Journal of Educational Research and Evaluation
Volume 5, Issue 3, 2021, pp. 429-435
P-ISSN: 2597-422x E-ISSN: 2549-2675
Open Access: https://ejournal.undiksha.ac.id/index.php/JERE

The Test of Construct Validity for One-Factor Model

Yahfizham[1], Irwan Yusti[2], Muhammad Luthfi Hamzah[3]

[1] UIN Sumatera Utara, Medan, Indonesia
[2] STTIND, Padang, Indonesia
[3] UIN Suska Riau, Pekan Baru, Indonesia

ARTICLE INFO

Article history:
Received March 25, 2021
Revised March 29, 2021
Accepted July 30, 2021
Available online August 25, 2021

Kata Kunci :
Validitas Konstruk, Model One-Factor, CBT

Keywords:
Construct Validity, One-Factor Model, CBT

ABSTRAK

Tes dalam bidang pendidikan merupakan alat ukur yang digunakan oleh pendidik dan untuk menunjang proses penilaian kemampuan peserta didik. Namun, semua item yang terkandung dalam instrumen perlu diuji. Penelitian ini bertujuan untuk menguji validitas konstruk instrumen angket Cognitive Abilities (CA) dengan model Computer-Based Test (CBT). Sampel penelitian artikel ini terdiri dari 31 mahasiswa. Metode yang digunakan untuk mengumpulkan data yaitu kuesioner. Instrument yang digunakan dalam mengumpulkan data yaitu angket. Penilaian instrumen angket dilakukan oleh lima orang ahli yang diundang sebagai validator dalam proses berlangsungnya teknik Focus Group Discussion (FGD). Pengecekan validitas konstruk kuesioner dilakukan dengan analisis faktor konfirmatori (CFA) menggunakan LISREL 9.30, dengan sampel kecil. Hasil tersebut menegaskan bahwa nilai aproksimasi fit (chi-square) = 0.00 0.03, degree of freedom (df) = 0 2.00, Root Mean Square Error of Approximation (RMSEA) = 0,000 0,08 dan p-value 1.000.000 > 0,80. Berdasarkan hasil, menunjukkan kecocokan data yang lebih baik dengan model satu faktor yang berkorelasi atau item tunggal atau kuesioner (isi, konstruksi, dan gaya bahasa). Pandangan bahwa validitas konstruk lebih menekankan pada seberapa jauh angket yang disusun berkaitan dengan pengukuran secara teoritis terhadap konsep-konsep yang telah disusun dengan meminta pertimbangan ahli.

ABSTRACT

The test in the field of education is a measuring tool used by educators and to support the process of assessing the abilities of students. However, all items contained in the instrument need to be tested. This study aims to test the construct validity of the Cognitive Abilities (CA) questionnaire instrument with the Computer-Based Test (CBT) model. The research sample of this study consists of 31 students. The method used to collect data is a questionnaire. The instrument used in collecting data is a questionnaire. The assessment of the questionnaire instrument was carried out by five experts who were invited as validators in the process of the Focus Group Discussion (FGD) technique. Checking the construct validity of the questionnaire was carried out by confirmatory factor analysis (CFA) using LISREL 9.30, with a small sample. The result confirmed that the value of approximate fit (chi-square) = 0.00 ≤ 0.03, degree of freedom (df) = 0 ≤ 2.00, the Root Mean Square Error of Approximation (RMSEA) = 0.000 ≤ 0.08 and p-value 1.000.000 > 0.80. Based on the results, the data shows a better fit with the correlated one-factor model or single item or questionnaire (content, construction, and language style). The point of view that construct validity emphasizes more on how far the questionnaire compiled is related to the theoretical measurement of the concepts that have been prepared by asking for expert judgment.

1. INTRODUCTION

The cognitive abilities are soft skills in the proses of knowledge to evaluation by Bloom. The cognitive abilities (intellectual tendency) in the applied, social, or behavioral sciences for many researchers is frequently regarding studying theoretical constructs that cannot be observed directly, as an example, in education is cognitive abilities. In order to measure cognitive skills, it can be done through the test (Duckworth & Yeager, 2015; Berger & Karabenick, 2016). The test in the field of education is a measuring tool used by educators in making judgments or assessing the learners. The test is conducted to support the process of assessing the abilities of students (Bromberek-Dyzman et al., 2021; Mastroleo et al., 2020; Roediger et al., 2011). Pragmatically, in the era digital literacy industry 4.0 revolution, CBT model is defeating the popularity of the paper-pencil based test (PPT) models (Boevé et al., 2015; Jeong, 2014; Khoshsima et al., 2017; Maier et al., 2016; Peters, 2017). The concept of assessment in the learning becomes a topic of cognitive abilities tendency using CBT model (Kaya & Leite, 2017; Lin et al, 2018). For educators, the CA-CBT model is used to classify or mapping the cognitive abilities of the students. Educators can find the weaknesses and strengthness of each learner so that it will be beneficial for the improvement strategies, methods, and techniques of learning (Moore & Miller, 2018). However, All items contained in instruments need to be tested. In theory and practice, the instruments used in the development of a model can do with

*Corresponding author.
E-mail addresses: yahfizham@uinsu.ac.id (Yahfizham)

a construct validity approach (Han & Han, 2016; Hayton et al., 2012). The process of accepting or rejecting the test model can be done through construct validity by (Marsh, Morin, Parker & Kaur, 2014; Perry, Nicholls, Clough & Crust, 2014). Empirically, construct validity is the main issue to establish the feasibility of the conceptual model proposed as stated (Calvo et al., 2019; Neill et al., 2016; Rochefort et al., 2018; Sellbom et al., 2015; Tamboer & Vorst, 2015; Taylor et al., 2018; Wong et al., 2013). Based on the point of view, construct validity is a technique to test the correctness of a designed measuring instrument (Li, 2016; Orr et al, 2017). Construct validity is testing some variables that correlate typical details of the indicators contained in a measuring instrument.

The CA-CBT model is the test tool commonly used by educator to administer several questions in expressing mastery or non-mastery cognitive abilities to the learner (Huebner, Finkelman & Weissman, 2018; Kaplan, de la Torre & Barrada, 2015; Ma & de la Torre, 2016). It's interesting to develop the test as a tool to measure the cognitive abilities of students (*1* , 1 , 2 1*, n.d.; Alla Belousova, 2015). The CA-CBT with the multiple-choice item quality question approach is carried out as a requirement to establish the construct validity (Blikstein, Kabayadondo, Martin & Fields, 2017). The proof of the quality of the CA-CBT model can be done with the item test (Lin & Chang, 2019; Huang, 2019) and construct validity. Another opinion belongs by Burcu, Alexander & Ng (2015) and Krumm et al (2016) stated that the construct validity can be done using the CFA approach with single items or a one-factor model. CFA is a technique for the test conformity between the latent variables and observed variables (Brigman et al, 2015; Cokley, 2015; Guille, Arias, Vicente & Badia, 2016). In theory and practice, the instruments used in the development of a model can do with a construct validity approach (Shuck, Adelson & Reio Jr, 2016). The process of accepting or rejecting the test model can be done through construct validity by (Marsh, Morin, Parker & Kaur, 2014; Perry, Nicholls, Clough & Crust, 2014). Empirically, construct validity is the main issue to establish the feasibility of the conceptual model proposed as stated (Neill et al., 2016; Rochefort et al., 2018; Sellbom et al., 2015; Tamboer & Vorst, 2015). Based on the point of view by (Li, 2016; Orr et al, 2017), construct validity is a technique to test the suitability between what is to be measured and the tools used to measure it.

Based on the opinion, construct validity is testing some variables that correlate typical details of the indicators contained in a measuring instrument (Richardson et al, 2016). CFA is typically evaluated on the basis of goodness of-fit indexes (Marsh, Morin, Parker & Kaur, 2014). In the face of their popularity, according what worth these indexes should reach to confidently specify between the acceptance and rejection of a model has been greatly debated (Lai, 2019; Brenner, Heath, Vogel, & Credé, 2017; Marcoulides & Yuan, 2017). Many experts argue that, using CFA to test the questionnaire instrument and calculated model fit indexes with 31 students (respondents), is very small size of sample for the model construction (Prudon, 2015; Hoofs et al, 2018; Lewis, 2017). However, (Wolf, Harrington, Clark, Miller, 2013; Jackson, Voth, & Frey, 2013; Ory, & Mokhtarian, 2010) have conducted studies with a small sample using CFA. Therefore, the main issue of this article is to evaluate the construct validity of the CA-CBT model to measure students cognitive abilities with the one-factor model, while CFA were used for verification. This article attempt to examining of the construct validity of the instrument questionnaire CA-CBT, using one-factor model. The research objective of this article was to make out that one validation instrument sheet containing (form) 23 statements with 31 respondents was able to meet the CFA aspect criterion (valid). This article purpose to examining the construct validity of the instrument questionnaire Cognitive Abilities (CA) with the Computer-Based Test (CBT) model.

2. METHODS

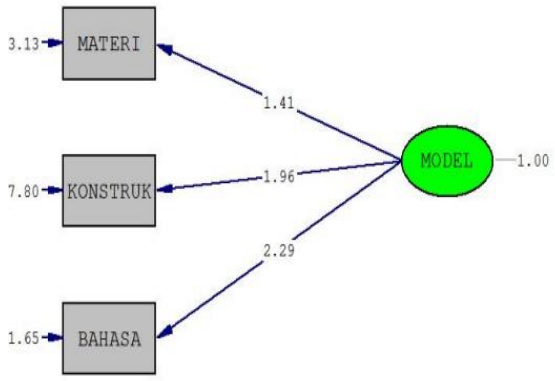
The sample of this article research consisted of 31 students from Universitas Islam Negeri Sumatera Utara, Medan, Indonesia. The students agreed and filled out the questionnaires completely. The questionnaire instrument was used to obtain data with items that rated on a 4-point Likert scale, covering from 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree. The sample of the brief description variable X_1 = content (item$_1$ to item$_7$), X_2 = construction (item$_8$ to item$_{17}$), and X_3 language style (item$_{18}$ to item$_{23}$) as the indicators. The assessment of the questionnaire instrument was carried out by five experts invited as validators in the ongoing process of the Focus Group Discussion (FGD) technique. The experts involved in various fields such as lecturers who teach the subject itself, assessment, evaluation, information system, and Indonesian language. The feature of instrument questionnaire was also observed through their empirical construct validity of indicators to show the level to which each indication a general latent construct. The CFA was used to assess the factor structure of the one-factor model with LISREL 9.3 student version to conduct our one-factor model analysis. The one-factor model was constructed with the criterion features such as each of the items is not equal to zero loading on both the general point and the group factor

that is outlined to measure but zero loadings on the other group factors. The one-factor model was fully saturated, so goodness-of-fit indices interpreted, and only factor loadings reported.

3. RESULT AND DISCUSSION

Results

This research was done quantitatively derived from the primary data collected by the written method to the respondents in the form of questionnaires, the sincerity of respondents giving answers essentially. The reliabilities of the questionnaire used in this research was depend on the construct validity of the CA-CBT model. The questionnaire was developed based on content, construct, and language style. The testimony of the validity should be carried out on items that have medium and high validity criteria, while for low efficacy, its discarded. The result of the analysis using Aiken's showed that items 14 and 15 belong to medium category with 0.70 and 0.75, and the overmeasure of the things were in the high class. The results of the validity test with the construct validity, with evaluating the goodness-of-fit model of the CA-CBT model was by performing a CFA. The standardized solution for the one-factor model of the CA-CBT (see Picture 1) is set out below.



Picture 1. One-factor model

The standard error variances estimates factor correlations in the one-factor model reported by default with minimum = 1.645. Also, all indicators of the variance, specifically R^2 values was ranged between 0.329 to 0.760, which indicates that the model describes the noticed correlations well. All of the loading factors forming the test models have met the requirements of a fit model, which obtained 1.41, 1.96, and 2.29 more significant than 0.5. The following table 3 shows the results of processing the LISREL data with CFA, which shows that all forming variables to the test the CA-CBT model meet the criteria of a good model. The analysis disclosed that the model is saturated; the fit is perfect — the goodness-of-fit indices of the CA-CBT model from the one-factor model presented in Table 3, and Table 4. Next (see Table 3), we elucidated maximum likelihood estimation (MLE) of CA-CBT model variable X_1, X_2, and X_3 ($item_1$ to $item_{23}$) with the standardized factor loadings, standard error var, and R^2. All levels of fitness between the model and the data with the standardized factor loadings were significant ($\lambda > 0.3$), this indicates that the CA-CBT model has met the requirements of a fit.

The results displayed an insufficient fit of the model to the data. The CFA was exercised to analysis the factor structure of the hypothesized one-factor model. Since the one-factor model is saturated, the goodness of fit indices was illustrated, and factor loadings revealed (see Figure 2). The values of chi-square (X^2) = -0.00 ≤ 0.03, df = 0 < 2.00, RMSEA = 0.000 ≤ 0.08 and p-value = 1.00000 > 0.80. Overall, the endorse of statistical fit for a unidimensional model. The unidimensional model is used to represent a specific kind of measurement scale. CA-CBT model should be designed to measure only cognitive abilities by narrowing the idea of a single construct. The results appointed that this brief, arrays good construct validity, reciting to measures of cognitive abilities. The indicators of the construct validity of the CA-CBT model is shown in Table 5. Total variance = 23.624, generalized of variance = 213.446, largest eigenvalue = 15.849, smallest eigenvalue = 2.605, and the condition number = 2.467 with number of iterations = 0. The CA-CBT model, positively correlated between all variables and items. On our small sample of the respondents and related instruments that measure the raised variables, to investigate or examine the construct validity of the cognitive abilities, values of the statistically significant correlations of the CA-CBT model with the CFA one-factor model. The one-factor model based on the supposition of multivariate normality for the observed data, which implies a specific assumption for the parameter of the factor loading. The results of the construct

validity test on the CA-CBT model instrument using the CFA approach revealed that all items are unidimensional or in other words, only measure one-factor loading. It concluded that the one-factor model theorized by the CA-CBT model instrument is acceptable to meet the criterion based on table 6. This general distribution belief implies one specific measurement of the tool used. This research brings out the issue of construct validity the test on CA-CBT model with a single device to use this model.

Discussion

To our proficiency, this is the first research to judge the construct validity of the CA-CBT model to measure the students cognitive abilities using one-factor model. The CA-CBT model is provided based on the marginal maximum likelihood estimation to enable an exact test of this assumption by the one-factor model with factor loadings. In CFA, there are latent variables and indicator forming variables. The latent variables are variables that directly observed because they are abstract, while signs create variables that observed and measured directly. The CA-CBT model was carried out for theoretical class subjects whose questions structured according to Bloom's cognitive taxonomy model, which focuses on aspects of the mastery of the material or subject matter. The Bloom's has separated soft skills in two classifications of the cognitive abilities (Sukajaya, Purnama, & Purnomo, 2015; Ramirez 2016). To get information on the level of cognitive abilities skills of the students on the material that has been presented, what has been usually done are PPT. The PPT very popular among lecturers/teachers/ educators before computer technology developed to measured low order thinking (LOT) abilities. The second is the abilities to think at a higher order thinking (HOT), which is the abilities to use understood theory, able to sort out public information, ready to provide synthesis results from data, and be able to rearrange all data that has been owned evaluation (Heong et al., 2011; Rapih & Sutaryadi, 2018; Widana, 2017). The identification of mastery of learning can be made with a variety of methods and techniques that can take, namely by testing diagnostic thinking skills assisted by computers as a tool before conducting an assessment of learning outcomes (Maddocks, 2018; Marques-Costa, Almiro & Simoes, 2018).

The point of view by that Administering CBT according to various research results that have been reported in many journal articles, both national and international, have the same level of validity, even better than the traditional administration of the test, which are PPT and CBT. It's interpretes that the process of supporting the assessment of cognitive abilities can be carried out involving information communication and technologies such as computer. The various studies on PPT with CBT have been carried out which revealed that the computer-based test have the same results or even better than PPT (Clariana, & Wallace, 2002; Yulianto, Wahyuni, & Eka, 2016; Septanto, 2018; Marastuti et al, 2020). Many advantages are obtained such as the speed of time to check the results of the exam participant's answers, the speed to determine the ability of the examinees without having to answer all the questions, the questions that are displayed are easy to randomize, reduce cheating and so on. The aspects of anxiety, habits, test atmosphere and form question safety were not the constraints. But, the validity of a measurement tool has several essential meanings. The accuracy of the interpretation of the test results, degrees that indicate categories that can include low, medium and high groups. The principle of a valid is not universal the test (Kane, 2015; Ouyang, Xin, & Chen,2016).

This issue needs further research because the standards for educational and psychological testing guidelines require that any effects due to computer administration be either eliminated or accounted for in the interpretation of test scores in any testing mode comparability study (Huseyin & Ozturan, 2018; Sorrel et al, 2016; Nikou, & Economides, 2016). Item analysis methods that are usually carried out on the two test models as classical and modern test theory like Rasch model and Item Response Theory (IRT) by including item difficulty and item discrimination index using the item fit index, and internal structure as evidence of construct validity. Thus, construct validity can be interpreted as validity judging from the terms of the structure, context, or fabrication. In terms of terminology, a learning outcome teststated as a test that has construction validity if the proof of learning outcomes reviewed in terms of composition, framework, or imagery can accurately reflect a construction in psychological theory. The validity of the development concept of an argument, is the basis for the preparation of tools (Kelava 2016; Hoogland & Tout, 2018)

Therefore, there must be a discussion about the theory, which is the basis for determining the construction of the instrument. The validity of construction is meant that the test concerned seen as having a proper sentence structure, or the sequence of the number of questions is coherent, but that the new learning outcome test can be said to have construction validity if the items that construct the test are correct and able to precisely measure aspects of thinking as determined in the purpose of carrying out the analysis. The CFA approach can prove that the construct validity of the instrument exists to presented empirically. CFA explores empirical data so that it finds characteristics and relationships between endogenous variables and exogenous variables without specifying the model in the data. Finally, some limitations of the attendance of this study need to be introduced. First, because the sample of this research is an appliance

sample from the students receiving from single-center setting or one university. It is not viable to popularized the findings directly to the other general. Second, the three variables and twenty-three indicators still need further development. Third, because of practical reasons, we did not get the measure of the test-retest reliabilities surrounded by respondents. Interestingly, a statistically significant correlation of the CA-CBT model with the cognitive abilities disclosed empirically.

4. CONCLUSION

The results of the CFA analysis shows that there is no second dimension in the testing of the CA-CBT model, this means that all indicators accepted. The research designates that the instrument (questionnaire) of the CA-CBT model is valid and may be used as a brief measure or to the support assess of the students cognitive abilities process of learning. Based on the CFA one-factor model, the CA-CBT model was perfect saturated, or one-factor loadings significant. However, the further study evaluated the psychometric belongings of the CA-CBT model in various populations and large sample, and it's test-retest reliabilities.

5. REFERENCES

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419.

Alla Belousova, V. P. (2015). Technique of Thinking Style Evaluating. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, *3*(2), 1–8. https://doi.org/10.23947/2334-8496-2015-3-2-1-8.

Berger, J. L., & Karabenick, S. A. (2016). Construct Validity of Self-Reported Metacognitive Learning Strategies. *Educational Assessment*, *21*(1), 19–33. https://doi.org/10.1080/10627197.2015.1127751.

Blikstein, P., Kabayadondo, Z., Martin, A., & Fields, D. (2017). *An Assessment Instrument of Technological Literacies in Makerspaces and FabLabs*. *106*(1), 149–175. https://doi.org/10.1002/jee.20156.

Boevé, A. J., Meijer, R. R., Albers, C. J., Beetsma, Y., & Bosker, R. J. (2015). *Introducing Computer-Based Testing in High- Stakes Exams in Higher Education : Results of a Field Experiment*. 1–13. https://doi.org/10.1371/journal.pone.0143616.

Brigman, G., Wells, C., Webb, L., Villares, E., Carey, J. C., & Harrington, K. (2015). Psychometric properties and confirmatory factor analysis of the student engagement in school success skills. *Measurement and Evaluation in Counseling and Development*, *48*(1), 3–14. https://doi.org/10.1177/0748175614544545.

Bromberek-Dyzman, K., Jankowiak, K., & Chełminiak, P. (2021). Modality matters: Testing bilingual irony comprehension in the textual, auditory, and audio-visual modality. *Journal of Pragmatics*, *180*. https://doi.org/10.1016/j.pragma.2021.05.007.

Burcu, M., Alexander, G. C., & Ng, X. (2015). Construct Validity and Factor Structure of Survey-based Assessment of Cost-related Medication Burden. *National Library of Medicine*, *53*(2), 199–206. https://doi.org/10.1097/MLR.0000000000000286.

Calvo, S., Morales, A., & Wade, J. (2019). The use of MOOCs in social enterprise education: an evaluation of a North–South collaborative FutureLearn program. *Journal of Small Business and Entrepreneurship*, *31*(3), 201–223. https://doi.org/10.1080/08276331.2018.1453241.

Cokley, K. (2015). A confirmatory factor analysis of the Academic Motivation Scale with Black college students. *Measurement and Evaluation in Counseling and Development*, *48*(2), 124–139. https://doi.org/10.1177/0748175614563316.

Duckworth, A. L., & Yeager, D. S. (2015). *Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes*. *May*, 237–251. https://doi.org/10.3102/0013189X15584327.

Guille, M., Arias, B., Vicente, E., Verdugo, M. A., & Badia, M. (2016). *Research in Developmental Disabilities Confirmatory factor analysis of the supports intensity scale for children*. *50*, 140–152. https://doi.org/10.1016/j.ridd.2015.11.022.

Han, B., & Han, B. (2016). Social Media Burnout : Definition , Measurement Instrument , and Why We Care Social Media Burnout : Definition , Measurement Instrument , and Why We Care. *Journal of Computer Information Systems*, *00*(00), 1–9. https://doi.org/10.1080/08874417.2016.1208064.

Hayton, J. C., Biron, M., & Christiansen, L. C. (2012). *Book review*. *51*(3), 457–459. https://doi.org/10.1002/hrm.

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. Bin, & Mohamad, M. M. B. (2011). The

Level of Marzano Higher Order Thinking Skills among Technical Education Students. *International Journal of Social Science and Humanity*, *1*(2), 121–125. https://doi.org/10.7763/ijssh.2011.v1.20.

Huebner, A., Finkelman, M., & Weissman, A. (2018). Factors Affecting the Classification Accuracy and Average Length of a Variable-Length Cognitive Diagnostic Computerized Test. *Journal of Computerized Adaptive Testing*, *6*(1), 1–14. https://doi.org/10.7333/1802-060101.

Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour and Information Technology*, *33*(4), 410–422. https://doi.org/10.1080/0144929X.2012.710647.

Kane, M. T. (2015). *Explicating validity Explicating validity*. *December*. https://doi.org/10.1080/0969594X.2015.1060192.

Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New Item Selection Methods for Cognitive Diagnosis Computerized Adaptive Testing. *Applied Psychological Measurement*, *39*(3), 167–188. https://doi.org/10.1177/0146621614554650.

Kaya, Y., & Leite, W. L. (2017). Assessing Change in Latent Skills Across Time With Longitudinal Cognitive Diagnosis Modeling: An Evaluation of Model Performance. *Educational and Psychological Measurement*, *77*(3), 369–388. https://doi.org/10.1177/0013164416659314.

Kelava, A. (2016). *A Review of Confirmatory Factor Analysis for Applied Research (Second Edition)*. *XX*(X), 1–5. https://doi.org/10.3102/1076998616631747.

Khoshsima, H., Morteza, S., Toroujeni, H., & Tefl, M. A. I. (2017). Comparability of Computer-Based Testing and Paper-Based Testing: Testing Mode Effect, Testing Mode Order, Computer Attitudes and Testing Mode preference. *International Journal of Computer (IJC) International Journal of Computer (IJC*, *24*(1), 80–99.

Krumm, G., Arán Filipppetti, V., Lemos, V., Koval, J., & Balabanian, C. (2016). Construct validity and factorial invariance across sex of the Torrance Test of Creative Thinking – Figural Form A in Spanish-speaking children. *Thinking Skills and Creativity*, *22*, 180–189. https://doi.org/10.1016/j.tsc.2016.10.003.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7.

Lin, Y., Jiang, Y., Gong, Y., Zhan, Z., & Zhang, J. (2018). A Discrete Multiobjective Particle Swarm Optimizer for Automated Assembly of Parallel Cognitive Diagnosis Tests. *IEEE Transactions on Cybernetics*, *PP*, 1–14. https://doi.org/10.1109/TCYB.2018.2836388.

Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *The British Journal of Mathematical and Statistical Psychology*, *69*(3), 253–275. https://doi.org/10.1111/bmsp.12070.

Maddocks, D. L. S. (2018). *The Identification of Students Who Are Gifted and Have a Learning Disability : A Comparison of Different Diagnostic Criteria*. https://doi.org/10.1177/0016986217752096.

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers and Education*, *95*, 85–98. https://doi.org/10.1016/j.compedu.2015.12.002.

Marsh, H. W., Morin, A. J. S., Parker, P., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis. *Ssrn*, *November*, 1–26. https://doi.org/10.1146/annurev-clinpsy-032813-153700.

Mastroleo, N. R., Humm, L., Williams, C. M., & Kiluk, B. D. (2020). Initial Testing of a Computer-Based Simulation Training Module to Support Clinicians' Acquisition of CBT Skills for Substance use Disorder Treatment. *Journal of Substance Abuse Treatment*, 114. https://doi.org/10.1016/j.jsat.2020.108014.

Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). the Use of Rasch Measurement Model in English Testing. *Jurnal Cakrawala Pendidikan*, *38*(1), 16–32. https://doi.org/10.21831/cp.v38i1.22750.

Neill, T. A. O., Lewis, R. J., Law, S. J., Larson, N., Hancock, S., Radan, J., Lee, N., & Carswell, J. J. (2016). Forced-choice pre-employment personality assessment : Construct validity and resistance to faking. *PAID*. https://doi.org/10.1016/j.paid.2016.03.075.

Nyoman Sukajaya, I., Ketut Eddy Purnama, I., & Purnomo, M. H. (2015). Intelligent classification of learner's cognitive domain using bayes net, naïve bayes, and j48 utilizing bloom's taxonomy-based serious game. *International Journal of Emerging Technologies in Learning*, *10*(2), 46–52. https://doi.org/10.3991/ijet.v10i1.4451.

Orr, M. T., Pecheone, R., Hollingworth, L., Beaudin, B., Snyder, J., & Murphy, J. (2017). *The Performance Assessment for Leaders : Construct Validity and Reliability Evidence*. https://doi.org/10.1177/1942775117742646.

Ouyang, X., Xin, T., & Chen, F. (2016). *Construct Validity of the Children ' s Coping Strategies Scale (CCSS): A*

Bifactor Model Approach. https://doi.org/10.1177/0033294116628362.

Perry, J. L., Nicholls, A. R., Clough, P. J., Crust, L., & Perry, J. L. (2014). *Measurement in Physical Education and Exercise Science Assessing Model Fit : Caveats and Recommendations for Confirmatory Factor Analysis and Exploratory Structural Equation Modeling Assessing Model Fit : Caveats and Recommendations for Con fi rmatory Fac*. *February 2015*, 37–41. https://doi.org/10.1080/1091367X.2014.952370.

Peters, M. A. (2017). Technological unemployment: Educating for the fourth industrial revolution. *Educational Philosophy and Theory*, *49*(1), 1–6. https://doi.org/10.1080/00131857.2016.1177412.

Ramirez, T. V. (2016). *On Pedagogy of Personality Assessment : Application of Bloom ' s Taxonomy of Educational Objectives On Pedagogy of Personality Assessment : Application of Bloom ' s Taxonomy of*. *3891*(June). https://doi.org/10.1080/00223891.2016.1167059.

Rapih, S., & Sutaryadi, S. (2018). Perpektif guru sekolah dasar terhadap Higher Order Tinking Skills (HOTS): pemahaman, penerapan dan hambatan. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, *8*(1), 78. https://doi.org/10.25273/pe.v8i1.2560.

Richardson, J., Iezzi, A., Khan, M. A., Chen, G., Maxwell, A., & Hons, B. (2016). *Measuring the Sensitivity and Construct Validity of 6 Utility Instruments in 7 Disease Areas*. 1–13. https://doi.org/10.1177/0272989X15613522.

Rochefort, C., Baldwin, A. S., & Chmielewski, M. (2018). Experiential Avoidance: An Examination of the Construct Validity of the AAQ-II and MEAQ. In *Behavior Therapy* (Vol. 49, Issue 3). Elsevier Ltd. https://doi.org/10.1016/j.beth.2017.08.008.

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten Benefits of Testing and Their Applications to Educational Practice. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 55). Elsevier Inc. https://doi.org/10.1016/B978-0-12-387691-1.00001-6.

Sellbom, M., Cooke, D. J., Hart, S. D., Sellbom, M., Cooke, D. J., Hart, S. D., Validity, C., Sellbom, M., Cooke, D. J., & Hart, S. D. (2015). *Construct Validity of the Comprehensive Assessment of Psychopathic Personality (CAPP) Concept Map : Getting Closer to the Core of Psychopathy Construct Validity of the Comprehensive Assessment of Psychopathic Personality (CAPP) Concept Map : Getting C*. *9013*(October). https://doi.org/10.1080/14999013.2015.1085112.

Tamboer, P., & Vorst, H. C. M. (2015). A new self-report inventory of dyslexia for students: criterion and construct validity. *Dyslexia (Chichester, England)*, *21*(1), 1–34. https://doi.org/10.1002/dys.1492.

Taylor, S., Girdler, S., Parsons, R., McLean, B., Falkmer, T., Carey, L., Blair, E., & Elliott, C. (2018). Construct validity and responsiveness of the functional Tactile Object Recognition Test for children with cerebral palsy. *Australian Occupational Therapy Journal*, *65*(5), 420–430. https://doi.org/10.1111/1440-1630.12508.

Widana, I. W. (2017). Higher Order Thinking Skills Assessment (HOTS). *JISAE (Journal of Indonesian Student Assessment and Evaluation)*, *3*(1), 32–44. https://doi.org/10.21009/JISAE.031.04.

Wong, I. H., Denkers, M., Urquhart, N., & Farrokhyar, F. (2013). *Construct validity testing of the Arthroscopic Knot Trainer (ArK)*. https://doi.org/10.1007/s00167-013-2524-x.