Jurnal Penelitian dan Pengukuran Psikologi
Volume 11, Nomor 01, April 2022 http://doi.org/10.21009/JPPP.111.08

CONSTRUCT VALIDITY OF THE *NIJMEEGSE SCHOOLBEKWAAMHEIDS* TEST (NST)

Medianta Tarigan[1] & Fadillah[2]
Department of Psychology, Universitas Pendidikan Indonesia [1]
Faculty of Art and Design, Institut Teknologi Bandung [2]

***Email**: medianta@upi.edu*

Abstract

This study aims to provides construct validity information about the psychometric properties of the Nijmeegse Schoolbekwaamheids Test (NST). NST is one of the instruments that is widely used to measure school readiness. The NST has 10 subtests in the form of pictures and stories that were adapted from the Netherlands. This study involved 471 pre-school students aged 4-6 years. This study used IRT to determine the quality of each item, Exploratory Factor Analysis to determine the new factors formed from the 10 available subtests, and Confirmatory Factor Analysis to determine to construct validity. The IRT 3PL model showed that all items in the NST are valid, but 54 tis necessary to review several items, especially in subtest 10. In the EFA factor analysis, 4 new factors were found that measure the level of factor: logic reasoning ability, memory, fine motor skill, and conceptual maturity in children. These are valid and reliable as evidenced by the RMSEA value on the CFA of 0.05 and the reliability value of 0.87.

Keywords: Nijmeegse Schoolbekwaamheids Test, Item Response Theory, Exploratory Factor Analysis, Confirmatory Factor Analysis

Abstrak

Penelitian ini bertujuan untuk memberikan informasi validitas konstruk terkait properti psikometri pada instrumen *Nijmeegse Schoolbekwaamheids Test* (NST). NST merupakan salah satu instrumen yang banyak digunakan untuk mengukur kesiapan anak dalam memasuki sekolah. Instrumen NST memiliki 10 subtes berbentuk gambar dan cerita yang diadaptasi dari Belanda. Penelitian ini melibatkan 471 partisipan yang merupakan siswa pra-sekolah berumur 4-6 tahun. Beberapa metode analisis yang digunakan yaitu analisis item IRT untuk mengetahui kualitas pada setiap item, Exploratory Factor Analysis untuk menentukan faktor baru yang terbentuk dari 10 subtes yang tersedia, dan Confirmatory Factor Analysis untuk mengetahui validitas konstruk. Hasil menunjukkan bahwa dengan analisis item IRT 3PL hampir seluruh aitem dinyatakan valid, keculi aitem yang ada pada subtest 10 yang membutuhkan telaah lebih lanjut. Pada analisis factor EFA didapatkan 4 faktor baru yang masing-masing factor mengukur tingkat pemahaman logis, motoric halus, memori, dan kematangan konseptual pada anak. Keempat faktor ini valid dan reliabel yang dibuktikan dengan nilai RMSEA pada CFA sebesar 0.05 dan nilai reliabilitas sebesar 0.87.

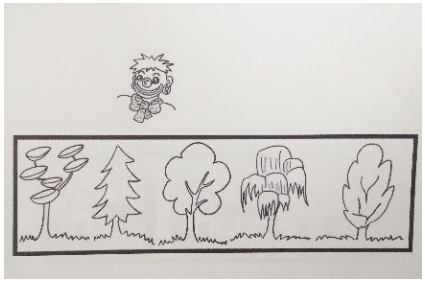
Kata Kunci: Nijmeegse Schoolbekwaamheids Test, Item Response Theory, Exploratory Factor Analysis, Confirmatory Factor Analysis

1. Introduction

The transition that children experience as they start entering school is complex. In this transition phase, children change their role as part of a small social group, the family, to a wider social group such as friends, teachers, and other parties. At the same time, children also prepare themselves for a structured and independent academic world. School readiness has become a topic that is widely raised by researchers to help schools, government and families know how well children can go through this transition phase to give effective responses. School readiness is the interaction between children's characteristics and schools' flexibility to provide for the needs of each child in their early years of formal schooling (Cuskelly & Detering, 2003).

Maturity progress is one of the indications to see a child's readiness. This includes 1) Physical Development, the progress of a child's visual-motoric coordination, which determines an individual's capability in writing. 2) Mental processes (cognitive), such as comparing, sorting, categorizing, finding hidden objects, and developing concepts, both in the form of verbal and non-verbal. 3) Social emotions; are individuals' ability to adapt to prevailing norms, such as playing play with peers, regulating emotional expression, and responding to other emotions (Papalia et al., 2010).

In knowing the extent to which children are ready, the school carries out various assessments to obtain the child's condition. In Indonesia, one of the most widely used assessment tools to determine children's school readiness is The Nijmeegse Schoolbekwaamheids Test (NST) (Mönks, Rost, & Coffie, 1978). This test was adapted from the Netherlands, and consists of 10 subtests: form perception, fine motor activity, number concepts, visual discrimination, critical perception, concentration, memory, object, and situation assessment, story retelling, and human pictures. The total score of 10 subtests is considered capable of measuring students' school readiness. NST is also considered to be substantially correlated with literacy and numeracy skills at the end of the first grade of primary school (ter Laak, 1991).



Picture 1. Preview of NST Subtest 1

However, the popularity of NST as one of the primary school entrance tests in Indonesia is not in line with the solid or robust information about this test, especially its psychometric quality. Is still unknown the theoretical construct that underlies the ten existing subtests. Often, practitioners especially psychologists, directly translate each of these 10 aspects only based on the name of the subtest, without knowing the interrelationships between aspects in interpreting the results. In the end, this baseless action in interpretation makes assessment results unreliable.

With limited information related to theories or hypotheses of the constructs that build the NST, this research will explore what constructs underlie NST, and we try to figure out how many factors exist in the NST and how they are grouped into that factor. To do this, the researcher will use the EFA (Exploratory Factor Analysis) analysis method to explore the NST's constructs to examine the NST's construct validity. EFA is a fundamental tool in developing and validating psychological theories and measurements through multivariate statistical methods (Watkins, 2018).

This study will also evaluate the item quality of the NST. In previous research, the item quality of the NST has been tested out through empirical analysis based on the Classical Test Theory (CTT), by including as many as 343 early childhood children who will enter elementary school as research participants (Mariyati & Affandi, 2016). This study recommends further research using Item Response Theory (IRT), to obtain comprehensive information. IRT can overcome the weakness of CTT which in its analysis depends on the group, item, and estimation of standard error measurement (SEM). The item quality of the NST will be tested using the IRT 3PL which consists of the level of difficulty, discrimination, and distractors.

2. Methods

Participants

This study uses primary data of the NST test scores from 471 pre-school students (4-6 years old, male=245, female=226) from various schools in Indonesia. In Tabel 1, an overview of participants' demographics is presented.

Table 1. Demographic Data

		Total	Percentage
Gender	Male	245	52%
	Female	226	48%
Age	4 y.o	16	3%
	5 y.o	348	74%
	6 y.o	107	23%

Instrument

The NST test consists of 10 subtests with pictures and stories and uses a dichotomy score (1-0) for all subtests except for subtests 7 and 9, which use a three values score (0,1,2). Subtest 1: consists of 8 questions with each containing 2-5 distractors and one answer, measuring the ability to distinguish shape. Subtest 2: consists of 8 questions containing instructions to imitate shapes, measuring fine motor skills. Subtest 3: consists of 8 questions containing 3-8 distractors, measuring the ability to understand the size, number, and comparison. Subtest 4: consists of 8 questions with instructions to find hidden objects in complex images, measuring the ability to observe in detail. Subtest 5: consists of 8 questions that contain an incomplete picture, measuring critical thinking skills. Subtest 6: consists of 8 questions with instructions to find the same object to a set of similar objects, measuring concentration. Subtest 7: consists of 8 questions with instructions to find the same image, has 8 distractors & 8 answers, measuring memorization ability. Subtest 8: consists of 8 questions each having 2-3 distractors and one answer, measuring the child's understanding of the situation and object. Subtest 9: consists of 8 questions, 8 distractors, and 8 answers with instructions to find the same object. Subtest 10: drawing the human figure, with a 1-8 scale score according to the completeness of the picture. The total score is carried out by adding up all the correct answers from 10 subtests.

Data Analysis

This quantitative research uses Items Respond Theory (IRT) analysis to emphasize the probability value of each participant in achieving a certain score to find out the quality of the items. IRT allows the efficient estimation of children's cognitive abilities while ensuring that children are given items with the right level of difficulty and in the end, the result score is more sensitive according to the theory and measurement model (Jacob & Rothstein, 2016), which can not be done by using the existing classical test theory (Guler et al., 2014). This study used 3 PL model IRT, which describes 3 parameters simultaneously; a/discrimination, b/difficulty, and c/pseudo-guessing (Hambleton et al., 1991). Software R 4.1.1 was used to measure the IRT item analysis.

Internal consistency of the NST was tested using McDonald's ω coefficient with an adequate level at >0.7 (Dunn et al., 2014). The construct validity was evaluated by exploratory factor analysis (EFA). EFA is used when the initial hypothesis about the factors is unknown or no prior knowledge of how many factors exist between items so the goal of the analysis is to determine which factor fits well into a group of items (Decoster & Hall, 1998; Hayton et al., 2004; Hurley et al., 1997).

Before running the EFA analysis, the Kaiser-Meyer-Olkin measure of sampling adequacy (> 0.50) and Bartlett's test of sphericity (p-value < 0.05) were used to assess whether the data were suitafore as EFA requirement. After determining the suitable data set, EFA analysis was run and varimax rotation was used to achieve the orthogonal approximation of the factor structure, and the eigenvalues (>1) are used to determine the initial set of factors (Maskey et al., 2018). CFA was also performed in this study to evaluate the initial structural model fit of the NST. The goodness of the model fit was used by chi-square, GFI, RMSR, RMSEA, NFI, NNFI, CFI, RFI, IFI, dan CN where values of the GFI, CFI, TLI, and NFI ,= .9the 0 and value of RMSEA and SRMR = .08 indicated an acceptable model fit (Joseph F Hair et al., 2010). JASP 0.13.1.0 was used for Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA).

Procedures

Before data gathering, the relevant teachers were contacted for consent, permission, and discussion about assessment plans. The teachers and parents were informed as to the nature and procedures of the assessment. Three psychologists with a psychometric background were recruited and familiarized themselves with the test procedures, class environment, and the participants days before testing was conducted. The participants were assessed in a group of classes consist of a maximum of 15 children per class size. Each psychologist was responsible for approx. 5 children. The standardized instruction was given in Bahasa. The duration of testing is approx one hour per class period.

3. Results

Descriptive analysis was conducted to provide a summary of the sample in this particular study. The description of the data was analyzed with several statistical measures, with the following results as presented in Table 2.

Table 2. Descriptive Analysis

Subtest	Valid	Mean	Std. Deviation	Variance	Skewness	Kurtosis	Minimum	Maximum
1	471	6.304	2.056	4.229	-1.589	2.105	0	8
2	471	2.584	1.821	3.316	0.27	-0.803	0	7
3	471	5.032	1.972	3.89	-0.883	0.331	0	8
4	471	4.972	2.324	5.401	-0.779	-0.283	0	8
5	471	5.372	2.078	4.319	-1.044	0.571	0	8
6	471	5.263	2.256	5.088	-0.901	0.012	0	8
7	471	20.03	4.914	24.148	-2.44	7.358	0	24
8	471	4.79	2.174	4.728	-0.591	-0.356	0	8
9	471	16.66	3.964	15.711	-2.734	9.463	0	23
10	471	3.015	2.038	4.155	-0.064	-1.005	0	8

The highest average score was in subtest 7 (20.03) followed by subtest 9 (16.66). As mentioned before, subtests 7 and 9 use a three values score (0,1,2), so the average score can be higher than other subtests using a dichotomy score (1,0). The range of different responses varied the most for subtest 7, with a variance value of 24,148. This value indicates that the diversity of scores on subtest 7 is quite high. But overall, the data is homogeneous because the standard deviation of each subtest is smaller than the mean value.

IRT Analysis Results

After analyzing the data, the results of the IRT-3PL model analysis are presented. The selection of the 3PL model is based on the consideration that this model produces estimates guessing parameter/c, which states the probability of guessing the correct answer. Tests rely on chi-square estimation and p-value as determining criteria for fit (significant) or not fit (not significant) items

Table 3. IRT 3PL Analysis Result

	a/discrimination		b/difficulty			c/pseudo-guessing	
Subtest	Normal	Abnormal	Easy	Average	Hard	Acceptable	Inacceptable
1	7 87.5%	1 12.5%	1 12.5%	7 87.5%	0 0%	8 100%	0 0%
2	8 100%	0 0%	0 0%	7 88%	1 13%	8 100%	0 0%
3	8 100%	0 0%	1 12.5%	7 87.5%	0 0%	8 100%	0 0%
4	6 75%	2 25%	0 0%	8 100%	0 0%	8 100%	0 0%
5	8 100%	0 0%	0 0%	8 100%	0 0%	8 100%	0 0%
6	6 75%	2 25%	0 0%	8 100%	0 0%	8 100%	0 0%
7	10 62.5%	6 37.5%	-	-	-	-	-
8	8 100%	0 0%	0 0%	8 100%	0 0%	8 100%	0 0%
9	7 46.7%	8 53.3%	-	-	-	-	-
10	1 12.5%	7 87.5%	0 0%	6 75%	2 25%	8 100%	0 0%

As presented in Table 3 above, more than 60% of all items in each subtest have the discriminant ability in the normal range. Only subtests 9 and 10 have below 50% (subtest 9= only 46,7% of items and subtest 10= only 12,5% of items). Meanwhile, the difficulty/b parameter showed that 80% of all items have acceptable difficulty levels and 100% of items indicate no pseudo-guessing. Since the difficulty parameter/b and guessing parameter/c is not as relevant for a polytomous score (Yang & Kao, 2014), this type of parameter cannot be estimated for subtest 7 and 9.

EFA Results

As mentioned earlier, the Kaiser-Meyer-Olkin measure of sampling adequacy (> 0.50) and Bartlett's test of sphericity (*p-value* < 0.05) were assessed before running EFA to test the validity of each attribute (Usman & Sobari, 2013).

Table 4. Bartlett's test

df	df	p
1607.902	45	<.001

Tabel 5. Kaiser-Meyer-Olkin test

Subtest	MSA	> 0.5
Overall MSA	0.927	v
Subtest 1	0.92	v
Subtest 2	0.944	v
Subtest 3	0.928	v
Subtest 4	0.933	v
Subtest 5	0.919	v
Subtest 6	0.928	v
Subtest 7	0.929	v
Subtest 8	0.931	v
Subtest 9	0.891	v
Subtest 10	0.941	v

In Bartlett's Test, the results showed in this study, the data meets the requirement for further analysis with a p-value <0.001 ($p<0.05$) and the Kaiser-Meyer-Olkin test also showed, fother total score NST and each subtest score, that the data meets the requirement (MSA > 0.5). As observed in Table 6, McDonald's ω for this test was .87. It is confirmed this test had adequate internal consistency.

Table 6. Internal consistency as measured by McDonald's ω coefficients

Estimate	McDonald's ω
Point estimate	0.87
95% CI lower bound	0.653
95% CI upper bound	0.697

The EFA was conducted on a sample of 471 students, and as observed 10 subtests were clustered in 4 factors (Tabel 7).

Table 7. Factor Loadings

Subtest	Factor 1	Factor 2	Factor 3	Factor 4
Subtest 1	0.672			
Subtest 2		0.935		
Subtest 3	0.668			
Subtest 4	0.594			
Subtest 5	0.707			
Subtest 6	0.49			0.467
Subtest 7	0.377		0.46	
Subtest 8	0.558		0.394	
Subtest 9			0.613	
Subtest 10				0.351

The factor structure is fit into a factor if it has a factor loading > 0.3 (Pauls & Daseking, 2021). Based on the analysis result, all factors have a factor loading> 0.3. It was found that factor 1 contained subtests 1, 3, 4, 5, 6, 8. Then factor 2 contained only subtests 2, factor 3 contained subtests 7 and 9, and subtest 4 contained only subtests 10.

CFA Results

In the next analysis, confirmatory factor analysis (CFA) was used for the four factors obtained from the previous EFA results. The Overall Model Fit showed a Chi-square index with a p-value <.001, where the statistical criteria are the

smaller the Chi-square value, the better the model fit with the data. The following Table 8 presented the results of Chi-square and another model.

Table 8. Fit Index Values Obtained in CFA

Fit Indices Examined	Criteria	Fit Indices Obtained	Keterangan
Chi-square	p-value ≥ 0.05 model fit	4.494 0.106	Model fit
GFI	GFI ≥ 0.90 good fit 0.80 ≤ GFI ≤ 0.90 marginal fit	0.995	Good Fit
RMSR	RMSR ≤ 0.05 good fit	0.019	Good Fit
RMSEA	RMSEA ≤ 0.05 good fit	0.05	Good Fit
NFI	NFI ≥ 0,90 *good fit* 0,80 ≤ NFI ≤ 0,90 *marginal fit*	0.992	Marginal Fit
NNFI	NNFI ≥ 0.90 good fit 0.80 ≤ NNFI ≤ 0.90 marginal fit	0.987	Good Fit
CFI	CFI ≥ 0.90 good fit	0.996	Good Fit
RFI	RFI ≥ 0,90 good fit 0,80 ≤ RFI ≤ 0,90 marginal fit	0.977	Marginal Fit
IFI	IFI ≥ 0,90 good fit 0,80 ≤ IFI ≤ 0,90 marginal fit	0.996	Good Fit
CN	CN ≥ 200 Good	628.874	Good

Chi-square has high sensitivity to sample size, the larger the sample size, the greater the Chi-square statistic that will be obtained. Therefore it is recommended to have other indices information to improve the accuracy of the fit model result. This study is also equipped with other tests; GFI, RMSR, RMSEA, NFI, NNFI, CFI, RFI, IFI, and CN. Form all criteria it can be concluded that this research model has a very good precision value and can be categorized as a fit model. The next step was to perform a Measurement Model Fit to evaluate the factor loading coefficient. The test was carried out by looking at the z-value on each factor loading coefficient. The following are the results of the measurement model fit test.

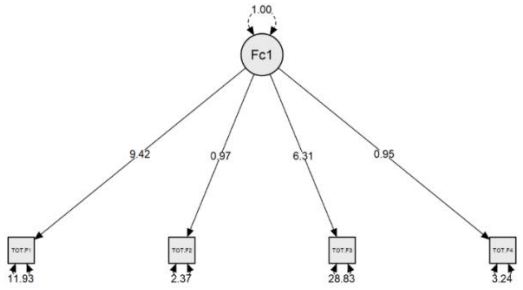


Figure 1. Four factors Model of the NST

Table 9. Factor Loading the NST

Indicator	Estimate	Std. Error	z-value	p-value	Significance
TOT.F1	11.929	4.686	2.546	0.011	V
TOT.F2	2.372	0.164	14.424	0	V
TOT.F3	28.835	2.819	10.227	0	V
TOT.F4	3.239	0.22	14.751	0	V

A variable is said to have good validity if the z value of the loading factor is greater than the critical value (or 1.96 or practically ≥ 2) and the standardized loading factor is 0.50 or ideally 0.70 or higher. However, it depends on different sample sizes, the standardized loading factor of 0.3 can be used for sample size > 300 (Joe F. Hair et al., 2014). As observed in Table 9, all subtests are significant positively, which is indicated by a z-value> 1.96 or p-value less than 0.05.

Table 10. Over All Model Fit for each Factor

Faktor	Chi-Square		p-value		RMSEA	
	Nilai	Ket,	Nilai	Ket.	Nilai	Ket.
1	1019.022	Good	0.188	Model fit	0.009	Good Fit
2	28.753	Good	0.07	Model fit	0.034	Good Fit
3	359.41	Good	0.059	Model fit	0.017	Good Fit
4	9.185	Good	0.24	Model fit	0.026	Good Fit

CFA analysis of each factor of the NST is needed to determine the internal validity of each factor. Results from the initial analysis on all subtests indicate that the model for each factor is not fit with p-value <0.05 and RMSEA> 0.05. After modifying by allowing errors in each item to be correlated with one another, a fit model is obtained with a p-value > 0.05, as seen in Table 10. Thus, these results indicate that all items in each factor are proven to measure only one thing, the subtest itself. The significance of the items on each factor is measured by looking at the z-value and p-value on each item, as seen in Table 11.

Tabel 11. Items Insignificance

Factor	Insignificant Items	Percentage	Item Number
1	0	0%	-
2	0	0%	-
3	14	45%	F3.2, F3.3, F3.5, F3.8, F3.10, F3.13, F3.14, F3.16, F3.17, F3.20, F3.26, F3.27, F3.29, F3.31
4	1	12.5%	F4.8

In factor 4 model, only 1 item was found insignificant. In factor 3, 50% of the items are significant, and the remaining 45% are insignificant. Further review for insignificant items is needed to reconsider whether the item is included in the model or not.

4. Discussion

This study showed that all items in the NST are valid based on item quality analysis 3 PL model IRT, which describes 3 parameters; a/discrimination, b/difficulty, and c/pseudo-guessing. In general, the level of difficulty of the NST items spread from easy to difficult and qualified to be used as a test to measure children's cognitive maturity. However, it is necessary to review several items, especially in subtest 10, with 87.5% of the discrimination index in the abnormal criteria.

EFA analysis results showed that 4 factors were found to be significant in the NST construct, where several former subtests are grouped in the same factor as they measure the same construct. Based on the analysis, all factors showed factor loading >0.3 where factor 1 consists of subtest 1, 3, 4, 6, and 8, factor 2 consists of subtest 2, factor 3 consists of subtest 7 and 9, and factor 4 consists only of subtest 10. Factor 1 grouped several subtests that have the same latent construct, logic reasoning which is usually found in fluid intelligence concepts. Ability to distinguish the shape; ability to understand the size, number, and comparison; ability to observe in detail; ability to concentrate; andunderstandg the situation and object were what Cattel and Horn mentioned as the inherited problem-solving capacity that involves the ability to see new relationships or educe relations in a logical manner (Kent, 2017). Fluid intelligence is the children's ability to solve new problems without using previous experience or knowledge that they have from the academic learning process (Özcan et al., 2021).

Meanwhile, factor 2 which only consists of subtest 2 does measure a typical area that is related specifically to fine motor development. As mentioned in previous research, two types of fine motor skills, i.e fine motor object manipulation and fine motor writing, predict academic achievement beyond language ability and other cognitive skills (Dinehart & Manfra, 2013). This previous study also reported that fine motor skill gives a significant effect on math and reading achievement.

Factor 3 consists of two subtests (subtest 7; memorization ability and subtest 9; memorization in the narrative) with a similar construct related to the role of memory ability. Although there is a high correlation between memory and fluid intelligence, several studies have shown that memory is not included in fluid intelligence but is more to crystalize intelligence that children needed during the learning process at school (Shipstead et al., 2016). It is suggested that memory intervention programs that start from 4 years old class can have lasting positive academic effects (Diamond & Lee, 2011).

The last factor, factor 4 consists only of subtest 10, which measures children's capacity to observe information from the environment and represented relevant information by drawing the human figure. This concept of cognitive capacity is similar to the idea proposed by Goodenough, where Goodenough believes that the construction and way children execute human drawings is a reflection of their knowledge and intelligence maturity in processing information and expressing used conceptually (Plubrukarn & Theeramanoparp, 2003). From this new NST construct, the practitioners or researchers can carry out the interpretation process of the NST test result based on proper and robust information. The discovery of 4 new factors that construct the NST raises further questions about the correlation between IQ score and each of the NST factors, considering that school readiness and intelligence have a significant influence on children's academic conditions (Izzaty et al., 2017). Previous research showed that intelligence scores from CPM have a positive correlation with the total score of the NST (Mariyati, 2019).

To ensure that the four factors of the NST do measure the same latent construct, the CFA was again carried out for the construct validity test. As observed from the analysis result, data showed that the four new factors of the NST measure the same latent construct, children's cognitive maturity. It can be said that this measuring instrument is valid for measuring children's cognitive maturity as part of school readiness indicators. To make reliable and valid interpretations of children's assessments, pa psychologists can now use these new four factors as a basis for interpreting data from the NST.

5. Conclusion

Psychometric properties of the NST were examined and this study showed most of the items were qualified and able to distinguish cognitive development between subjects. However, some items needed careful attention and were necessary review, especially subtest 10, with 87.5% of the discrimination index in the abnormal criteria. In construct validity examination with EFA and CFA, the NST consists of four new factors (logic reasoning ability, memory, fine motor skill, and conceptual maturity), which these four factors measured the same latent construct; children's cognitive maturity.

6. References

Cuskelly, M., & Detering, N. (2003). Teacher and Student Teacher Perspectives of School Readiness. *Australasian Journal of Early Childhood*. https://doi.org/10.1177/183693910302800208

Decoster, J., & Hall, G. P. (1998). Overview of Factor Analysis. *In Practice*. https://doi.org/10.2307/2685875

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. In *Science*. https://doi.org/10.1126/science.1204529

Dinehart, L., & Manfra, L. (2013). Associations Between Low-Income Children's Fine Motor Skills in Preschool and Academic Performance in Second Grade. *Early Education and Development*. https://doi.org/10.1080/10409289.2011.636729

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12046

Guler, N., Uyanik, G. K., & Teker, G. T. (2014). Comparison of Classical Test Theory and Item Response Theory in Terms of Item Parameters. *Internatinal Association of Social Science Research*.

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, Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis. In *Vectors*. https://doi.org/10.1016/j.ijpharm.2011.02.019

Hambleton, R. K., Swaminathan, H., & Rogers, H. (1991). Measurement methods for the social sciences series, Vol. 2.Fundamentals of item response theory. In *Sage Publications, Inc.*

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. In *Organizational Research Methods*. https://doi.org/10.1177/1094428104263675

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. In *Journal of Organizational Behavior*. https://doi.org/10.1002/(SICI)1099-1379(199711)18:6<667::AID-JOB874>3.0.CO;2-T

Izzaty, R. E., Cholimah, N., & Wulandari, R. (2017). Pengembangan Buku Cerita Tematik Sebagai Media Pembelajaran Pengenalan Membaca Pada Anak Prasekolah. *Jurnal Pendidikan Anak*. Https://doi.org/10.21831/Jpa.V3i2.11704

Jacob, B., & Rothstein, J. (2016). The measurement of student ability in modern assessment systems. *Journal of Economic Perspectives*. https://doi.org/10.1257/jep.30.3.85

Kent, P. (2017). Fluid intelligence: A brief history. *Applied Neuropsychology: Child*. https://doi.org/10.1080/21622965.2017.1317480

Mariyati, L. I. (2019). Inteligensi Dan Kesiapan Anak Masuk Sekolah Dasar. *Psyche:Jurnal Psikolog*.

Mariyati, L. I., & Affandi, G. R. (2016). Tepatkah Nijmeegse Schoolbekwaamheids Test (Nst) Untuk Mengukur Kesiapan Sekolah Siswa Sekolah Dasar Awal Pada Konteks Indonesia? (Analisis Empirik Berdasar Teori Tes Klasik). *Jurnal Ilmiah Psikologi Terapan*, *4*(2).

Monks, F. J., Rost, H., & Coffie, N. H. (1978). Nijmeegse Schoolbekwaamheids Test.

Özcan, M. Ş., Çetinkaya, E., Göksun, T., & Kisbu-Sakarya, Y. (2021). Does learning to code influence cognitive skills of elementary school children? Findings from a randomized experiment. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.12429

Papalia, D. E., Olds, S. W., & Feldman, R. D. (2010). *Human Development : Psikologi Perkembangan bagian I s/d IV*.

Pauls, F., & Daseking, M. (2021). Revisiting the Factor Structure of the German WISC-V for Clinical Interpretability: An Exploratory and Confirmatory Approach on the 10 Primary Subtests. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2021.710929

Plubrukarn, R., & Theeramanoparp, S. (2003). Human Figure Drawing Test: Validity in Assessing Intelligence in Children Aged 3-10 Years. *Journal of the Medical Association of Thailand*.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691616650647

ter Laak, J. J. F. (1991). Presuppositions and relation of cognitive functions in a schoolreadiness curriculum. *European Journal of Psychology of Education*. https://doi.org/10.1007/BF03173143

Usman, & Sobari. (2013). *Aplikasi Teknik Multivariate*. Rajawali Pers.

Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*. https://doi.org/10.1177/0095798418771807

Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry*. https://doi.org/10.3969/j.issn.1002-0829.2014.03