

This study aims to test the validity of the IST 2000R which has been adapted into Indonesian. IST or *Intelligenz-Structure-Tests* is an intelligence test that has been used by practicing psychologists in Indonesia, both for individual and mass examinations, in industrial and organizational, educational, or clinical settings (Sirodj, 2018). The IST commonly used in Indonesia is the initial version of IST, namely IST 70 which consists of 9 subtests (Rahmawati, 2014). IST 70 is in great demand because it is considered to have a comprehensive examination element, which can measure aspects of verbal, numerical, and figural-spatial intelligence (Akmal et al., 2021). IST 70 was first adapted by the Faculty of Psychology, the University of Padjadjaran around the 1970s for examination purposes in the Psychology Division of Indonesian Army (Agustin & Sirodj, 2018a).

The IST 70 Indonesian version has been tested for its construct validity by Agung & Fitri (2020). It was found that the 9 subtests construct a fit model or contribute significantly to the model, where the highest loading factor is in the ZR subtest and the lowest is in the WU subtest (Agung & Fitri, 2020). Other studies also showed that IST 70 has appropriate construct validity or fit to measure intelligence with the factor loading of each significant subtest ranging from 0.449 to 0.730 ($p < 0.05$) (Akmal et al., 2021). In other countries, IST 70 has received criticism. It is considered not to have a systematic theoretical basis and a strong psychometric explanation regarding the model of the relationship between factors (Brocke et al., 1998). In Indonesia, previous studies recommend reviewing or revising the items due to their low quality of items (Adinugroho, 2016; Agustin & Sirodj, 2018a; Rahmawati, 2014).

Until now, the quality of the IST 70 is not only questioned by psychologists but also by experts in general. In the industrial and organizational settings, for example, many companies use this test as the primary pre-hire test (Rosa et al., 2019). Many people try to find information about the IST 70 content, that can help them pass the pre-hire test. This leads some parties to open access to IST 70 content on the internet or even sell the manual book . The confidentiality of the test is not maintained. This certainly undermines the validity of the IST 70 because the score result cannot guarantee a reflection of the individual's ability due to content leakage (Rahmawati, 2014). Not only that, many argue that this test has entered the obsolescence or expiration period. Some items are considered irrelevant to the current context, especially questions related to verbal information. So, more effort is needed to develop measuring instruments in Indonesia (Suwartono, 2016).

IST has undergone various developments and the latest version is available, namely IST 2000R. This latest version is still using the postulates proposed by Amtheur regarding the hierarchical framework or prototype of intelligence structure research (HPI model) as in the early development of IST (Amthauer et al., 1999). The HPI model captures the components of intelligence that are repeatedly tested to correlate and converge with the intelligence structure model, namely verbal, numerical and spatial-figural components (Buehner et al., 2006). The IST 2000R can measure four of the seven HPI convergence factors, namely verbal intelligence, numerical intelligence, figural intelligence, and reasoning (André Beauducel et al., 2001). The three components are measured in the basic module and reasoning is measured in other additional modules. Although the postulates used are not different from those in IST 70, further testing of the strength of the IST 2000R construct is needed because

there are some changes, especially related to the items being built, which are completely new forms of items and different subtest construct arrangements from the previous version.

The IST 2000R consists of 3 modules that can be used separately. The first module is a basic module consisting of 9 subtests that measure 3 aspects of intelligence, namely verbal, numerical and figural-spatial abilities (Schmidt-Atzert, 2008). These three aspects are also measured on the IST 70. The difference is that the IST 2000R provides a more proportional number of subtests; 3 subtests for each of the intelligence aspects. In addition, 2 subtests that measure memory ability are provided, which are only used when there is a need to examine this aspect (Liepmann et al., 2007). This is unlike the previous edition, IST 70, which required the memory subtest to be performed to obtain a total IQ score. For IST 2000R, the memory subtest was excluded from the basic module, so this subtest was not included in the measurement of the total IQ score. This is because the memory subtest has a different construct in measuring general intelligence (A Beauducel et al., 2010). As part of the test development, the IST 2000R provides an additional module, a module that measures aspects of general knowledge. Items in this module measure an individual's insight into geography/history, economics, arts/culture, mathematics, natural sciences, and everyday life (Liepmann et al., 2007). This development is to accommodate the need for the measurement of fluid intelligence (gf) and crystallized intelligence (gc) as the basis for the theory of intelligence proposed by Cattel. However, it should be noted that in this additional module, the items contain a cultural dependence aspect which is not suitable for use by individuals who do not live or grew up in Germany (Liepmann et al., 2007). In this study, the analysis was only carried out on 9 subtests that are part of the basic module of the IST 2000R.

Previously, psychometric analysis of IST in Indonesian had been carried out in several studies (Agung & Fitri, 2020; Agustin & Sirodj, 2018b; Akmal et al., 2021; Kumolohadi & Suseno, 2012), but research on psychometric property testing IST 2000R Indonesian version are still limited. In this study, CFA (Confirmatory Factor Analysis) analysis was used to test the validity of the IST 2000R construct. In a complex scale development process, CFA is an important tool for assessing the internal, or latent, structure of an instrument (Brown & Croudace, 2015). The CFA analysis technique is an a priori measurement model in which both the number of factors and indicators are determined explicitly, which is then used to assess the suitability of the factor structure by testing the fit model with the

data (Alhija, 2010). The fit of the item structure is measured by the loading factor. The greater the loading factor value, the more it shows the appropriate dimension (Thompson, 2006).

Reliability was also measured in this study to determine the level of internal consistency on each item. The validity and reliability of the test are the most important things to consider when dealing with measurement (Ahmad & Ahlan, 2015). There are two types of reliability considered, namely internal reliability, and construct reliability (CR). Internal reliability is a concept that refers to the extent to which all items measure the same basic construct (Pallant, 2007) while construct reliability is a concept to assess the extent to which a measuring instrument accurately measures the theoretical construct it designed (Jackson, 2003). By knowing the information about the psychometric properties of the IST 2000R, this test can be believed to be an appropriate intelligence test tool to be used by practitioners and researchers in Indonesia.

Method

Participants

This study included 919 students, from State University with undergraduate education levels (L=539; P=380), aged 17-20 years, in several cities in Indonesia; Bandung, Bogor, Jakarta, Malang, Medan, Padang, Pekanbaru, Palembang, Semarang, Surabaya, and Yogyakarta. The following table is the distribution of research participant data.

Table 1
Participant Demography

Type of demography		Total Participants	%
Gender	Female	380	41.35%
	Male	539	58.65%
Age	17 y.o	98	10.12%
	18 y.o	620	67.46%
	19 y.o	188	20.46%
	20 y.o	17	1.85%
City	Bandung	136	14.80%
	Bogor	116	12.62%
	Jakarta	105	11.43%
	Malang	73	7.94%
	Medan	59	6.42%
	Padang	27	2.94%
	Pekanbaru	31	3.37%
	Palembang	52	5.66%
	Semarang	100	10.88%
	Surabaya	120	13.06%
	Yogyakarta	100	10.88%
Total		919	100%

Instrument

The instrument used in this study is the IST 2000R, which consists of 3 group aspects; Verbal, Numerical, and Figural. Each group aspect consists of 3 subtests and each subtest consists of 20 questions. The time needed to complete this test is 77 minutes. Eight of the nine subtests on the IST 2000R are multiple choice questions, while the CA subtest is free answer questions. After testing, scoring is carried out according to the provisions of the IST 2000R module. In Table 3, the explanation of instrument content can be seen more detail.

Table 2

IST 2000R Structure

Aspect		Subtest	Description	Total items	Time	Typo of quest.	Type of Score
Verbal	1.	*Sentence Completion* (SC)	Contains sentences with one word missing.	20	*6*	*Multiple choice*	Right(1)/ Wrong(0)
	2.	*Verbal Analogies* (VA)	The relationship between two words and find a word that has a similar relationship to another word.	20	*7*	*Multiple choice*	Right(1)/ Wrong(0)
	3.	*Similarities Subtest* (VS)	Presents six word groups with the task of finding two words with the same term.	20	*8*	*Multiple choice*	Right(1)/ Wrong(0)
Numerical	4.	*Numerical Calculations* (CA)	Contains arithmetic tasks with real numbers.	20	*10*	*Free Answer*	Right(1)/ Wrong(0)
	5.	*Number Series* (NS)	Presents a series of numbers that are formed according to a certain pattern and are asked to continue the pattern.	20	*10*	*Multiple choice*	Right(1)/ Wrong(0)
	6.	*Numerical Signs* (SI)	Choosing the correct mathematical operators for mathematical equations.	20	*10*	*Multiple choice*	Right(1)/ Wrong(0)
Figural	7.	*Figure Selection* (FS)	Geometric shapes presented with multiple pieces resulting from cutting one of the shapes with the task and identifying all the shapes that can be built	20	*7*	*Multiple choice*	Right(1)/ Wrong(0)
	8.	*Cubes* (CU)	Identify the rotated cube	20	*9*	*Multiple choice*	Right(1)/ Wrong(0)
	9.	*Matrices* (MA)	Presented a set of images arranged according to certain rules	20	*10*	*Multiple choice*	Right(1)/ Wrong(0)

Procedure

Data collection was carried out from June-December 2019. The administration of the test was carried out referring to IST 2000R module by 2 psychologists in each area, who had previously attended workshops related to the use of IST 2000R. The test was administrated with a maximum number of 20 participants per session.

Data analysis

This study used an analytical approach with the CFA method to determine the extent to which all subtests in the IST 2000R measure the same construct or are unidimensional in nature. Some of the steps carried out in the CFA method are establishing model specifications; describing the theoretical model and creating a path diagram. After the model is set, parameter estimation is carried out using the maximum likelihood method. The next step is to match the model with some parameter indices. Because it is confirmatory, the index that will be used is chi-square. However, to anticipate because the chi-square is sensitive to the size of the sample, another fit model index is used; Root Mean Squared Error of Approximation (RMSEA) < 0.05; Comparative Fit Index (CFI), NFI, & GFI > 0.90; to indicate an acceptable fit (L.T Hu & Bentler, 1995). The overall model fit assessment is done by looking at the size of the model fit index (goodness of fit index). If the model is accepted, then it can be continued with the significance test. CFA analysis was carried out with LISREL 8.80 software.

Result

Descriptive analysis

Based on the results of the descriptive analysis, it can be seen that none of the participants was able to achieve the maximum score (180). The highest total score was 154, with the highest average score obtained from subtest 4 Calculation (CA): 18.36, and the smallest average score was in subtest 1 Sentence Completion (SC): 9.36.

Table 3

Descriptive analysis

	1_SC	2_VA	3_VS	4_CA	5_NS	6_SI	7_FS	8_CU	9_MA	Total
Mean	9.36	11.11	11.17	18.36	16.09	16.75	11.73	11.24	12.22	118.04
Standard Error	0.08	0.06	0.09	0.06	0.13	0.11	0.12	0.13	0.09	0.50
Median	9.00	11.00	12.00	19.00	17.00	18.00	12.00	11.00	12.00	119.00
Mode	9.00	11.00	12.00	19.00	20.00	20.00	10.00	12.00	12.00	124.00
Standard Deviation	2.39	1.91	2.85	1.74	3.93	3.29	3.49	3.81	2.64	15.09
Sample Variance	5.72	3.64	8.11	3.04	15.42	10.84	12.20	14.54	6.98	227.76
Kurtosis	-0.30	0.09	1.52	9.91	0.23	1.09	-0.54	-0.32	0.36	0.42
Skewness	0.18	0.02	-1.02	-2.28	-1.03	-1.12	-0.10	-0.13	-0.41	-0.49
Range	15.00	13.00	16.00	17.00	18.00	19.00	18.00	20.00	17.00	105.00
Minimum	3.00	5.00	1.00	3.00	2.00	1.00	2.00	0.00	2.00	49.00
Maximum	18.00	18.00	17.00	20.00	20.00	20.00	20.00	20.00	19.00	154.00
Sum	8601	10208	10265	16875	14790	15395	10782	10333	11227	108476
Count	919	919	919	919	919	919	919	919	919	919

The widest standard deviation was found in the subtest 5 Serial Number (NS) (3.93), which means that in this subtest the data obtained deviates the furthest from zero or can be said as heterogeneous. The narrowest standard deviation was found in subtest 4 Calculation (CA) (1.74) so the data is assumed to be quite homogeneous. Meanwhile, a fairly large range was found in the subtest 8 Cube (CU) with a value of 20, which means that there were participants who had a score of 0 to a perfect score of 20. In addition, descriptive analysis data found that the highest average score is obtained in the numerical group subtest. (CA, NS, and SI), then the figural group subtest (FS, CU, MA), and finally the verbal group subtest (SC, VA, VS).

Confirmatory Factor Analysis (CFA)

CFA is used to determine how a set of measured items represents a construct. Key relationship links are constructed to variables (factor loading estimates) and constructed to each other (correlation constructs). By estimating this relationship, researchers can conduct an empirical examination of the proposed measurement theory (Joe F. Hair et al., 2014). Initial testing on each dimension showed unfit results because some of the fit model criteria were not met, so a modification of the model was carried out by correlating the errors in each item.

Table 4

Overall Model Fit Test

GOF	Acceptable Match Rate	Model Index	Notes
Chi-square	The smaller value the better (p-value ≥ 0.05)	0.05675	Good
GFI	GFI ≥ 0.90 good fit	0.99	Good Fit
	0.80 ≤ 0.90 marginal fit		
RMSEA	RMSEA ≤ 0.05 good fit	0.023	Good Fit
	0.80 ≤ NNFI ≤ 0.90 marginal fit		
CFI	CFI ≥ 0.90 good fit	0.99	Good Fit

The cut-off values used for goodness of fit are Chi-Square/DF < 5, GFI > .90, and RMSEA ≤ .08 (Li Tze Hu & Bentler, 1999; Schermelleh-Engel et al., 2003). Based on the model fit test results, it is found that the model indexes match the acceptable rate. This result indicates that the 1st Order CFA model is fit in measuring the construct validity of the IST 2000R Indonesian version. Furthermore, the validity of the subtest was tested to evaluate the latent variables with several indicators to see if the subtest significantly measured the factors to be measured using the 1st order CFA model, as shown in the following figure:

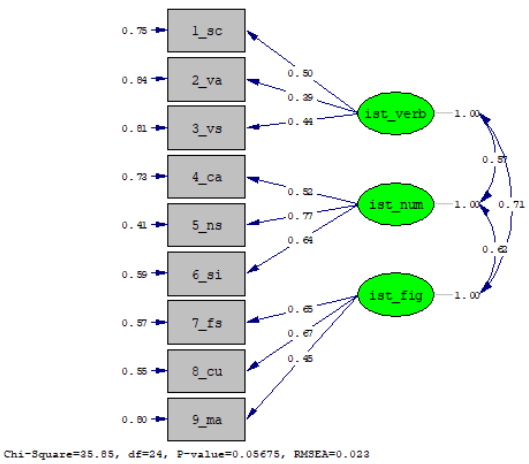

Figure 1. Results of Standardized Loading Factors

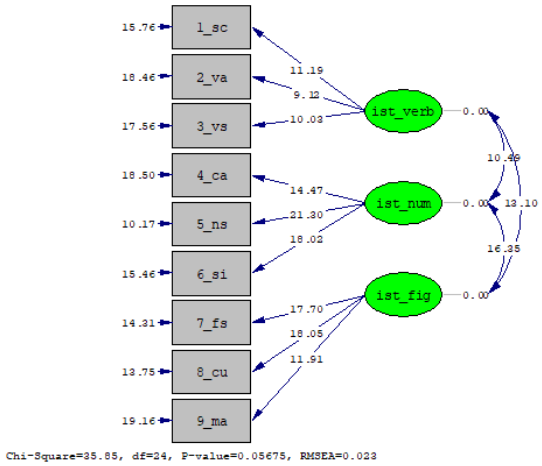

Figure 2. T-value results

The magnitude of the significance of the subtest on each latent variable is measured based on the value of the standardized loading factor. The subtest is said to be significant if the standardized loading factor value is 0.3 (Thompson, 2006). The greater the value of the standardized loading factor, the more it indicates that the subtest measures the appropriate group of subtests.

In testing the reliability, this study refers to the recommendations of Hair et al. (1998) that the value of Construct Reliability (CR) is in the range of 0.60 and 0.70. In addition, to determine the contribution of exogenous variables to endogenous variables, it can be seen from the R Square value in each subtest. The coefficient of determination (R^2) measures how far the model's ability to explain variations in endogenous (latent) variables (Ghozali & Fuad, 2008).

Table 5

Results of Subtest Validity and Reliability

Latent Variable	Subtest	Valid if		Note	Reliable if		R Square
		Standardized *Loading factor* (SLF) ≥ 0.30	***t-value* > 1.96**		**CR ≥ 0.6**	**Note**	
Verbal	*Sentence Completion* (SC)	0.50	11.19	Valid	0.42	Low reliability	0.25 (25%)
	Verbal Analogies (VA)	0.39	9.12	Valid			0.16 (16%)
	Similarities Subtest (VS)	0.44	10.03	Valid			0.19 (19%)
Numerik	*Numerical Calculations* (CA)	0.52	14.47	Valid	0.68	Reliable	0.27 (27%)
	Number Series (NS)	0.77	21.30	Valid			0.59 (59%)
	Numerical Signs (SI)	0.64	18.02	Valid			0.41 (41%)
Figural	*Figure Selection* (FS)	0.65	17.70	Valid	0.62	Reliable	0.43 (43%)
	Cubes (CU)	0.67	18.05	Valid			0.45 (45%)
	Matrices (MA)	0.45	11.91	Valid			0.20 (20%)

Based on table 5, it can be found that the IST 2000R has good construct validity and good reliability for subtests measuring numerical intelligence and figural intelligence. Meanwhile, the subtest that measures verbal intelligence is still classified as having low reliability (CR 0.44; <0.6). In addition, the result showed that the largest R^2 value is given by the serial number subtest (NS) in the Numerical subtest group (0.59), which means that this subtest can explain the variation from the Numerical

subtest group by 59%. In the Verbal group, the largest contribution was obtained from the SC subtest, which was 25%. As for the Figural group, the largest contribution was obtained from the Cube subtest, which was 45%.

Reliability Analysis

Based on the reliability results table above, it can be seen that the SC, VA, and MA subtests have low reliability (< 0.5). The VS, CA, SI, FS, and CU subtests have moderate reliability (0.5 - 0.7) and the NS subtests have high reliability (0.7 - 0.9).

Table 6

Reliability Test Results of Each IST 2000R Subtest

IST 2000R Subtest	*Split Half Spearman-Brown Coefficient*	*Cronbach Alpha*
SC Subtest (*Sentence Completion*)	0.319	0.353
VA Subtest (*Verbal Analogies*)	0.167	0.186
VS Subtest (*Verbal Similarities*)	**0.665**	**0.639**
CA Subtest (*Calculations*)	**0.605**	**0.632**
NS Subtest (*Number Series*)	**0.734**	**0.864**
SI Subtest (*Numerical Signs*)	**0.692**	**0.806**
FS Subtest (*Figure Selection*)	**0.568**	**0.679**
CU Subtest (*Cubes*)	**0.569**	**0.744**
MA Subtest (*Matrices*)	0.365	**0.543**

Discussion

This study aims to see the consistency of the IST 2000R Indonesian version structure. Confirmatory factor analysis was used to confirm the 9 subtests referring to the 3 intelligence factors. The overall model fit test results showed that the model meets the acceptable fit level index. The index that shows the model residue can be seen through the RMSEA index. The smaller the residue, the better the model. In the overall IST 2000R model, the RMSEA index shows 0.023 (≤ 0.05), a small index which means that the model is closed and the role of other factors in explaining the model is very low. This is also supported by the chi-square value resulting in a p-value > 0.05 (not significant), which

means that the model only measures one factor or it is unidimensional. That is, all items only measure one factor, namely intelligence.

The GFI value in the model showed 0.99 (≥ 0.90), which means that the compiled model is adequate if it is returned to the research population. Likewise, the CFI value showed an index of 0.99 (≥ 0.90). That is, the comparison value of the model compiled with the ideal model is close to the value 1 or the comparison is adequate. This result is in line with the construct validity test conducted on the German version of the IST 2000R through confirmatory factor analysis (Liepmann et al., 2007). The analysis carried out confirmed the presence of three factors in the IST 2000R, namely the verbal, numerical, and figural factors group. The model fit index obtained in this study is as follows AGFI of 0.96; RMSR of 0.030 and CFI of 0.98.

This study also obtained information related to the factor loading of each subtest. In the verbal group, the largest factor load was in the Sentence Completion (SC) subtest of 0.50, while the smallest factor load was the Verbal Analogy (VA) subtest, which was 0.39. This result is in contrast to the results of the analysis conducted on the German version of the IST 2000R (Liepmann et al., 2007), wherein the verbal group the Verbal Analogy subtest had the largest factor load of 0.80, while the subtest with the smallest factor load was the Sentence Completion subtest (SC) of 0.62. In the numerical group, the largest factor load is in the Serial Number (NS) subtest of 0.77, while the smallest is the Calculation (CA) subtest of 0.52. A different result is also shown in this case from the German version where the Serial Number (NS) subtest has the smallest index while the Numerical Signs (SI) subtest has the largest factor loading index. In the figural group, the results of the factor analysis of this study indicate that the Cube subtest (CU) has the largest factor load index of 0.67 while the Matrix (MA) subtest has the smallest factor load of 0.45. In the German version, the largest factor load is in the Image Selection (FS) subtest, while the smallest is the Matrix (MA) subtest of 0.66. Further research is needed to determine the cause of this difference.

Table 7

Reliability Test Results of Each IST 2000R Indonesian Version and German Version

Latent Variable	Subtest	German version (Liepmann et al., 2007)		Indonesian version	
		Cronbach Alpha	Split Half Spearman-Brown	Cronbach Alpha	Split Half Spearman-Brown
Verbal	*Sentence Completion* (SC)	0.69	0.71	0.35	0.32
	Verbal Analogies (VA)	0.70	0.74	0.19	0.17
	Similarities Subtest (VS)	0.77	0.81	0.64	0.66
Numerik	*Numerical Calculations* (CA)	0.87	0.91	0.63	0.60
	Number Series (NS)	0.91	0.94	0.86	0.73
	Numerical Signs (SI)	0.89	0.91	0.81	0.69
Figural	*Figure Selection* (FS)	0.77	0.81	0.68	0.57
	Cubes (CU)	0.81	0.86	0.74	0.57
	Matrices (MA)	0.73	0.77	0.54	0.36

This study also compares the reliability of each subtest with the results of previous studies. The results of the German version, of the reliability test, gave quite high, which were in the range of 0.86 to 0.95 on the Cronbach alpha reliability coefficient, and 0.89 to 0.96 on the split-half reliability coefficient (Liepmann et al., 2007). Overall, the German version of the IST 2000R showed a higher reliability coefficient than the Indonesian version. The biggest difference is in the Verbal Analogy subtest, where the index of the Indonesian version shows 0.19 (Cronbach Alpha) and 0.17 (Split Half Spearman-Brown) while the German version shows 0.70 (Cronbach Alpha) and 0.74 (Split Half Spearman-Brown). The difference in results may be caused by the heterogeneity of the subjects involved. (Chairunisa, 2016) in her research showed that there are significant differences in the reliability coefficients produced by homogeneous groups and heterogeneous groups. The heterogeneous group produces a higher reliability coefficient than the homogeneous group. The subjects used to measure the validity of the IST 2000R German version were 15-56 years old, with 30.1% not meeting the university entry requirements and the remaining 69.9% meeting the university entry requirements (Liepmann et al., 2007). While in this study, the age range of participants was 17-20 years, with the same educational level background (S1). This difference indicates that the German version is more heterogeneous than the subjects used in this study, so the reliability coefficient of

the German version of the IST 2000R is higher than the Indonesian version. Further research is encouraged to test the reliability of this test by involving more heterogeneous subjects.

Although requires further research on several issues, the result of the current study might have important implications for the development of intelligence tests in Indonesia. The IST 2000R Indonesian version has a fit model and each factor is confirmed to be unidimensional in measuring intelligence. This tool can be used as an alternative measurement for the relatively newer intelligence measurement.

Conclusion

Based on the results of the study, it can be concluded that the IST 2000R Indonesian version measures aspects that should be measured according to the construct. The Goodness of Fit test using the Maximum Likelihood approach showed that the significance level meets the criteria and can be considered fit. The results of the Composite Reliability (CR) index on the IST 2000R showed that the subtest group that measures numerical and abstract intelligence has good measurement consistency, while the subtest group for verbal intelligence is still classified as having a low reliability. This study recommends the use of a more heterogeneous sample for further research in the future.

References

Adinugroho, I. (2016). Pengujian Properti Psikometrik Intelligenz Struktur Test Subtes Kemampuan Spasial Dua Dimensi (Form Auswahl): Studi Pada Dua Sma Swasta Di Jakarta. *Jurnal Ilmiah Psikologi MANASA*, *5*(2), 165. https://ejournal.atmajaya.ac.id/index.php/Manasa/article/view/181

Agung, I. M., & Fitri, A. R. (2020). Analisis Psikometri Intelligenz Structur Test (IST) pada Mahasiswa. *Psikobuletin:Buletin Ilmiah Psikologi*, *1*(1), 1–10. https://doi.org/10.24014/pib.v1i1.8174

Agustin, D., & Sirodj, N. (2018a). Analisis Kualitas Aitem Intelligence Structure Test (IST) melalui Metode Item Response Theory (IRT). In *SCHEMA (Journal of Psychological Research*. https://doi.org/https://doi.org/10.29313/schema.v4i2.4420

Agustin, D., & Sirodj, N. (2018b). Analisis Kualitas Aitem Intelligence Structure Test (IST) melalui Metode Item Response Theory (IRT). In *SCHEMA (Journal of Psychological Research*.

Ahmad, B. I., & Ahlan, A. R. (2015). Reliability And Validity Of A Questionnaire To Evaluate Diabetic Patients' Intention To Adopt Health Information Technology: A Pilot Study. *Journal of Theoretical and Applied Information Technology*, *72*(2), 253–264. http://www.jatit.org/volumes/Vol72No2/12Vol72No2.pdf

Akmal, N., Widyastuti, & Nur, H. (2021). Psychometric Proporties Analysis of IST (Intelligenz Structur Test). *Indonesian Journal of Educational Studies (IJES)*, *24*(2), 126–142. https://doi.org/10.26858/ijes.v24i2.29797

Alhija, F. A. N. (2010). Factor analysis: An overview and some contemporary advances. In *International Encyclopedia of Education* (pp. 162–170). https://doi.org/10.1016/B978-0-08-044894-7.01328-2

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (1999). Intelligenz-Struktur-Test 2000. *Hogrefe, Göttingen*.

Beauducel, A, Liepmann, D., Horn, S., & Brocke, B. (2010). *Intelligence Structure Test*. Hogrefe.

Beauducel, André, Brocke, B., & Liepmann, D. (2001). Perspectives On Fluid And Crystallized Intelligence: Facets For Verbal, Numerical, And Figural Intelligence. *Personality and Individual Differences*, *30*(6), 977–994. https://doi.org/10.1016/S0191-8869(00)00087-8

Brocke, B., Beauducel, A., & Tasche, K. (1998). Der intelligenz-struktur-test: Analysen zur theoretischen grundlage und technischen güte. *Diagnostica*, *44*(2).

Brown, A., & Croudace, T. J. (2015). Scoring and estimating score precision using IRT. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment (a volume in the Multivariate Applications Series)*. Routledge. https://www.researchgate.net/publication/273685295_Scoring_and_estimating_score_precision_using_IRT

Buehner, M., Krumm, S., Ziegler, M., & Pluecken, T. (2006). Cognitive Abilities and Their Interplay. *Journal of Individual Differences*, *27*(2), 57–72. https://doi.org/10.1027/1614-0001.27.2.57

Chairunisa, E. D. (2016). Komparasi Estimasi Reliabilitas Pada Mata Pelajaran Sejarah Ditinjau Dari Homogenitas Dan Heterogenitas Kelompok. Jurnal Pendidikan Ilmu Sosial. *JURNAL PENDIDIKAN ILMU SOSIAL*, *24*(2), 179–180. https://doi.org/10.17509/jpis.v24i2.1454

Ghozali, I., & Fuad, F. (2008). *Structural Equation Modeling*. Universitas Diponegoro.

Hair, Joe F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. In *European Business Review*. https://doi.org/10.1108/EBR-10-2013-0128

Hair, Joseph F., JR, Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis (5th ed.)*. Prentice Hall , Inc.

http://opac.lib.unlam.ac.id/id/opac/detail.php?q1=515.7&q2=Jos&q3=M&q4=-

Hu, L.T, & Bentler, P. . (1995). Evaluating Model Fit In R. H. Hoyle (Ed.) Structural Equation Modeling: Concepts, Issues, and Application. In *Sage Publication*. Sage Publication.

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

Jackson, S. L. (2003). *Research Methods and Statistics, A Critical Thinking Approach*. Thomson Wadsworth.

Kumolohadi, R., & Suseno, miftahun ni'mah. (2012). Intelligenz Struktur Test Dan Standard Progressive Matrices : (Dari Konsep Inteligensi Yang Berbeda Menghasilkan Tingkat Inteligensi Yang Sama). *Jurnal Inovasi Dan Kewirausahaan*.

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). In *Verlag für Psychologie Hofgrefe*.

Pallant, J. (2007). *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows, 3rd ed*. McGraw Hill Open University Press.

Rahmawati, E. (2014). Evaluasi Karakteristik Psikometri Intelligenz Struktur Test (IST). *Proceeding Seminar Nasional Psikometri*, 270–282. http://hdl.handle.net/11617/6418

Rosa, S. J., Fitri, A. R., & Agung, I. M. (2019). Psychometric properties of the SE-Revised: A rasch model analysis. *HUMANITAS: Indonesian Psychological Journal*, *16*(1), 54. https://doi.org/10.26555/humanitas.v16i1.10919

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *MPR-Online*.

Schmidt-Atzert, L. (2008). Intelligenz-Struktur-Test 2000 R. *Zeitschrift Für Personalpsychologie*.

Sirodj, D. A. N. (2018). Analisis Kualitas Aitem Intelligence Structure Test (IST) Melalui Metode Item Response Theory (IRT). *SCHEMA (Journal of Psychological Research)*, 98–108.

Suwartono, C. (2016). Alat Tes Psikologi Konteks Indonesia: Tantangan Psikologi di Era MEA. *Jurnal Psikologi Ulayat*, *3*(1), 1–6. https://doi.org/10.24854/jpu12016-51

Thompson, B. (2006). Exploratory and confirmatory factor analysis: Understanding concepts and applications. In *American Psychological Association*. https://doi.org/10.1037/10694-000