International Journal of Evaluation and Research in Education (IJERE)
Vol. 12, No. 1, March 2023, pp. 357~368
ISSN: 2252-8822, DOI: 10.11591/ijere.v12i1.22736

The instrument development to measure the verbal ability of prospective high school students

Muhammad Rais Ridwan[1,2], Samsul Hadi[3], Jailani Jailani[4], Heri Retnawati[4]
[1]Study Program of Research and Evaluation Education, Graduate School, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
[2]Study Program of Mathematics Education, STKIP YPUP Makassar, Makassar, Indonesia
[3]Department of Electrical Engineering Education, Faculty of Engineering, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
[4]Department of Mathematics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

Article Info

Article history:

Received Aug 18, 2021
Revised Sep 6, 2022
Accepted Oct 2, 2022

Keywords:

Composite reliability
Confirmatory factor analysis
Construct validity and reliability
Scholastic aptitude test
Verbal ability test

ABSTRACT

An alternative for determining an accurate major for prospective high school students is not only based on academic scores but also on the results of the scholastic aptitude test (SAT). Verbal ability is an SAT subtest that assesses language management, vocabulary, and problem-solving abilities through a complete language study. This study developed a verbal ability test instrument for junior high school students consisting of the ability of synonyms, antonyms, and analogies. The data was collected from 300 junior high school students in grade nine who took a test with dichotomous data. The data analysis approach used one-order confirmatory factor analysis (CFA) with correlation factors. The results showed that CFA with correlation factors indicated the construct validity of the instrument was valid with the index criteria value χ^2=446.80, *df*=389, *p*-value=0.02267, root mean squared error of approximation (RMSEA)=0.022, goodness of fit index (GFI)=0.91, adjusted goodness-of-fit (AGFI)=0.89, and comparative fit index (CFI)=0.98. Then, construct reliability has good reliability with coefficient values for each dimension of 0.93, 0.95, and 0.84. As for the composite reliability of 0.88. It shows that using the verbal ability test instrument is feasible and has a reliable scale to measure the ability of junior high school students.

Corresponding Author:

Samsul Hadi
Department of Electrical and Engineering Education, Faculty of Engineering,
Universitas Negeri Yogyakarta
Karang Malang, Caturtunggal, 55281 Yogyakarta, Indonesia
Email: samsul_hd@uny.ac.id

1. INTRODUCTION

The education system in Indonesia implements a school-based education structure over 16 years, which is six years at the elementary school level, three years at junior high school, three years at senior high school, and four years at the university level. Indonesian government policy requires all children aged between 7 to 15 years to be given primary education at the elementary and junior high school levels [1]. The next level of education is high school which consists of general education and vocational schools. In learning activities at the school, the main focus is on teachers who are ready and able to carry out the teaching and learning process by the curriculum. The history of the curriculum in Indonesia has changed from year to year since independence until now. The education curriculum has undergone several revisions in 1947, 1952, 1962, 1968, 1975, 1984, 1994, 2004, 2006, and 2013 [2].

The fundamental goal of the 1947 curriculum was to implement educational reform in Indonesia based on the Dutch educational model, which attempts to produce pupils with character [3]. Then, the 1947 curriculum revision in 1952 aimed to form-wise, democratic and responsible citizens for the welfare of society and the nation. The 1964 revisions focused on fostering patriotism and nationalism based on national, international, and religious principles to improve students' intelligence, emotions, and bodies. Subsequent reforms were carried out in 1968, which were in line with political reforms from the 'Old Order' to the 'New Order' [2] and continued revisions in 1975, which are teacher-oriented guidelines to understand the goals of students learning so that they have skills to do active learning [4]. The 1975 curriculum was then revised in 1984, where learning continues to apply the concept of active learning to students and focuses on communicative competence [4]. In 1994, the 1984 curriculum revision was then carried out based on the Law of the Republic of Indonesia No. 2 of 1989. A significant change in this curriculum was adapting the subjects of education to the history of the nation's struggle as compulsory subjects, which were updated in the 1994 curriculum [5]. In addition, government policies in this curriculum priorities science-based education to dominate science subjects compared to social studies subjects [6].

A curriculum modification, known as the competency-based curriculum or the 2004 curriculum, was carried out due to structural changes in the Indonesian government system from centralized to decentralized [7], [8]. Competencies consist of graduate, general, and essential competencies. The following curriculum revision was carried out in 2006, where there was no fundamental difference between the previous curriculum. Nevertheless, the 2006 curriculum or school-based curriculum has met national education criteria to assure students' educational success [9]. Based on the difference between expectations and results, the following curriculum change uses the 2013 curriculum. Some of the problems in implementing the previous curriculum implementation include teacher competence, evaluation process, school facilities, and subject matter [6], [10], [11]. A significant change in the 2013 curriculum is a separate assessment process based on knowledge, skills, and attitude competencies. Changes in the Indonesian curriculum over the last 15 years have also emphasized the goal of knowledge and skill competency for each grade level and student-centered learning. However, the implementation of the 2013 curriculum requires the teacher's role to carry out learning and be busy with student assessments. At the end of 2014, the government reimposed the 2006 curriculum (education unit level curriculum), and the school stopped implementing the 2013 curriculum, which ran for one semester [12]. Schools with three semesters may use the 2013 curriculum. Several Special Capital District of Jakarta and West Java schools still use both curricula [13]. Although the 2006 curriculum was not long-term, the 2013 curriculum was again implemented in 2017 with revisions based on content and flexible and moderate implementation [14].

In addition, government policy in implementing the 2013 curriculum stipulates that the determination of majors at the high school level starts from grade 10, so the school must determine majors based on academic values and student talent abilities. Talent is one of the unique and distinctive characteristics that distinguish individuals [15]–[17]. According to Salkind and Rasmussen [18], talent is a set of characteristics related to an individual's ability to acquire knowledge or skills. The suitability of the department with the ability of students' talents will result in student learning being directed so that it can determine the continuity of student success in the future.

The school can track students' talent abilities to obtain accurate information in determining program majors for prospective high school students. It follows one of the reference criteria for program majors for grade 10 who will advance to grade 11 in the guide for preparation of learning outcomes reports for the education unit level curriculum. However, in implementing the 2013 curriculum, there is a policy revision where the majors are made to select junior high school students who want to continue in grade 10, high school level. The school can obtain information based on academic grades and student interests based on questionnaires, interviews, or other means to identify interests and talents [19]. Then, the 2013 curriculum document based on the Indonesian Minister of Education and Culture Regulation in 2013 [20] allows students to choose specialization groups (program majors) consisting of groups of mathematics and natural sciences, social sciences, and cultural and language sciences and can also choose subjects between specialization groups. Education units can also add criteria for majors according to the characteristics and needs of the school. According to the ministerial regulation, the criteria for program majors can be carried out by taking into account report cards and national exam scores for junior high schools or *Madrasah Tsanawiyah*. Other criteria are the recommendation of guidance and counseling teachers in junior high school, placement tests result when enrolling in high schools, and aptitude and interest tests by psychologists.

A person's talent in a particular field can be assessed by using specific potential tests such as potential academic test. The potential academic test aims to describe the strengths and weaknesses of the test-takers in conducting the test. The potential test to measure the actual ability of students based on the assessment of interests and talents in specific fields can use the scholastic aptitude test (SAT). Aptitude tests are cognitive measures used to predict future performance in activities such as school learning [21]–[23].

Talent ability is a predictor in predicting individual success in academics [18], [21], [24], [25]. The design of aptitude tests is focused on measuring verbal and numerical abilities, which are not directly related to the curriculum [26]. These two abilities explain the variance of talent ability of 38.8%, with a correlation between verbal and numerical abilities of 0.713 [27]. According to Oyetunde [21], the items used in the aptitude test measure specific abilities such as verbal and numerical abilities to predict academic performance in educational programs. Verbal and numerical abilities are also predictors of student performance, namely learning outcomes in the field of economics for junior high school students [28].

Verbal ability is one component of the scholastic aptitude test that predicts individual success. Various studies have stated that verbal ability predicts student achievement in learning [29]–[34]. Verbal ability aims to measure language management, vocabulary, and problem-solving skills by conducting a comprehensive language review. Verbal ability tests can measure various domains of vocabulary knowledge such as synonyms, antonyms, analogies, substitutions, understanding, and composing and completing sentences [35]–[37]. Janssen, Boeck, and Steene [38] stated that various measures of verbal ability such as vocabulary, reading comprehension, verbal analysis, and verbal tests are included in most intelligence tests. Then, verbal skills consist of the ability to find equivalent words (synonyms), opposites (antonyms), and connect vocabulary based on word relationships (analogy) [38]–[40]. According to Jigau [41], verbal ability consists of items containing four sets of words where test participants show two words that are the same or the meaning of words that are opposite (word analogy).

The variable indicators are observed to measure the ability to find equivalents and opposites, focusing on discussing the representation of the two types of word relationships using adjectives [42]–[45]. Then, Gross, Fischer, and Miller [46] describe the procedures for solving problems related to synonyms and antonyms. Solving synonyms items only require a conceptual step, while antonyms consist of direct antonyms and indirect antonyms, which require an associative step first and then involve direct antonyms. The following verbal ability is the ability to connect vocabulary based on word relationships or word analogies. In their research, Sternberg and Nigro [47] compared the responses of elementary school, middle school, and college students to verbal analogy items based on five different semantic relationship patterns, namely synonymous, antonymous, functional relationships, linear ordering, and category membership. Meanwhile, according to Jigau [41], word analogy points to connecting the same word or the meaning of the opposite word. In this study, the development of a test instrument to measure the verbal ability of junior high school students consisted of the ability of synonyms, antonyms, and word analogies.

So far, research using verbal ability test instruments is still lacking to predict the performance outcomes of junior high school students. The use of aptitude test instruments such as the differential aptitude test (DAT) by Santos and Boyon [48] consists of verbal and numerical reasoning abilities to predict the performance of 11th grade students of science, technology, engineering, and mathematics (STEM) interest, namely mathematics learning outcomes related to limits and continuity. The results showed that verbal and numerical reasoning skills significantly affected predicting students' performance with STEM interests, with the most significant predictor being numerical ability. Then, Amusa's research [49] also used a verbal ability test instrument to see the effect of high school students' performance in physics subjects. The instrument was used to test students' cognitive abilities in English grammar, verbal analogy, understanding, verbal deduction, sentence completion, and word groups. The results showed that verbal ability had a practical and functional effect on determining the performance results of high school students in physics. Furthermore, verbal ability test instruments were used to see differences in verbal abilities based on students' level of education and gender [50]. The results showed differences in verbal ability based on students' education level and gender. The students' verbal ability scores with the lowest average to the highest are students with junior high school, senior high school or vocational high school, and college levels. In addition, there are also differences in students' verbal ability scores based on gender, where the average verbal ability subtest score of female students is higher than male students.

It is necessary to develop a construct of the aptitude test instrument because the ability of scholastic talent impacts student performance and helps the school obtain information in determining majors for prospective high school students. These instruments can be employed consistently in selecting prospective students to identify the placement of program majors based on scholastic talent abilities. However, the development of the aptitude ability instrument construct is still lacking to measure students' verbal and numerical abilities at the junior high school level. Research on the development of the analogy ability test instrument, a sub-test of verbal ability, was only carried out to measure analogy ability at the undergraduate students [51]. The results showed that the test instrument items were valid and reliable so that the test instrument was feasible and consistent to measure student-level analogy abilities. Then, Candiasa, Natajaya, and Widiartini [52] researched by validating the vocational aptitude test instrument for prospective vocational high school students. The aptitude test instrument consists of numerical, analytical, visual, and communication skills.

The construct validity test used a data suitability index based on the criteria of Chi-square value, significance value (*p*-value), and comparative fit index (CFI) value. Construct validity was obtained valid so that the test instrument could be used to identify prospective vocational high school students' initial ability to determine the appropriate program majors. However, the construct of the aptitude test instrument can be used by students with lower levels of education. The development of the test instrument construct is not related to the curriculum that affects student learning materials for each level. However, the construct of the aptitude test instrument can be used for prospective high school students at the junior high school education level. The SAT is part of the DAT [53]. The DAT is one of the most widely used multiple-aptitude test series in education in educational counseling for middle and high school-aged children [54]. Therefore, SAT can also be carried out on prospective high school students who obtain information on specializations or program majors based on talent abilities. In addition, the development of the aptitude test instrument construct is not related to the curriculum, which has no effect on student learning materials for each level.

The process of developing the test instrument construct needs to test the reliability and validity of the construct. The test was conducted to determine the feasibility and consistency of the test instrument capable of measuring the test's ability. Testing the validity and reliability of an instrument can use factor analysis. Factor analysis is a statistical process that identifies the covariance between observed variables to reduce them to a minimal number of latent variables [55]. Researchers can use factor analysis to get practical information about the measurement instrument's internal structure, specifically the relationship between latent variables and factors on observable variables [56]. Then, McCoach, Gable, and Madura [57] added that factor analysis would allow researchers to identify the number of constructs in an instrument and explain the pattern of relationships between observed variables and constructs and between constructs and latent variables. The factor analysis approach used in this study is confirmatory factor analysis (CFA). CFA investigates the connection between observable and unobserved variables by having researchers apply a fictitious model to estimate the population covariance matrix and then compare it to the covariance matrix of the experimental sample [58]. This study aimed to develop a verbal ability test instrument for prospective high school students. The construct of the test instrument obtained based on the literature review was tested for validity and reliability to obtain a construct by the conceptual model. The test uses a first-order confirmatory factor analysis approach with a correlation factor model.

2. RESEARCH METHOD

This development research aims to examine the reliability and construct validity of the verbal ability test instrument for ninth-grade junior high school students, based on a literature review and conceptual model. The observed verbal ability construct variables consisted of synonyms, antonyms, and analogy. According to Retnawati [59], developing a test instrument consists of three stages. The initial stage is to determine the purpose of the instrument preparation, look for the relevant theory or scope of material, formulate the instrument item indicators, and then develop the instrument items. The next stage is content validation, revision based on expert validator input, making the final assessment instrument, and collecting test data. The last stage is the trial results' data analysis using a CFA approach.

The population in this study was 9th-grade junior high school students consisting of 300 students. The sampling technique used is the proportional random sampling technique [60], consisting of 41 students from private schools and 259 students from public schools. The sample size for CFA is based on the number of variables or items observed. According to Hair *et al*. [61], the sample size using the Maximum Likelihood (ML) estimate was around 100-200 respondents. Meanwhile, Anderson and Gerbing [62] proposed that a sample size of 150 or greater be used to achieve the minimum standard error. Furthermore, Schumacher and Lomax [63] argue that 250-500 respondents are required for a researcher to perform precise calculations with CFA. Then, Comrey and Lee [64] determined the sample sizes of 50, 100, 200, 300, and 500, respectively, with very bad, bad, moderate, good, very good, and perfect criteria. Data collection in this study was conducted online using google forms. Responses to the test instrument to obtain data on the verbal ability by working on each test item using time settings. The test instrument consists of two parts, with the first part containing questions about student demographics. At the same time, the second part contains questions related to verbal ability consisting of the ability of synonyms, antonyms, and analogies. The verbal ability test instrument consists of 30 items with ten items each for the ability of synonyms, antonyms, and analogies. Student responses to the test instrument in dichotomous data with the correct answer for each item are worth 1, and the wrong answer is 0. Student responses to demographic questions are given in Table 1.

Data analysis in this study was used to test the reliability and validity of the test instrument construct using the CFA analysis approach. CFA is a qualitative and statistical process consisting of testing the reliability of items or indicators of the observed variables, construct reliability, face validity and qualitative content, quantitative measures of convergent and discriminant validity, and goodness of fit [65]. CFA aims to

evaluate latent structures based on a priori development and theoretical support [66], [67]. In addition, according to Hair *et al*. [65], CFA aims to confirm the nature of measurements consisting of observed variables containing item indicators used to measure latent constructions that have been defined and operationally defined. Testing using CFA is an essential part of the validation process for constructs obtained based on evidence in a theoretical model [68].

The theoretical model is tested by determining the relationship between the observed variables and latent factors and the relationship between the factors themselves. The relationship between the factors themselves is referred to as the correlation factor. The results of the model consist of parameters that: i) Remain at a specific value; ii) Are limited so that their values are the same as other parameters; and iii) Are free to take unknown values [69]. The CFA method uses the first order with the correlation factor model. The study used LISREL 8.71 software for data analysis. The criteria for the conformity index of the CFA model are based on the value of $\chi^{\wedge 2}$, *p*-value, root mean squared error of approximation (RMSEA), goodness of fit index (GFI), adjusted goodness of fit index (AGFI), and comparative fit index (CFI). According to Arbuckle [70], the criterion for the Chi-square value is $\chi^2 \leq 2df$ while for the *p*-value using a value greater than 0.05 [71], [72]. The RMSEA value of less than 0.08 indicates a good fit model, but a value less than 0.05 indicates a perfect fit model [73]. Then, the GFI value is more significant than 0.90, AGFI is greater than 0.80, and the RMSEA value is less than 0.08 [74]. Then, a CFI value greater than 0.90 indicates a suitable model [75].

Table 1. Demographic data of research sample

School status	School name	Location	Number of students
Private school	Pondok Pesantren MDIA Bontoala	Makassar	41
State school	SMPN 1 Makassar	Makassar	70
	SMPN 1 Gunungjati	Cirebon	41
	SMPN 1 Kuranji	Tanah Bumbu	18
	SMPN 4 Tamalatea	Jeneponto	36
	SMPN 1 Pangsid	Sidenreng Rappang	16
	SMPN 1 Malili	Luwu Timur	16
	SMPN 1 Kajuara	Bone	31
	SMPN 1 Bajeng Barat	Gowa	24
	SMPN 1 Cempa	Pinrang	7
Percentage of number of students by gender			174 (M: 58%) 126 (F: 42%)

3. RESULTS AND DISCUSSION

The variables observed in this study were the ability of synonyms (SIN), antonyms (ANT), and analogies (ANL). Testing the validity and construct reliability of the test instrument using the first-order CFA results with the correlation factor model. The CFA results describe the suitability of the measurement construct model with data on the verbal ability of junior high school students. Another CFA result is to identify the effect and contribution of the items on the observed variables.

3.1. Confirmatory factor analysis of the factor correlation model

Data analysis of the verbal ability of junior high school students with the CFA approach used a factor correlation model, namely analyzing the observed variables simultaneously. The following CFA results consist of the goodness fit of test criteria in a standardized solution diagram. Then, identify the significant effect for each item on the observed variable based on the standardized value of the factor load and the *t*-value. At the same time, the coefficient of determination defines the magnitude of the item's contribution to the observed variable. The standardized solution diagram in Figure 1 results from first-order CFA using the correlation factor model by analyzing the observed variables simultaneously.

The results of the first-order CFA using the correlation factor in Figure 1 obtained initial results showing that there are items of the verbal ability test instrument measuring more than one dimension that significantly affect the observed variables with a *p*-value of less than 0.05. Then, the analysis uses modification twice by doing correlations between items. Correlation between items for the first modification is done by connecting items SIN1.2-SIN1.3, SIN1.3-SIN3.10, ANT1.1-ANT1.2, ANT1.3-ANT3.6, SIN2.8-ANT3.7, SIN1.3-ANT3.8, SIN1.2-ANL2.3, and ANL4.7-ANL5.9, and ANL4.8-ANL5.9. As for the second modification, connect items SIN1.2-SIN3.10, ANT3.7-ANT3.8, SIN3.10-ANL3.6, and ANL4.8-ANL5.9. However, the modification results show that the constructed model of the test instrument satisfies the same conditions as the test items, which have a significant effect on the observed variables with a *p*-value of 0.02267, which is smaller than 0.05. The results of the CFA indicate that there are items of verbal ability test instruments that measure more than two observed ability variables. Other CFA results consist of the GFI, AGFI, and CFI. Table 1 shows the CFA findings before and after making modifications based on the goodness of fit of test criteria.

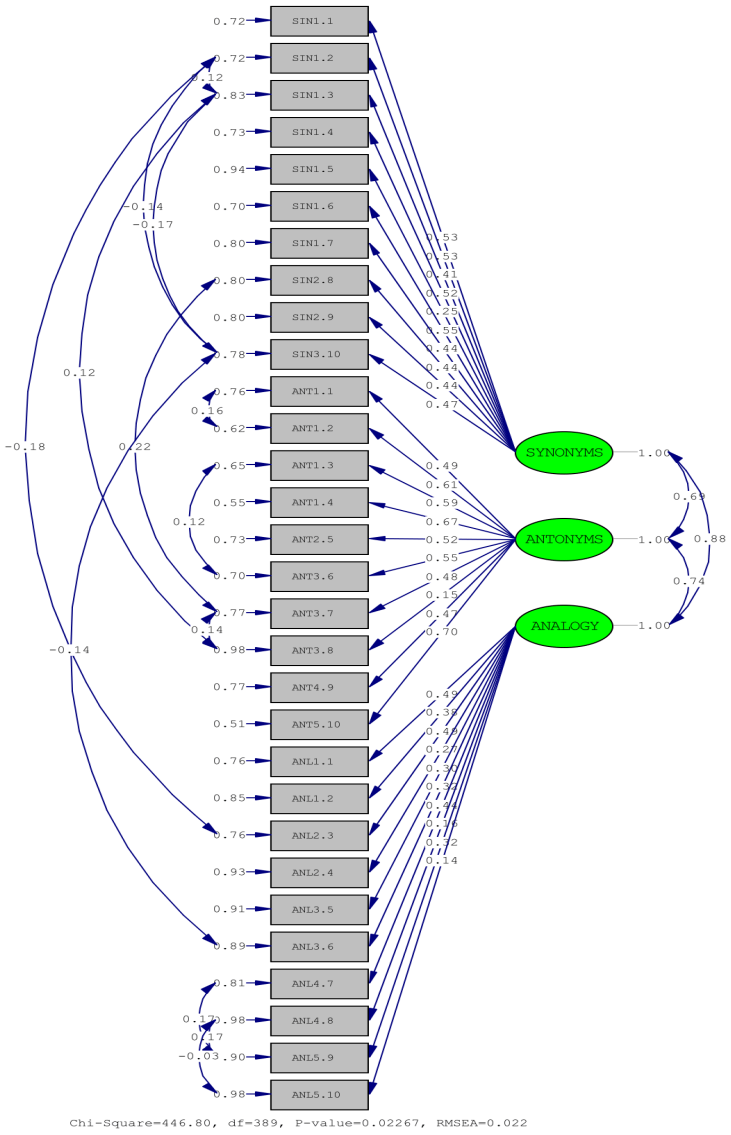

Figure 1. Standardized solution diagram of first-order CFA results

Table 1. Goodness fit of test criteria value

Criteria	Cut off value	Before modification	Status	After modification	Status
χ^2	≤2*df*	584.31	Fit	446.80	Fit
p-value	>0.05	0.00000	Not fit	0.02267	*
RMSEA	<0.05	0.039	Fit	0.022	Fit
GFI	≥0.90	0.88	Not fit	0.91	Fit
AGFI	≥0.80	0.87	Fit	0.89	Fit
CFI	≥0.90	0.95	Fit	0.98	Fit

*=Marginal fit with *p*-value<0.05

The initial CFA results in Table 1 show the results before modifying the constructed model of the verbal ability test instrument that did not match the data. The instrument's construct model was updated after a *p*-value of 0.00, which was less than 0.05, was achieved using the goodness fit of test criterion. Modification by comparing items twice to reduce the value of χ^2 of 137.51. However, the constructed model of the test instrument still has a *p*-value of 0.02267, which is smaller than 0.05, so the fit condition is marginally fulfilled. Other index criteria based on the suitability index of the measurement model in Table 1 show for the criteria the value of χ^2=446.80, *df*=389 with $\chi^2 \leq 2df$ [70] and the value of RMSEA=0.022 with RMSEA<0.05 [73], [76]. The results of the next index, namely the GFI value=0.91, AGFI=0.89, and

CFI=0.98 each meet the GFI value>0.90 and AGFI>0.80 [74] and the CFI value>0.90 [75]. These results imply that the constructed model for measuring the verbal ability test instrument is consistent with the theoretical model. The constructed model for measuring verbal ability is feasible to measure junior high school students' ability to use synonyms, antonyms, and analogies.

Furthermore, the results of the test items have a significant effect on the respective abilities of the observed variables based on the standardized factor load (SFL) values and *t*-values given in Table 2. Items or indicators have good validity on the construct or latent variable if the SFL value is greater or equal to 0.30 [77], [78]. In addition, the t-value must be greater than or equal to the critical value of 1.96 [61], [79].

The CFA results in Table 2 show that the standardized value of factor loading for each item of the observed variable has a positive value and *t*-value greater than 1.96. The item has a significant effect on the observed variable. However, there are items from the observed variables, each of which has a SFL value of less than 0.30. These items are SIN1.5 and ANT4.8 for the synonym and antonym ability variables, respectively. At the same time, the items for analogy ability are ANL2.4, ANL4.8, and ANL5.10 items. So that the five items are not valid for measuring each construct variable of the verbal ability test instrument for junior high school students. Then, identify the next item related to the contribution of each item to the observed variables based on the coefficient of determination. The item SIN1.6 has the most significant contribution to the synonym ability variable of 30.00%. In terms of the antonym and analogy ability variables, the most critical item was ANT1.4, which provided 49.00%, followed by ANL1.1 and ANL2.3, which each contributed 24.00%.

Table 2. SFL values and t-values of the observed variable test items

Dimensions	Items	SFL value	*t*-value
Synonyms	SIN1.1	0.53	9.04
	SIN1.2	0.53	8.65
	SIN1.3	0.41	6.54
	SIN1.4	0.52	8.69
	SIN1.5	**	4.00
	SIN1.6	0.55	9.37
	SIN1.7	0.44	7.34
	SIN2.8	0.44	7.39
	SIN2.9	0.44	7.33
	SIN3.10	0.47	7.66
Antonyms	ANT1.1	0.49	8.25
	ANT1.2	0.61	10.77
	ANT1.3	0.59	10.24
	ANT1.4	0.67	12.02
	ANT2.5	0.52	8.96
	ANT3.6	0.55	9.38
	ANT3.7	0.48	8.28
	ANT4.8	**	2.30
	ANT5.9	0.47	7.99
	ANT3.10	0.70	12.85
Analogy	ANL1.1	0.49	7.90
	ANL1.2	0.38	6.03
	ANL2.3	0.49	7.82
	ANL2.4	**	4.20
	ANL3.5	0.30	4.74
	ANL3.6	0.32	5.05
	ANL4.7	0.44	7.00
	ANL4.8	**	2.42
	ANL5.9	0.32	5.04
	ANL5.10	**	2.09

**=Reference item, SFL value does not meet the criteria

3.2. Construct validity and reliability

The measurement stage that occurs after the measurement model of an instrument construct model reaches conformance with the theoretical model is called construct validation and reliability determination and identification. According to Hair *et al*. [61], once the measurement model's conformity requirements have been met, the constructed model must be tested for reliability and validity before proceeding with the structural equation model. Construct validity assures that a set of data appropriately reflects the investigated hidden theoretical notion [80]. The validity of the latent construct measurement model consists of construct validity, convergent validity, and discriminant validity [81]. Construct validity is achieved when the instrument's construct model meets the fit index conditions [82]. Table 2 shows that all index criteria are met in the construct validation of the verbal ability test instrument based on the measurement model's

appropriateness index. In this case, the construct validity of the verbal ability test instrument is valid to measure the verbal ability of junior high school students. The analysis results, namely testing of discriminant and convergent validity, omega reliability (OR), and construct reliability (CR) are presented in Table 3.

Table 3. Results of calculation of validity and reliability of verbal ability constructs

Observed variable	Dimensions	AVE	Square of AVE	CR	OR
Verbal ability	Synonyms	0.22	0.47	0.93	
	Antonyms	0.29	0.54	0.95	0.88
	Analogy	0.12	0.35	0.85	

The calculation results in Table 3 show the average variance extracted (AVE) value based on the standardized value of factor loading for each variable item observed in the verbal ability test instrument. The synonym, antonym, and analogy ability variables had AVE values of 0.22, 0.29, and 0.12, respectively. According to Hair *et al*. [80], a measure of convergent validity explains how items measure constructs in structural equation modeling (SEM) based on an AVE value greater than or equal to 0.50. Then, according to Shkeer and Awang [83], it is stated that the validity criteria converge with an AVE value greater than 0.60. Meanwhile, according to Ghozali [84], the criterion of a convergent validity value of 0.70 is considered good validity, while a value of 0.50 – 0.60 is still acceptable for early-stage research. The study results in Table 2 show that each observed variable has an AVE value of 0.50 smaller. The latent construct's variance for each item indicator collected is lower than the error variance. As a result, the verbal ability test instrument's items are less valid to measure each observable variable. Thus, the verbal ability test instrument has poor convergent validity to measure each variable observed in the verbal ability test instrument.

The subsequent identification of validity is discriminant validity. Discriminant validity is required to develop instruments that involve latent variables [85]. Discriminant validity is also called "divergent validity" [86], which means that two ideas must be very different. The discriminant validity test shows that one thing is different from another [87]. According to Hair *et al*. [80], discriminant validity is defined as a correlation between two constructs. If the correlation value of the two constructs is less than 0.85, then the instrument construct meets discriminant validity [88]. The CFA model correlation of the observed variables of the verbal ability test instrument simultaneously obtained correlations between the observed variables, namely the SIN-ANT, ANT-ANL, and SIN-ANL variables, respectively 0.69, 0.74, and 0.88. The results showed that the verbal ability test instrument met discriminant validity with each correlation value smaller than 0.85 except for the correlation between the synonym and analogy ability variables. Thus, the verbal ability test instrument has a construct of the observed variable that is different from the other observed variables except for the variable ability of synonyms and the ability of analogy.

The result of the following test is the analysis of the calculation of the reliability of the verbal ability test instrument to measure the ability of junior high school students. An instrument is said to be reliable if the instrument can measure the same ability repeatedly with relatively consistent results. The reliability coefficient determines the degree of consistency. The construct and composite reliability were discussed in this study. The SFL values from the CFA results can be used to estimate construct reliability [89]. According to Retnawati [59], construct reliability is the consistency of the variables that constitute the latent construct to be tested. In addition, construct dependability is used to assess how well the factors underlying the construct are represented in structural equation modeling [90]. The test instrument's reliability meets reliability and has a construct reliability coefficient (CR) greater than 0.70 [91], [92]. The results showed that the three variables observed in the verbal ability test instrument had a construct reliability coefficient of 0.93, 0.95, and 0.85. It shows that the value of the construct reliability coefficient for each construct has a value greater than 0.70. So that for each variable, the verbal ability test instrument has good reliability for measuring the ability of junior high school students. The following reliability calculation analysis is composite reliability. Composite reliability is also called internal consistency, a combination of the reliability of the latent construction that underlies the measurement scale [80], [89]. Composite reliability was obtained based on the reliability coefficient omega (ω). The results showed that the verbal ability test instrument had an omega reliability coefficient of 0.88, indicating a reliability coefficient value greater than 0.70. It signifies that the verbal ability instrument has a consistent scale for measuring the ability of junior high school students.

This study's development of talent instruments was primarily focused on verbal ability assessment. Many structures can be constructed to assess junior high school students' scholastic talent abilities, such as numerical calculations analysis. However, research on the development of the scholastic talent instrument construct still lacks to measure the ability of junior high school students. The development and validation of talent skills instruments by Wulandari *et al*. [93] consists of visual, numerical, verbal, spatial reasoning, and

vocabulary tests. Instrument validity only uses expert judgment content validity in this study, and unidimensional testing uses exploratory factor analysis based on eigenvalues. The validity of the instrument in this study has not used CFA. The results of the first-order CFA with a one-factor model can also be used to identify items that only measure one component in unidimensional testing. The following study by Anazia [28] used quantitative and verbal aptitude tests to identify the influence of these talents on the performance of high school students. The analysis of the instrument in this study is only to calculate the reliability coefficient. The Pearson product-moment correlation coefficient was used to get the reliability coefficient. Then, research by Candiasa, Natajaya, and Widiartini [52] validates the aptitude test instrument for prospective vocational high school students consisting of sub-tests of numerical, analytical, visual, and communication skills in using language. The instrument validity analysis consisted of content and construct validation. The construct validity used a conformity index based on chi-squared value, significance value (*p*-value), and CFI value. Meanwhile, in this study, we pay attention to additional criteria, namely the value of RMSEA, AGFI, and GFI. Construct validity was obtained valid so that the test instrument could be used to identify the initial abilities of prospective students based on the results of the aptitude test that corresponded to the program majors at the vocational high school level.

Testing the validity and construct reliability of this study's verbal ability test instrument only used first-order CFA with the correlation factor model. The study has not tested the reliability and validity of the construct using second-order CFA. This study proves that the conceptual model constructed is by the theoretical model it is valid in measuring the verbal abilities of junior high school students. The results of the construct reliability test also show that each observed variable has good reliability so that the aptitude test instrument is consistent when measuring for a relatively long time to measure the ability of junior high school students. The construct validity of the verbal ability instrument was only assessed on a few research subjects with a limited number of samples, which is a flaw in this study. As a result, experimenting with a bigger sample size is recommended for better results.

4. CONCLUSION

This study used first-order CFA with the correlation factor model to examine the validity and construct reliability of the verbal ability test instrument. The validity and construct reliability of the verbal ability test instrument were valid and reliable for measuring 9th-grade junior high school students. Then, the items in the instrument are also valid for measuring the ability variables of synonyms, antonyms, and analogies. However, in the testing of this research, several estimates were made by making modifications to obtain a measurement model following the test's ability data. As a result, this research should be continued by analyzing the findings of a second-order CFA test to acquire thorough validity and reliability test results.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to all who contributed to this study. The Ministry of Education and Culture of the Republic of Indonesia provided financial assistance for this study through scholarships. The authors also thank the Directorate of Research, Technology, and Community Service, Directorate General of Higher Education, Research and Technology, Ministry of Education, Culture, Research and Technology for funding sources to implement doctoral dissertation grant research in 2022.

REFERENCES

[1] S. N. Azkiyah and A. Mukminin, "In search of teaching quality of efl student teachers through teaching practicum: Lessons from a teacher education program," *Center for Educational Policy Studies Journal*, vol. 7, no. 4, pp. 105–124, 2017, doi: 10.26529/cepsj.366.

[2] A. Mukminin, A. Habibi, L. D. Prasojo, A. Idi, and A. Hamidah, "Curriculum reform in indonesia: Moving from an exclusive to inclusive curriculum," *Center for Educational Policy Studies Journal*, vol. 9, no. 2, pp. 53–72, 2019, doi: 10.26529/cepsj.543.

[3] K. H. Dewantara, *The work of Ki Hadjar Dewantara in the first part*. Yogjakarta: Majelis Luhur Persatuan Taman Siswa (in Indonesian), 1977.

[4] A. H. Gunawan, *Educational policies in Indonesia]*. Jakarta: Bina Aksara (in Indonesian), 1986.

[5] S. H. Hasan, "Curriculum development: Pedagogical ideological and theoretical developments (in Indonesian)," 2007. [Online]. Available: www.geocities.ws/konferensinasionalsejarah/s _hamid_hasan.

[6] A. Habibi, M.Pd., A. Mukminin, M. Sofwan, and U. Sulistiyo, "Implementation of classroom management by English teachers at high schools in Jambi, Indonesia," *Studies in English Language and Education*, vol. 4, no. 2, p. 172, 2017, doi: 10.24815/siele.v4i2.6104.

[7] C. Bjork, *Indonesian education: Teachers, schools, and central bureaucracy*. London, UK: Routledge, 2005.

[8] S. Kristiansen and Pratikno, "Decentralising education in Indonesia," *International Journal of Educational Development*, vol. 26, no. 5, pp. 513–531, 2006, doi: 10.1016/j.ijedudev.2005.12.003.

[9] R. Raihani, "Education for multicultural citizens in Indonesia: policies and practices," *Compare*, vol. 48, no. 6, pp. 992–1009, 2018, doi: 10.1080/03057925.2017.1399250.

[10] R. W. Febriya and W. Nuryono, "Survey on counselors' perceptions and readiness for guidance and counseling based on the 2013 curriculum at south Surabaya senior high school," (in Indonesian), *Jurnal BK UNESA*, vol. 04, no. 03, pp. 1–10, 2014.
[11] W. S. Gershon, "Introduction: Towards a Sensual Curriculum," *Journal of Curriculum Theorizing*, vol. 27, no. 2, pp. 1–16, 2011.
[12] S. H. Hasan, "History education in curriculum 2013: A new approach to teaching history," *Historia: Jurnal Pendidik dan Peneliti Sejarah*, vol. 14, no. 1, p. 163, 2013, doi: 10.17509/historia.v14i1.2023.
[13] H. Prasetyono, A. Abdillah, T. Djuhartono, I. P. Ramdayana, and L. Desnaranti, "Improvement of teacher's professional competency in strengthening learning methods to maximize curriculum implementation," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 10, no. 2, pp. 720–727, 2021, doi: 10.11591/ijere.v10i2.21010.
[14] M. Nurtanto, N. Kholifah, A. Masek, P. Sudira, and A. Samsudin, "Crucial problems in arranged the lesson plan of vocational teacher," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 10, no. 1, pp. 345–354, 2021, doi: 10.11591/ijere.v10i1.20604.
[15] M. Kincaid, "A study of individual differences in educational situations," *Psychological Review*, vol. 32, no. 1, pp. 34–53, 1925, doi: 10.1037/h0073540.
[16] R. E. Jung, S. G. Ryman, A. A. Vakhtin, J. Carrasco, C. Wertz, and R. A. Flores, "Subcortical correlates of individual differences in aptitude," *PLoS ONE*, vol. 9, no. 2, 2014, doi: 10.1371/journal.pone.0089425.
[17] P. Robinson, "Individual differences, aptitude complexes, SLA processes, and aptitude test development," *Second Language Learning and Teaching*, vol. 4, pp. 57–75, 2012, doi: 10.1007/978-3-642-20850-8_4.
[18] N. J. Salkind and K. Rasmussen, *Encyclopedia of measurement and statistics*. London: SAGE Publications, Inc., 2007, doi: 10.4135/9781412952644.
[19] Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah, *Guidelines for Preparing Student Learning Outcomes Reports based on the High School Education Unit Level Curriculum (KTSP)*. Jakarta: Departemen Pendidikan Nasional Republik Indonesia (in Indonesian), 2006.
[20] Minister of Education and Culture Republic Indonesia, "Regulation of The Minister of Education and Culture of The Republic of Indonesia Number 59 of 2014 Concerning The 2013 Curriculum of High School/Madrasah Aliyah," Minister of Education and Culture (in Indonesian), 2013.
[21] A. A. Oyetunde, "Construction and validation of a general science aptitude test (GSAT) for Nigerian junior secondary school graduate," *Iorin Journal of Education*, vol. 27, pp. 22–23, 2007.
[22] L. R. Aiken, *Psychological testing and assessment*, 6th ed. Boston: Allyn and Bacon, Inc., 1988.
[23] N. E. Gronlund, *Measurement and evaluation in teaching*. New York: Macmillan Publishing Co, 1981.
[24] M. Curabay, "Meta-analysis of the predictive validity of scholastic aptitude test (SAT) and american college testing (ACT) scores for college GPA," University of Denver, 2016.
[25] P. Pyari, K. Misha, and B. Dua, "A study of impact of aptitude in mathematics as stream selection at higher secondary level," *Issues and Ideas in Education*, vol. 4, no. 2, pp. 141–149, 2016, doi: 10.15415/iie.2016.32011.
[26] M. Koljatic, M. Silva, and R. Cofré, "Achievement versus aptitude in college admissions: A cautionary note based on evidence from Chile," *International Journal of Educational Development*, vol. 33, no. 1, pp. 106–115, 2013, doi: 10.1016/j.ijedudev.2012.03.001.
[27] R. A. Olatoye and A. A. Aderogba, "Performance of Senior Secondary School Science Students in Aptitude Test : The Role of Student Verbal and Numerical Abilities," *Journal of Emerging Trends in Educational Research and Policy Studies*, vol. 2, no. 6, pp. 431–435, 2011.
[28] I. U. Anazia, "Quantitative and verbal aptitudes as predictors of senior secondary school students' performance in economics," *IAFOR Journal of Education*, vol. 7, no. 1, pp. 7–18, 2019, doi: 10.22492/ije.7.1.01.
[29] A. Toomela, E. Kikas, and E. Mõttus, "Ability grouping in schools:a study of academic achievement in five schools in Estonia," *Trames*, vol. 10, no. 1, pp. 32–43, 2006.
[30] D. O. Fakeye, "Componential analysis as a model of ESL vocabulary instruction," *African Journal of Educational Research*, vol. 10, no. 1, pp. 14–24, 2006.
[31] A. J. Abiodun and A. O. Folaranmi, "Effects of verbal ability on second language writers' achievement in essay writing in english language," *International Journal of African & African American Studies*, vol. 6, no. 1, pp. 61–67, 2007.
[32] M. G. Vaughn, K. M. Beaver, J. Wexler, M. DeLisi, and G. J. Roberts, "The Effect of School Dropout on Verbal Ability in Adulthood: A Propensity Score Matching Approach," *Journal of Youth and Adolescence*, vol. 40, no. 2, pp. 197–206, 2011, doi: 10.1007/s10964-009-9501-1.
[33] A. Porojan, A. Morgan, and C. Avila, "Predicting academic performance and attrition in undergraduate students," *The International Journal of Health, Wellness, and Society*, vol. 3, no. 2, pp. 103–116, 2013, doi: 10.18848/2156-8960/cgp/v03i02/41070.
[34] A. I. Gambari, A. U. Kutigi, and P. O. Fagbemi, "Effectiveness of computer-assisted pronunciation teaching and verbal ability on the achievement of senior secondary school students in oral english," *GIST Education and Learning Research Journal*, vol. 8, no. 8, pp. 11–28, 2014.
[35] D. F. Lohman and J. M. Lakin, "Intelligence and reasoning," in *The Cambridge handbook of intelligence*, Cambridge University Press, 2011.
[36] W. Widhiarso and Haryanta, "Examining method effect of synonym and antonym test in verbal abilities measure," *Europe's Journal of Psychology*, vol. 11, no. 3, pp. 419–431, 2015, doi: 10.5964/ejop.v11i3.865.
[37] H. Wang, F. Tian, B. Gao, C. Zhu, J. Bian, and T. Y. Liu, "Solving verbal questions in IQ test by knowledge-powered word embedding," *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016, pp. 541–550, doi: 10.18653/v1/d16-1052.
[38] R. Janssen, P. De Boeck, and G. Vander Steene, "Verbal fluency and verbal comprehension abilities in synonym tasks," *Intelligence*, vol. 22, no. 3, pp. 291–310, 1996, doi: 10.1016/S0160-2896(96)90024-0.
[39] S. Azwar, R. Suhapti, Haryanta, and W. Widhiarso, *Postgraduate Academic Potential Test (A1)*. Yogyakarta: Fakultas Psikologi, Universitas Gadjah Mada (in Indonesian), 2009.
[40] I. I. Krisna, D. Mardapi, and S. Azwar, "Determining standard of academic potential based on the Indonesian Scholastic Aptitude Test (TBS) benchmark," *Research and Evaluation in Education*, vol. 2, no. 2, p. 165, 2016, doi: 10.21831/reid.v2i2.8465.
[41] M. Jigău, *Career Counselling: Compendium of Methods and Techniques*. Bucharest: AFIR, 2007.
[42] W. G. Charles, M. A. Reed, and D. Derryberry, "Conceptual and associative processing in antonymy and synonymy," *Applied Psycholinguistics*, vol. 15, no. 3, pp. 329–354, 1994, doi: 10.1017/S0142716400065929.
[43] S. Jones, C. Paradis, M. L. Murphy, and C. Willners, "Googling for 'opposites': A web-based study of antonym canonicity," *Corpora*, vol. 2, no. 2, pp. 129–155, 2007, doi: 10.3366/cor.2007.2.2.129.

[44] C. Paradis, "Good, better and superb antonyms," *The annual texts by foreign guest professors*, vol. 3, pp. 385–402, 2010.
[45] C. Paradis and C. Willners, "Antonymy: From convention to meaning making," *Review of Cognitive Linguistics*, vol. 9, no. 2, pp. 367–391, 2011, doi: 10.1075/rcl.9.2.02par.
[46] D. Gross, U. Fischer, and G. A. Miller, "The organization of adjectival meanings," *Journal of Memory and Language*, vol. 28, no. 1, pp. 92–106, 1989, doi: 10.1016/0749-596X(89)90030-2.
[47] R. J. Sternberg and G. Nigro, "Developmental Patterns in the Solution of Verbal Analogies," *Child Development*, vol. 51, no. 1, p. 27, 1980, doi: 10.2307/1129586.
[48] J. C. D. Santos and M. C. L. Boyon, "Numerical and verbal reasoning aptitudes as predictors of STEM students' performance on limits and continuity," *Educational Measurement and Evaluation Review*, vol. 11, pp. 14–24, 2020.
[49] J. O. Amusa, "Verbal ability and critical thinking skills as determinants of students' academic achievement in secondary school physics," *International Journal of Educational Research and Policy Making*, vol. 3, no. 1, pp. 395–405, 2020.
[50] R. E. Izzaty and F. A. Setiawati, "Influence of educational level and gender on students," *International Conference on Education, Social Sciences and Humanities* 2019, pp. 156–165, doi: 10.2991/icesshum-19.2019.25.
[51] E. Septianingsih and M. A. Jerusalem, "Developing instrument of academic potential test analogy verbal ability for undergraduate students," *Journal of Education and Learning (EduLearn)*, vol. 15, no. 2, pp. 234–241, 2021, doi: 10.11591/edulearn.v15i2.14220.
[52] I. M. Candiasa, N. Natajaya, and K. Widiartini, "Vocational aptitude test," *SHS Web of Conferences*, vol. 42, p.1-7, 2018, doi: 10.1051/shsconf/20184200044.
[53] A. Anastasi and S. Urbina, *Psychological testing*, 7th ed. New Jersey: Prentice-Hall, Inc., 1997.
[54] G. K. Bennett, H. G. Seashore, and A. G. Wesman, *Differential aptitude tests*. Psychological Corporation, 1947.
[55] R. L. Worthington and T. A. Whittaker, "Scale development research: a content analysis and recommendations for best practices," *The Counseling Psychologist*, vol. 34, no. 6, pp. 806–838, 2006, doi: 10.1177/0011000006288127.
[56] M. de los Á. Morata-Ramírez and F. P. Holgado-Tello, "Construct validity of likert scales through confirmatory factor analysis: A simulation study comparing different methods of estimation based on pearson and polychoric correlations," *International Journal of Social Science Studies*, vol. 1, no. 1, 2013, doi: 10.11114/ijsss.v1i1.27.
[57] D. B. McCoach, R. K. Gable, and J. P. Madura, *Instrument Development in the Affective Domain*, 3rd ed. New York: Springer, 1986.
[58] J. B. Schreiber, F. K. Stage, J. King, A. Nora, and E. A. Barlow, "Reporting structural equation modeling and confirmatory factor analysis results: A review," *Journal of Educational Research*, vol. 99, no. 6, pp. 323–338, 2006, doi: 10.3200/JOER.99.6.323-338.
[59] H. Retnawati, *Validity, Reliability & Item Characteristics (Guide for Researchers, Students, and Psychometrics)*. Yogyakarta: Parama Publishing (in Indonesian), 2017.
[60] W. G. Cohran, *Sampling techniques*. New York: John Wiley & Sons, 2010.
[61] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis*, 6th ed. Pearson Prentice Hall, 2006.
[62] J. C. Anderson and D. W. Gerbing, "Structural equation modeling in practice: a review and recommended two-step approach," *Psychological Bulletin*, vol. 103, no. 3, pp. 411–423, 1988, doi: 10.1037/0033-2909.103.3.411.
[63] R. E. Schumacker and R. G. Lomax, *A beginner's Guide to Structural Equation Modeling*, 3rd ed. New York: Routledge, 2010.
[64] A. L. Comrey and H. B. Lee, *A first course in factor analysis*, 2nd ed. Lawrence Erlbaum Associates, Inc., 1992.
[65] J. F. Hair, M. C. Howard, and C. Nitzl, "Assessing measurement model quality in PLS-SEM using confirmatory composite analysis," *Journal of Business Research*, vol. 109, pp. 101–110, 2020, doi: 10.1016/j.jbusres.2019.11.069.
[66] B. Bryne, *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum, 1998.
[67] R. H. Hoyle and A. T. Panter, "Writing about structural equation models," in *Structural equation modeling: Concepts, issues and applications*, R. H. Hoyle, Ed. Newbury Park, CA: Sage, 1993, pp. 158–176.
[68] C. DiStefano and B. Hess, "Using confirmatory factor analysis for construct validation: An empirical review," *Journal of Psychoeducational Assessment*, vol. 23, no. 3, pp. 225–241, 2005, doi: 10.1177/073428290502300303.
[69] J. W. van Prooijen and W. A. van der Kloot, "Confirmatory analysis of exploratively obtained factor structures," *Educational and Psychological Measurement*, vol. 61, no. 5, pp. 777–792, 2001, doi: 10.1177/00131640121971518.
[70] J. L. Arbuckle, *Amos user's guide: version 3.6*. Chicago: Small Waters Cooperation, 1997.
[71] J. D. Elashoff, F. N. Kerlinger, and E. J. Pedhazur, "Multiple regression in behavioral research," *Journal of the American Statistical Association*, vol. 70, no. 352, p. 959, 1975, doi: 10.2307/2285468.
[72] K. G. Joreskog and D. Sorbom, *Lisrel 8.54 help file*. Chicago: Scientific Software International, 2003.
[73] M. W. Browne and R. Cudeck, *Alternative ways of assessing model fit*. Newbury Park, CA: Sage, 1993.
[74] S. Sharma, *Applied multivariate techniques*. New York: John Wiley & Sons, Inc, 1996.
[75] L. T. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural Equation Modeling*, vol. 6, no. 1, pp. 1–55, 1999, doi: 10.1080/10705519909540118.
[76] J. H. Steiger, *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT, 1989.
[77] R. L. Gorsuch, *Factor analysis*, 2nd ed. Hillsdale: Erlbaum, 1983.
[78] E. Mooi and M. Sarstedt, *A concise guide to market research: the process, data, and methods using IBM SPSS Statistics*. Berlin: Springer-Verlag, 2011.
[79] W. J. Doll, W. Xia, and G. Torkzadeh, "A confirmatory factor analysis of the end-user computing satisfaction instrument," *MIS Quarterly: Management Information Systems*, vol. 18, no. 4, pp. 453–460, 1994, doi: 10.2307/249524.
[80] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis*, 8th Ed. Hampshire: Cengage Learning, EMEA, 2019.
[81] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a silver bullet," *Journal of Marketing Theory and Practice*, vol. 19, no. 2, pp. 139–152, 2011, doi: 10.2753/MTP1069-6679190202.
[82] Z. Awang, S. H. Lim, and N. F. S. Zainudin, *Pendekatan Mudah SEM- Structural Equation Modelling*. Bandar BaruBangi: MPWS Rich Resources, 2018.
[83] A. S. Shkeer and Z. Awang, "the impact of marketing information system components on organizational decision making: A case of jordanian five star hotels," *International Review of Management and Marketing*, vol. 9, no. 6, pp. 197–204, 2019, [Online]. Available: https://www.econjournals.com/index.php/irmm/article/view/8873.
[84] I. Ghozali, *Structural equation models: Concepts and applications with the AMOS 24 bayesian SEM update program* (Indonesian version). Semarang: Badan Penerbit Universitas Diponegoro, 2017.
[85] M. R. Ab Hamid, W. Sami, and M. H. Mohmad Sidek, "Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion," *Journal of Physics: Conference Series*, vol. 890, no. 1, 2017, doi: 10.1088/1742-6596/890/1/012163.

[86] R. F. DeVellis, *Scale development: Theory and applications*, 4th Ed. Los Angeles: Sage, 2016.
[87] C. M. Voorhees, M. K. Brady, R. Calantone, and E. Ramirez, "Discriminant validity testing in marketing: an Analysis, causes for concern, and proposed remedies," *Journal of the Academy of Marketing Science*, vol. 44, no. 1, pp. 119–134, 2016, doi: 10.1007/s11747-015-0455-4.
[88] N. M. Noor, A. A. Aziz, M. R. Mostapa, and Z. Awang, "Validation of the Malay version of the inventory of functional status after childbirth questionnaire," *BioMed Research International*, vol. 2015, 2015, doi: 10.1155/2015/972728.
[89] G. J. Geldhof, K. J. Preacher, and M. J. Zyphur, "Reliability estimation in a multilevel confirmatory factor analysis framework," *Psychological Methods*, vol. 19, no. 1, pp. 72–91, 2014, doi: 10.1037/a0032138.
[90] R. E. Zinbarg, W. Revelle, I. Yovel, and W. Li, "Cronbach's, α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability," *Psychometrika*, vol. 70, no. 1, pp. 123–133, 2005, doi: 10.1007/s11336-003-0974-7.
[91] D. Gefen, D. Straub, and M.-C. Boudreau, "Structural equation modeling and regression: Guidelines for research practice," *Communications of the Association for Information Systems*, vol. 4, 2000, doi: 10.17705/1cais.00407.
[92] C. Viladrich, A. Angulo-Brunet, and E. Doval, "A trip around alpha and omega to estimate internal consistency reliability," *Anales de Psicologia*, vol. 33, no. 3, pp. 755–782, 2017, doi: 10.6018/analesps.33.3.268401.
[93] F. Wulandari, D. Mardapi, and Haryanto, "The development of students' aptitude test in online and multimedia based interests group selection," *Journal of Physics: Conference Series*, vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012110.

BIOGRAPHIES OF AUTHORS

Muhammad Rais Ridwan is a Ph.D. Candidate, Study Program of Educational Research and Evaluation, Graduate School, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia and a Lecturer, Study Program of Mathematics Education, STKIP YPUP Makassar, Indonesia. His current research interest includes the development of instruments at various levels and areas of education. His publication topics include meta-analysis, learning of mathematics, and the development of an instrument to measure students' ability. His research focuses on applied mathematics and mathematics education. He can be reached at email: mrais.2019@student.uny.ac.id; mraisridwan@stkip.ypup.ac.id.

Samsul Hadi is a professor and lecturer at the Department of Electrical Engineering Education, Faculty of Engineering, Universitas Negeri Yogyakarta. He received a Ph.D. degree in research and evaluation education from the Universitas Negeri Yogyakarta. He is passionate about measuring, assessing, and evaluating vocational education. His research interests include students' instrument learning and development at various educational levels and areas. His publications cover topics such as vocational learning assessment and computer-based assessment development. He can be reached at email: samsul_hd@uny.ac.id.

Jailani received a Ph.D. degree in research and evaluation education from the Universitas Negeri Jakarta. He has worked as an Academician at Universitas Negeri Yogyakarta for over 36 years. He is currently a Professor and a lecturer at the Department of Mathematics Education, Faculty of Mathematics and Natural Sciences. He was appointed as an Assistant Professor in 2001 and as Associate Professor in 2005. His research interests include student learning and development at various educational levels and areas. His publications cover qualitative research, instrument development, and mathematics learning assessment. He can be reached at email: jailani@uny.ac.id.

Heri Retnawati is a professor and lecturer at the Study Program of Research and Evaluation Education, Graduate School, Universitas Negeri Yogyakarta. She was hired as a lecturer at the institution in 2000 after completing her master's degree in research and evaluation education at Universitas Negeri Yogyakarta. She was appointed as Assistant Professor in 2012 and Associate Professor in 2016. She is passionate about the measurement and assessment of education. Her current research focuses on students' mathematics learning and instrument development at many levels and domains of education. Her publications include qualitative research, meta-analysis, development of the instrument, and mathematics learning assessment. She can be reached at email: heri_retnawati@uny.ac.id.