Research Article https://doi.org/10.12973/eu-jer.11.3.1441

European Journal of Educational Research
Volume 11, Issue 3, 1441 - 1462.
ISSN: 2165-8714
https://www.eu-jer.com/

Developing Assessment Instrument Using Polytomous Response in Mathematics

Sugeng Sutiarso*
University of Lampung, INDONESIA

Undang Rosidin
University of Lampung, INDONESIA

Aan Sulistiawan
Vocational School, INDONESIA

Received: November 25, 2021▪ Revised: February 2, 2022▪ Accepted: May 17, 2022

Abstract: This research is a developmental research aiming at developing a good mathematical test instrument using polytomous responses based on classical and modern theories. This research design uses the Plomp model, which consists of five stages, (1) preliminary investigation, (2) design, (3) realization/construction, (4) revision, and (5) implementation (testing). The study was conducted in three vocational schools in Lampung Province, Indonesia. The study involved 413 students, consisting of 191 male and 222 female students. The data were collected through questionnaire and test. The questionnaire was used to identify the assessment instruments currently employed by teachers and to be validated by the experts of mathematics and educational evaluation. The test used an open polytomous response test numbering of 40 items. The data were analyzed using both classical and modern theories. The results show that (1) the open polytomous response test has a good category according to classical and modern theory. However, the discrimination power of test items in classical theory needs several revisions, (2) the assessment instrument using the polytomous response of open multiple choice can guarantee information on the actual competence of students. This is proven by the fact that there is a harmony between the analysis result obtained from classical and modern theory from the students' arguments when giving reasons for their choices. Therefore, the open polytomous response test can be used as an alternative to learning assessment.

Keywords: *Assessment instrument, classical and modern theory, vocational school, polytomous responses.*

To cite this article: Sutiarso, S., Rosidin, U., & Sulistiawan, A. (2022). Developing assessment instrument using polytomous response in mathematics. *European Journal of Educational Research, 11*(3), 1441-1462. https://doi.org/10.12973/eu-jer.11.3.1441

Introduction

Assessment is an important activity that needs to be administered by teachers in schools. The conventional paradigm often interprets assessment as a way to find out the achievement of student learning outcomes as a whole, so that the assessment is positioned as a separate activity from the learning process (Syaifuddin, 2020). Referring to the current paradigm, assessment in schools was divided into three types, e.g., assessment as learning, assessment for learning, and assessment of learning (Wulan, 2018). The three types of assessments aim to provide recognition of the achievement of student learning outcomes after the learning process (Earl, 2013). Below is the assessment pyramid (Figure 1).

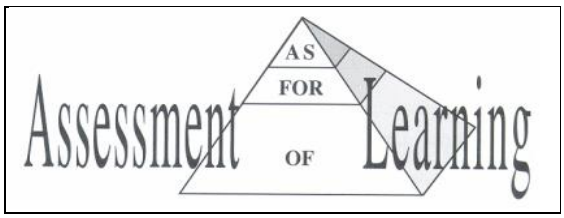


Figure 1. Assessment Pyramid

Assessment can be administered through a test. A test is a tool or procedure used to find out or measure students' abilities in particular areas with specific rules (Arikunto, 2012). A test consists of two types, namely multiple-choice and essay. A multiple-choice test is a form of assessment in which each item provides options, and one of the options is the correct answer. An essay test is a form of assessment that requires answers in sentences or words. Each type of test has its own strengths or weaknesses. The strengths of multiple-choice test over essay test are firstly, the test can be conducted for

***Corresponding author:**
Sugeng Sutiarso, University of Lampung, Sumantri Brojonegoro, Bandar Lampung, Indonesia. ✉ sugeng.sutiarso@fkip.unila.ac.id

many students, it is more objective, and the test results can be obtained more quickly. In addition to the strengths, it has several weaknesses. The multiple-choice test is not able to portray the actual abilities of students, the answers of this test tend to be guessing games or trial and error (Rosidin, 2017). In addition, the strength of the multiple-choice test has a scoring certainty compared to the essay test, namely 1 and 0 (score 1 for the correct answer, and 0 for the wrong answer). A multiple-choice test with only two answer choices is called a "dichotomous test," and a multiple-choice test with more than two answer choices is called a "polytomous test"(Kartono, 2008).

Until quite recently, multiple-choice test has been widely used by teachers to assess students' abilities, especially those with a large number and a wide area of expertise. To reduce the weakness of multiple-choice test, in the last four decades, experts have developed multiple-choice test by combining multiple-choice test and essay into multiple-choice test with reasons or hence forth called as polytomous response test (Suwarto, 2012). The polytomous response test score ranges from 1 – 4in which score 4 for the correct answer and reason, score 3 for the correct answer but the wrong reason, score 2 for the wrong answer but the correct reason, and score 1 for the wrong answer and reason (Kartono, 2008).

In the 1980s, the first test that was focused on and developed by experts was the closed polytomous response test, also known as the two-tier test (Treagust, 1988). This test consists of two levels: the first is choosing answers on the multiple-choice test, and the second level is choosing reasons based on the answer choices at the first level (Chandrasegaran et al., 2007). Several studies on the closed polytomous response test have been carried out, such as a test on mathematical ability in middle school (Hilton et al., 2013; Rovita et al., 2020), a test on calculus material (Khiyarunnisa & Retnawati, 2018), test on higher order thinking mathematical skills (Sundari, et al., 2021), and test on mathematical connection material (Lestari et al., 2021). Although the closed polytomous response test has been widely developed, researchers have found weaknesses in the test, such as students' misconceptions or students' actual competence cannot be identified in detail (Antara et al., 2019), the test instrument is difficult to construct (Khusnah, 2019), and student answers are still guessed (Myanda et al., 2020). However, there are strengths in the closed polytomous response test, such as the consistency of student answers errors, which is easily observed (Treagust, 1988), and the suitability between the student's answer choices and the reason is easy to know (Diani et al., 2019).

To reduce the weakness of the closed polytomous response test, the experts modified the test into an open polytomous response test. The open polytomous response test is a form of multiple-choice test that provides a place to write arguments for the answer choices (Retnawati, 2014). The studies on the open polytomous response test that have been carried out are test on calculus material in universities (Yang et al., 2017) and test on mathematics material in senior schools (Ayanwale, 2021) and junior schools (Falani et al., 2020). These studies developed open polytomous response tests for students in college and senior or junior school. Students in college and senior or junior school learn mathematics as a primary subject, while students in vocational schools learn mathematics as a secondary subject (Oktaria, 2016). In addition, students in vocational schools are more oriented towards practical abilities and skills, in contrast to students in college or high school who are more academically oriented, including in Indonesia (Permendikbud, 2016).

Currently, the Indonesian government expects that vocational schools is not academically left behind especially in mathematics. The government's commitment is to improve the method of assessing student learning in vocational school, and the current assessment method is using polytomous test. Often, students pay less attention during math exams for several reasons, such as considering mathematics as an unimportant subject (Putri et al., 2017), mathematics as a complicated subject (Vani et al., 2019), and mathematics as a boring subject (Ikmawati, 2020). Therefore, the students tend to answer the test by guessing. To avoid the tendencies, it is necessary to develop a polytomous response test (closed or open). By considering the disadvantages of the closed polytomous response test, it is reasonable to conduct research by developing open polytomous response test for students in vocational schools.

The test instrument developed must be reliable as a good test, and therefore it is necessary to analyze the quality of test items (Rosidin, 2017). There are two theories for analyzing the item quality, namely classical and modern. Classical theory is a measurement theory for assessing test based on the assumption of measurement errors between actual results and observations, and from the assumption, a formula for calculating the level of difficulty and item discrimination was developed (Hambleton & Jones, 1993). The modern theory is a measurement theory to assess students' abilities by comparing students' abilities with their group abilities, and it is known as Item Response Theory/IRT (Hambleton & Linden, 1982). Classical theory is widely used by teachers because it is easy to apply. However, this classical theory has a weakness, namely, it cannot separate the characteristics of students and items. The modern theory is a solution to overcome the weaknesses of the classical theory because, in the modern theory, an item does not affect other items (local independence), items only measure one dimension (unidimensional) (Anisa, 2013), and an item eliminates the relationship between respondents and items (parameter invariance) (Saepuzaman et al., 2021). Therefore, experts suggest that the test instrument is accountable; the quality of the items must be good according to the analysis of classical and modern theory (Retnawati, 2014).

Therefore, the aim of the research is to develop a good mathematical test instrument using polytomous responses according to classical and modern theories in vocational schools. The research problems are stated as follows: (1) Does the open polytomous response test developed have a good category so that it can be used as an assessment instrument

in vocational schools based on classical and modern theory? and (2) Does the open polytomous response test instrument developed provide information on students' actual competence in vocational schools?

Methodology

Research Design

This research is a research and development model that refers to Plomp's (2013) model, with the research procedure consisting of five stages: preliminary investigation, design, realization or construction, test phase, revision, and implementation (test).

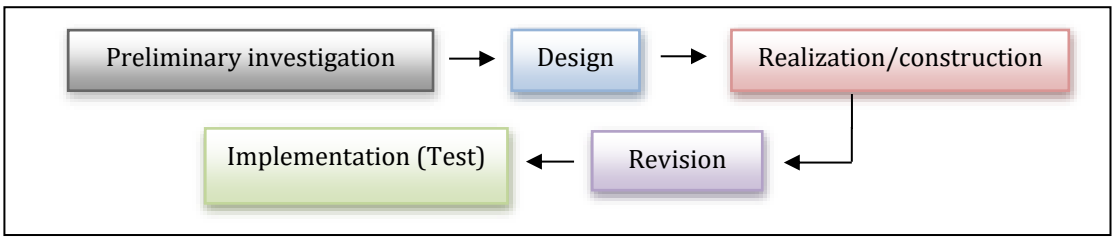


Figure 2. Stages of Research Design

The preliminary investigation stage is to identify the current assessment instruments used by teachers. The design stage is to make the open polytomous response test grid according to the basic competencies of mathematical concepts and to make a quality assessment questionnaire sheet. The realization/construction stage is developing the items test validated by an expert process for the items test. The revision stage is the improvement of items on the test based on expert advice. The implementation (testing) stage is to try out the test on students and to analyze the results of the test.

Research Subject

The subjects of the study were students of vocational schools situated in the province of Lampung, Indonesia. The research sample was determined using a non-probability sampling technique in the form of accidental sampling, which means taking a subject based on a subject that is easy to find and ready to be a respondent (Malhotra, 2006). The selected schools were three schools as representatives, namely the Mitra Bhakti vocational school, the Praja Utama vocational school, and the Ma'arif NU vocational school. The research subjects were 413 students in grade I (male students = 191, and female students = 222), whose mathematical abilities on the National Exam (NE) were categorized as moderate (average 64.67 out of the ideal score of 100). The following are the characteristics of the research subjects in detail shown in Table 1.

Table 1. Research Subjects

Vocational Schools	Grade	The Number of Students			Average of NE
		Total	Male	Female	
Mitra Bhakti	I-A	13	5	8	64.24
	I-B	23	10	13	62.4
	I-C	29	9	20	65
Praja Utama	I-A	40	40	0	64
	I-B	39	39	0	64.62
	I-C	43	15	28	68.72
	I-D	38	15	23	67.9
	I-E	41	14	27	67.56
	I-F	40	17	23	63.38
	I-G	40	8	23	67.74
Ma'arif NU	I-A	31	14	17	62.1
	I-B	36	5	51	58.34
Total		413	191	222	64.67

Data Collection Techniques

Data were collected using a questionnaire and a test. The questionnaire contains several questions to the teacher about the instrument for assessment used by the teacher and also expert validity of the instrument developed to determine content validity (Suhaini et al., 2021). The instrument was validated by two raters who have expertise in mathematics and educational evaluation. The three aspects assessed by the expert were the suitability of the items with the indicators, language, and alternative answers to the questions. The score validating the instrument follows the criteria as in Table 2 below.

Table 2. Content Validity Score

Description	Score
Irrelevant	1
Quite relevant	2
Relevant	3
Very relevant	4

After determining the content validity, the instrument was tested to students. Then, it was continued by determining the validity of the construct and its reliability to ensure that the instrument could be further analyzed.

The instrument used was the open polytomous response test, which consisted of 40 questions on the concepts of sequences and series (arithmetic and geometry), quadratic equations and functions, and matrices. Each item contained five answer choices along with the reasons. Student scores are referred to the polytomous score in the Partial Credit Model, where answer choices and reasons were related (Retnawati, 2014), as shown in Table 3 below.

Table 3. Scores for Student Answers

Student Answers		Score
Answer Options	**Reason**	
False	False	1
False	Right	2
Right	False	3
Right	Right	4

Data Analysis

The collected data were analyzed in two stages: (1) questionnaire data analysis (qualitative analysis) and (2) test data analysis (empirical analysis). The following is an explanation of each data analysis:

1. *Questionnaire data analysis (qualitative analysis)*

There are two sets of questionnaire data namely, identification of assessment instruments in schools and expert assessment of the assessment instruments developed. The results of the two questionnaires were analyzed descriptively. Specifically, for expert judgment, it was continued with an analysis of expert agreement that used the Gregory index formula (Gregory, 2015), namely:

$$V = \frac{D}{A+B+C+D}$$

Description:

V = Content Validity

A = Rater 1 and 2 are weak

B = Rater 1 is strong, and rater 2 is weak

C = Rater 1 is weak, and rater 2 is strong

D = Rater 1 and 2 are strong

The interpretation of Gregory's formula is that the number V is in the range of 0 to 1. The higher the number V (close to 1 or equal to 1), the higher the value of the validity of an item. Conversely, the lower the number V (close to 0 or equal to 0), the lower the validity of an item.

2. *Test data analysis (empirical analysis)*

After conducting the content validity test, the researchers conducted the construct validity and reliability test. The construct validation test used exploratory factor analysis. The instrument is considered to having good construct if the explained Kaiser-Meyer-Olkin (KMO) value is greater than 0.5 (Retnawati, 2014). Reliability test using Cronbach's alpha formula. The instrument is said to have good reliability if the coefficient value of Cronbach's alpha is 0.60 (Arikunto, 2012). If the instrument has good construct validation, further test can be analyzed, namely the level of difficulty and item discrimination. The reason for analyzing the level of difficulty and item discrimination is that they are both preliminary analyses of the assumptions of measurement theory (Hambleton & Jones, 1993). To simplify the process of analyzing the level of difficulty and item discrimination, the Iteman program was used for classical theory and the Winsteps program for modern theory (Sarea & Ruslan, 2019). The Winsteps program was used because it had several

advantages; namely, it can analyze polytomous data and calculate the maximum likelihood model using a 1-parameter logistic model (Untary et al., 2020).

2.1. Analysis of test data with classical theory

a. The item difficulty level is the percentage of the number of students who answered correctly or incorrectly. If the item has an index of 0.3-0.7, then the item is good; if the item has an index below 0.3, then the item is difficult; and if the item has an index above 0.7, then the item is easy.

b. Discrimination is the ability of a test to distinguish between high-ability students and low-ability students. Discrimination is said to be good if it has an index above 0.3, and if the discrimination index is below 0.3, then the question needs to be revised (Arikunto, 2012).

2.2. Analysis of test data with modern theory

a. The item difficulty level is the level of the student's latent trait towards the item. The difficulty of the items determines the ability of about 50% of the respondents who are expected to answer items correctly (DeMars, 2010). An item is said to be good if it has an index of between -2 and +2 (Hambleton & Swaminathan, 1985). If the index is close to -2, then the item is classified as very easy, and if the index is close to +2, the item is classified as very difficult (Retnawati, 2014). In the Winsteps program, the item difficulty level is in the Measure column.

b. Item discrimination is indicated by the slope of the curve on the item characteristics. The item is said to have good discrimination if the slope of the curve is moderate (not too gentle or steep) because if the slope of the curve is too gentle or steep, then the item is not good. Another opinion states that a good index is above 0.4 (Crocker & Algina, 1986). In the Winsteps program, the item discrimination is in the Pt-Measure Correlation column.

According to modern theory, before analyzing the item difficulty level and discrimination, three assumptions must be tested, namely unidimensionality, local independence, and model fit (Hambleton et al., 1991). Unidimensional means that each test item only measures one ability. There are three ways that are often used to test unidimensionality, namely the analysis of the Eigenvalue of the correlation matrix between items, the Stout-test on the unidimensional assumption test, and the index based on the residuals of the unidimensional solution (DeMars, 2010). In this study, the dimensional test used the Eigenvalue analysis of the correlation matrix between items.

Local independence is the state of the respondent's answer to an item that is not influenced by other items. Local independence test by proving that the probability of the respondent's answer pattern is the same as the probability of the respondent's answer to each item. If the unidimensional assumption is accepted, the local independence assumption will also be accepted (DeMars, 2010). Use the Model Fit Test to find out whether the model used is in accordance with the items. Test the fit of the model by measuring the outfit mean square (MNSQ) and PT-Measure. If the outfit's MNSQ value is 0.5 to 1.5 and the Pt-Measure Correlation is positive, it is said that the item fits the model (Linacre, 2012). In addition, the information function and standard error measurement (SEM) are analyzed, which aims to further explain the latent ability as measured by using a test that is expressed through item donations.

Results

Analysis of Questionnaire Data

Based on the results of the questionnaire, it was found that the teacher had never used the polytomous response. As many as 80% of teachers used essay tests and 20% of teachers used multiple-choice tests, with each instrument consisting of 2-5 items. In addition, about 10% of teachers used this assessment as a means for learning improvement, such as improving lesson plans and teaching methods. The results of the questionnaire stated that 90% of teachers who did not use assessment as an improvement in learning were caused by several aspects, such as teachers did not understand assessment (20%), teachers did not know how to analyze assessments (50%), and teachers did not know how to develop good assessment questions (30%). The following is the summary of the questionnaire from the identification data.

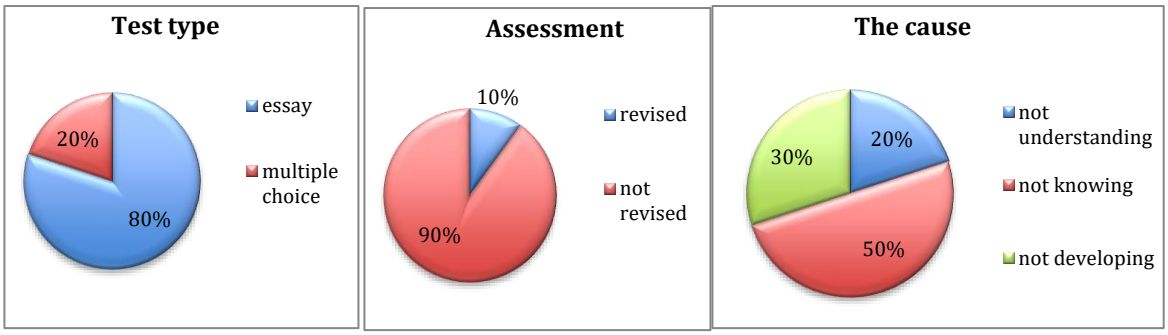


Figure 3. Description of Teacher Conditions in Assessment

Content Validity

The results of the two expert assessments showed that the content validation instrument is good. Furthermore, the analysis of expert judgment agreement was obtained as shown in Table 4 below.

Table 4. The Gregory Item Index

		Rater 1	
		Weak	**strong**
Rater 2	weak	0	0
	strong	0	40
Index Gregory		1	

Based on the results of the assessment in Table 4, it can be concluded that the instrument is valid because the value of V reaches a value of 1. Therefore, the instrument test can be continued. In addition to providing assessments, the experts also provided some suggestions for improvements to the instrument, namely the preparation of questions using the ABCD format (Audience, Behavior, Competence, and Degree), avoiding the use of ambiguous language or statements, improving mathematical concepts, making alternative answer choices that are misleading, and arranging them in order.

Analysis of Test Data

Construct Validity

After testing the instrument, it was followed by a construct validity test. The results of the test with exploratory factor analysis are shown in Table 5.

Table 5. Exploratory Factor Analysis

KMO and Bartlett's Test		**Score**
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.760
Bartlett's Test of Sphericity	Approx. Chi-Square	1936.378
	Df	780
	Sig.	.000

Based on Table 5, the explained KMO value is 0.76 (more than 0.5), and it can be concluded that the variables and samples used to allow for further analysis.

Reliability

The results of the estimation of the reliability of the instrument obtained a Cronbach's alpha coefficient value is 0.89 (more than 0.6). It means that the instrument has good reliability so that the analysis of the level of difficulty and item discrimination can be continued according to classical and modern methods.

Table 6. Reliability of Items

Cronbach's Alpha	**N of Items**
0.892	40

Analysis of Test Data with Classical Theory

Analysis of test data in the classical way did not require testing assumptions, but the analysis of the difficulty level of the items and the distinguishing power of items could be directly calculated if the validity and reliability have been met. The results of the two analyses are presented in Table 7 below.

Table 7. The Item Difficulty Level and Discrimination

Item	Difficulty Level	Category	Discrimination	Category	Item	Difficulty Level	Category	Discrimination	Category
1	0.538	Good	0.196	Revised	21	0.482	Good	0.143	Revised
2	0.487	Good	0.179	Revised	22	0.535	Good	0.429	Good
3	0.528	Good	0.214	Revised	23	0.492	Good	0.250	Revised
4	0.540	Good	0.304	Good	24	0.438	Good	-0.071	Revised
5	0.489	Good	0.089	Revised	25	0.436	Good	-0.107	Revised
6	0.446	Good	-0.161	Revised	26	0.385	Good	-0.286	Revised
7	0.438	Good	-0.232	Revised	27	0.383	Good	-0.321	Revised
8	0.453	Good	-0.143	Revised	28	0.416	Good	-0.143	Revised
9	0.414	Good	-0.143	Revised	29	0.458	Good	-0.125	Revised
10	0.409	Good	-0.339	Revised	30	0.385	Good	-0.375	Revised
11	0.438	Good	-0.143	Revised	31	0.404	Good	-0.321	Revised
12	0.436	Good	-0.036	Revised	32	0.433	Good	-0.250	Revised
13	0.400	Good	-0.321	Revised	33	0.441	Good	0,036	Revised
14	0.450	Good	-0.036	Revised	34	0.424	Good	-0.268	Revised
15	0.462	Good	0.250	Revised	35	0.412	Good	-0.321	Revised
16	0.453	Good	-0.089	Revised	36	0.431	Good	-0.304	Revised
17	0.416	Good	-0.143	Revised	37	0.404	Good	-0.232	Revised
18	0.419	Good	-0.196	Revised	38	0.363	Good	-0.482	Revised
19	0.431	Good	-0.232	Revised	39	0.230	Good	-0.929	Revised
20	0.441	Good	-0.089	Revised	40	0.211	Good	-1.071	Revised

Based on Table 7, it was found that all items have an item difficulty level in the index range of 0.3 to 0.7, so they are categorized in the good category. Meanwhile, only two items on discrimination had good categories, and the remaining items needed to be revised. The results indicated that all items were good based on the level of difficulty, but almost all items needed to be revised for item discrimination.

Analysis of Test Data with Modern Theory

The Unidimensional Assumption Test

The unidimensional assumption test is the first assumption test with factor analysis. Factor analysis begins by testing the adequacy of the sample to be used in the analysis, constructing a variance-covariance matrix, and then calculating the Eigenvalue. The Eigenvalue was then used to calculate the percentage of explained variance, as well as to describe the scree plot (Retnawati, 2014). The output of factor analysis was the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) statistic and scree plot.

Table 8. The KMO Test

KMO Test	Score
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	.760
Sig.	.000

The unidimensional test is seen based on the cumulative percentage of Eigenvalue and scree plot analysis results. If the cumulative percentage of the first-factor Eigenvalue is greater than 20%, then the unidimensional assumption is fulfilled (Retnawati, 2014). In Table 9, it can be seen that the cumulative percentage of the first-factor Eigenvalue is 20.220%. Because the Eigenvalue is more than 20%, this instrument is proven to only measure one factor or dimension.

Table 9. Total Variance Explained

Component	Initial Eigenvalue		
	Total	Variance (%)	Cumulative (%)
1	8.088	20.220	20.220
2	1.458	3.646	23.865

In addition, the unidimensional test can also be seen on the scree plot, which is based on the number of factors marked by the steepness of the graph with the acquisition of Eigenvalue.

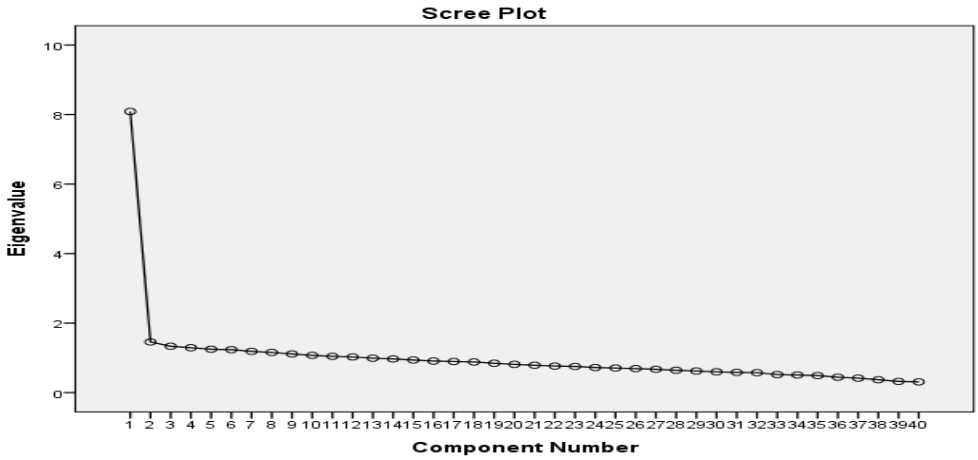


Figure 4. Scree Plot Unidimensional

Based on the scree plot, it is known that the Eigenvalue immediately slope on the second factor. It demonstrates that the developed instrument has only one dominant factor. The results prove that the test kit meets the unidimensional assumption, or in other words, only measures one dominant factor.

Local Independence Assumption Test

The local independence assumption test will be fulfilled if the student's answer to one item does not affect the student's answer to another item. Thus, the score of one item should not be determined or dependent on the scores of other items. This confirms that this assumption automatically proves that students' answers do not affect answers to other items (Retnawati, 2014).

Table 10. The Covariance Matrix.

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
K1	0.0486									
K2	0.012	0.0042								
K3	0.0105	0.0034	0.0032							
K4	0.0317	0.0106	0.0093	0.0293						
K5	0.01	0.0032	0.0029	0.0085	0.0028					
K6	0.0069	0.0023	0.0021	0.0062	0.002	0.0016				
K7	0.0036	0.0011	0.0009	0.0028	0.0009	0.0007	0.0004			
K8	0.0041	0.0015	0.0012	0.0037	0.0011	0.0008	0.0004	0.0006		
K9	0.0074	0.0024	0.0022	0.0065	0.002	0.0015	0.0007	0.0009	0.0017	
K10	0.021	0.0061	0.0054	0.0148	0.0051	0.0039	0.0017	0.0025	0.0043	0.0196

Table 10 shows the results of the variance-covariance values between groups of students' abilities. In the table, it can be seen that the covariance value between the ability interval groups located on the diagonal is small and close to zero. This result shows that there is no correlation, so it can be said that the local independence assumption test is accepted.

Model Fit

The model fit test was analyzed using the Winsteps program. The item requirements are called "fit to the model."If the Outfit MNSQ value is 0.5 to 1.5 and the Outfit ZSTD value is -2 to 2, and the Pt-Measure Correlation is positive, then it can be said that the item fits the model (Sumintono & Widhiarso, 2015). An item is considered fit if one of the conditions is accepted. In addition, it can also be seen from the MNSQ infit of 0.77 to 1.3, but at this stage, the fit of the model is only taken on the MNSQ and Pt-Measure outfit values. Based on the results of the analysis, all items matched the model or fit (Figure 5).

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
1	1134	413	-.70	.07	1.16	2.5	1.16	2.5	.74	.43	35.1	47.9	Q1
2	1079	413	-.45	.07	.97	-.4	.98	-.3	.75	.43	49.2	49.1	Q2
3	1066	413	-.39	.07	.93	-1.0	.94	-1.0	.67	.43	49.4	49.5	Q3
4	1064	413	-.38	.07	.82	-3.0	.82	-3.0	.46	.43	50.1	49.6	Q4
5	1021	413	-.19	.07	.84	-2.6	.84	-2.6	.41	.44	53.5	50.6	Q5
6	998	413	-.09	.07	1.16	2.4	1.16	2.4	.44	.44	42.4	50.9	Q6
7	977	413	.01	.07	1.03	.5	1.03	.5	.45	.44	40.7	51.1	Q7
8	966	413	.06	.07	1.16	2.4	1.16	2.5	.51	.44	42.1	51.1	Q8
9	965	413	.06	.07	.97	-.5	.97	-.5	.43	.44	49.9	51.0	Q9
10	964	413	.07	.07	.93	-1.0	.93	-1.0	.40	.44	49.2	51.0	Q10
11	986	413	-.03	.07	.97	-.4	.98	-.4	.38	.44	44.1	51.0	Q11
12	988	413	-.04	.07	.90	-1.7	.90	-1.7	.39	.44	52.3	51.0	Q12
13	957	413	.10	.07	1.07	1.1	1.07	1.1	.44	.44	47.5	51.0	Q13
14	1003	413	-.11	.07	1.01	.2	1.01	.3	.42	.44	48.7	50.9	Q14
15	1000	413	-.09	.07	1.10	1.5	1.10	1.6	.49	.44	42.9	50.9	Q15
16	994	413	-.07	.07	1.08	1.2	1.08	1.2	.44	.44	46.5	51.0	Q16
17	964	413	.07	.07	.96	-.7	.96	-.6	.48	.44	52.8	51.0	Q17
18	972	413	.03	.07	1.12	1.8	1.12	1.8	.35	.44	43.1	51.1	Q18
19	974	413	.02	.07	1.04	.7	1.05	.8	.26	.44	47.7	51.1	Q19
20	997	413	-.08	.07	.97	-.4	.97	-.4	.47	.44	50.8	50.9	Q20
21	1010	413	-.14	.07	1.01	.3	1.02	.3	.44	.44	47.5	50.8	Q21
22	1050	413	-.32	.07	1.08	1.2	1.08	1.2	.56	.44	43.6	50.0	Q22
23	1037	413	-.26	.07	1.05	.8	1.05	.8	.53	.44	47.2	50.3	Q23
24	974	413	.02	.07	1.06	1.0	1.07	1.1	.37	.44	46.7	51.1	Q24
25	968	413	.05	.07	1.00	.0	1.00	.0	.41	.44	47.9	51.1	Q25
26	943	413	.16	.07	.97	-.4	.97	-.4	.31	.44	50.8	50.9	Q26
27	936	413	.20	.07	1.02	.3	1.02	.4	.32	.44	46.0	50.8	Q27
28	975	413	.02	.07	.89	-1.8	.88	-1.9	.40	.44	54.5	51.1	Q28
29	981	413	-.01	.07	.97	-.5	.97	-.5	.33	.44	49.2	51.0	Q29
30	949	413	.14	.07	1.06	.9	1.05	.9	.30	.44	45.5	50.9	Q30
31	955	413	.11	.07	1.07	1.1	1.07	1.1	.23	.44	44.8	51.0	Q31
32	974	413	.02	.07	1.12	1.8	1.12	1.8	.35	.44	43.3	51.1	Q32
33	981	413	-.01	.07	1.04	.6	1.04	.6	.41	.44	48.2	51.0	Q33
34	959	413	.09	.07	1.06	1.0	1.07	1.1	.33	.44	46.5	51.0	Q34
35	955	413	.11	.07	1.04	.7	1.04	.7	.40	.44	47.2	51.0	Q35
36	961	413	.08	.07	1.04	.6	1.05	.8	.34	.44	47.7	51.0	Q36
37	955	413	.11	.07	.96	-.6	.96	-.6	.40	.44	51.1	51.0	Q37
38	921	413	.27	.07	.89	-1.7	.89	-1.8	.43	.44	52.8	50.6	Q38
39	825	413	.72	.07	.65	-6.3	.66	-6.2	.59	.43	58.1	48.6	Q39
40	801	413	.84	.07	.74	-4.5	.74	-4.6	.54	.43	55.4	48.2	Q40
MEAN	979.5	413.0	.00	.07	1.00	-.1	1.00	-.1			47.8	50.6	
S.D.	56.8	.0	.26	.00	.11	1.8	.11	1.8			4.3	.8	

Figure 5. Item Fit on Model

The Item Difficulty Level

The item difficulty level was analyzed using the Winsteps program, and the results obtained were presented in Table 12 (Measure column). The item difficulty level is in the range of -0.70 to 0.84. From Table 12, the highest difficulty level is item 40 (difficulty level 0.84), and the lowest difficulty level is item 1 (difficulty level -0.70). Since the difficulty level is in the range of 2 and 2, it can be concluded that all items are in the good category. If further divided into three categories, the difficulty level of items in the range of -0.7 to 0.84 is moderate (Sumintono & Widhiarso, 2015). It can also be seen on the difficult map items, namely that the difficulty level is in the range of -2 and 2.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
1	1134	413	-.70	.07	1.16	2.5	1.16	2.5	.74	.43	35.1	47.9	Q1
2	1079	413	-.45	.07	.97	-.4	.98	-.3	.75	.43	49.2	49.1	Q2
3	1066	413	-.39	.07	.93	-1.0	.94	-1.0	.67	.43	49.4	49.5	Q3
4	1064	413	-.38	.07	.82	-3.0	.82	-3.0	.46	.43	50.1	49.6	Q4
5	1021	413	-.19	.07	.84	-2.6	.84	-2.6	.41	.44	53.5	50.6	Q5
6	998	413	-.09	.07	1.16	2.4	1.16	2.4	.44	.44	42.4	50.9	Q6
7	977	413	.01	.07	1.03	.5	1.03	.5	.45	.44	40.7	51.1	Q7
8	966	413	.06	.07	1.16	2.4	1.16	2.5	.51	.44	42.1	51.1	Q8
9	965	413	.06	.07	.97	-.5	.97	-.5	.43	.44	49.9	51.0	Q9
10	964	413	.07	.07	.93	-1.0	.93	-1.0	.40	.44	49.2	51.0	Q10
11	986	413	-.03	.07	.97	-.4	.98	-.4	.38	.44	44.1	51.0	Q11
12	988	413	-.04	.07	.90	-1.7	.90	-1.7	.39	.44	52.3	51.0	Q12
13	957	413	.10	.07	1.07	1.1	1.07	1.1	.44	.44	47.5	51.0	Q13
14	1003	413	-.11	.07	1.01	.2	1.01	.3	.42	.44	48.7	50.9	Q14
15	1000	413	-.09	.07	1.10	1.5	1.10	1.6	.49	.44	42.9	50.9	Q15
16	994	413	-.07	.07	1.08	1.2	1.08	1.2	.44	.44	46.5	51.0	Q16
17	964	413	.07	.07	.96	-.7	.96	-.6	.48	.44	52.8	51.0	Q17
18	972	413	.03	.07	1.12	1.8	1.12	1.8	.35	.44	43.1	51.1	Q18
19	974	413	.02	.07	1.04	.7	1.05	.8	.26	.44	47.7	51.1	Q19
20	997	413	-.08	.07	.97	-.4	.97	-.4	.47	.44	50.8	50.9	Q20
21	1010	413	-.14	.07	1.01	.3	1.02	.3	.44	.44	47.5	50.8	Q21
22	1050	413	-.32	.07	1.08	1.2	1.08	1.2	.56	.44	43.6	50.0	Q22
23	1037	413	-.26	.07	1.05	.8	1.05	.8	.53	.44	47.2	50.3	Q23
24	974	413	.02	.07	1.06	1.0	1.07	1.1	.37	.44	46.7	51.1	Q24
25	968	413	.05	.07	1.00	.0	1.00	.0	.41	.44	47.9	51.1	Q25
26	943	413	.16	.07	.97	-.4	.97	-.4	.31	.44	50.8	50.9	Q26
27	936	413	.20	.07	1.02	.3	1.02	.4	.32	.44	46.0	50.8	Q27
28	975	413	.02	.07	.89	-1.8	.88	-1.9	.40	.44	54.5	51.1	Q28
29	981	413	-.01	.07	.97	-.5	.97	-.5	.33	.44	49.2	51.0	Q29
30	949	413	.14	.07	1.06	.9	1.05	.9	.30	.44	45.5	50.9	Q30
31	955	413	.11	.07	1.07	1.1	1.07	1.1	.23	.44	44.8	51.0	Q31
32	974	413	.02	.07	1.12	1.8	1.12	1.8	.35	.44	43.3	51.1	Q32
33	981	413	-.01	.07	1.04	.6	1.04	.6	.41	.44	48.2	51.0	Q33
34	959	413	.09	.07	1.06	1.0	1.07	1.1	.33	.44	46.5	51.0	Q34
35	955	413	.11	.07	1.04	.7	1.04	.7	.40	.44	47.2	51.0	Q35
36	961	413	.08	.07	1.04	.6	1.05	.8	.34	.44	47.7	51.0	Q36
37	955	413	.11	.07	.96	-.6	.96	-.6	.40	.44	51.1	51.0	Q37
38	921	413	.27	.07	.89	-1.7	.89	-1.8	.43	.44	52.8	50.6	Q38
39	825	413	.72	.07	.65	-6.3	.66	-6.2	.59	.43	58.1	48.6	Q39
40	801	413	.84	.07	.74	-4.5	.74	-4.6	.54	.43	55.4	48.2	Q40
MEAN	979.5	413.0	.00	.07	1.00	-.1	1.00	-.1			47.8	50.6	
S.D.	56.8	.0	.26	.00	.11	1.8	.11	1.8			4.3	.8	

Figure 6. Item Difficulty Level

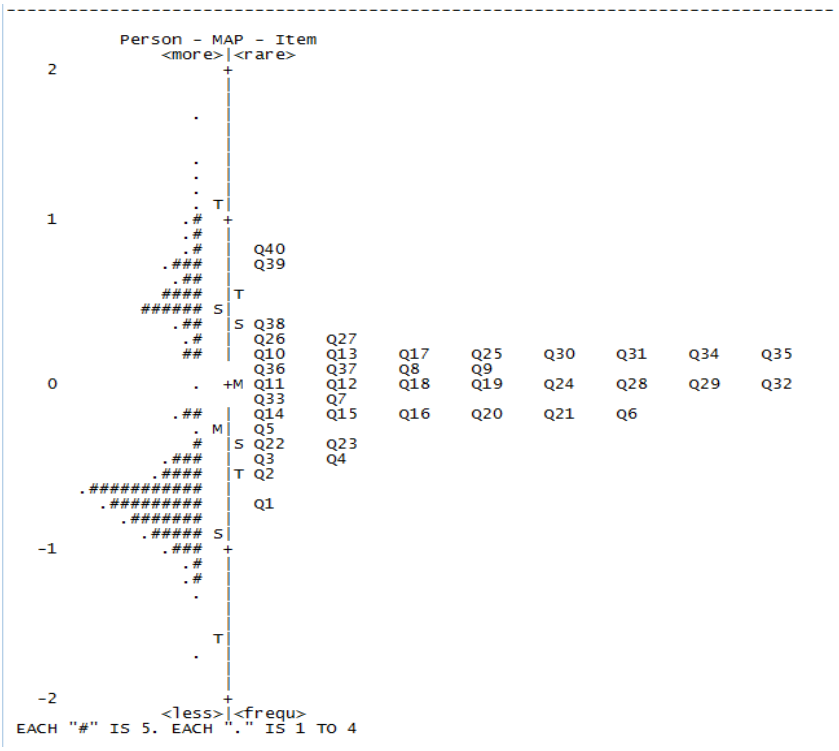


Figure 7. Item Difficulty Map

The Item Discrimination

The item discrimination analysis used the Winsteps program, and the results are presented in Figure 8 (PT-Measure Correlation column). Item discrimination is in the range of 0.23 to 0.74, with details of 27 items having an index above 0.4 (good category), and 13 items having an index below 0.4 (bad category). However, the 13 items can be used as an instrument as long as the index is above 0 (Alagumalai et al., 2005).

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
1	1134	413	-.70	.07	1.16	2.5	1.16	2.5	.74	.43	35.1	47.9	Q1
2	1079	413	-.45	.07	.97	-.4	.98	-.3	.75	.43	49.2	49.1	Q2
3	1066	413	-.39	.07	.93	-1.0	.94	-1.0	.67	.43	49.4	49.5	Q3
4	1064	413	-.38	.07	.82	-3.0	.82	-3.0	.46	.43	50.1	49.6	Q4
5	1021	413	-.19	.07	.84	-2.6	.84	-2.6	.41	.44	53.5	50.6	Q5
6	998	413	-.09	.07	1.16	2.4	1.16	2.4	.44	.44	42.4	50.9	Q6
7	977	413	.01	.07	1.03	.5	1.03	.5	.45	.44	40.7	51.1	Q7
8	966	413	.06	.07	1.16	2.4	1.16	2.5	.51	.44	42.1	51.1	Q8
9	965	413	.06	.07	.97	-.5	.97	-.5	.43	.44	49.9	51.0	Q9
10	964	413	.07	.07	.93	-1.0	.93	-1.0	.40	.44	49.2	51.0	Q10
11	986	413	-.03	.07	.97	-.4	.98	-.4	.38	.44	44.1	51.0	Q11
12	988	413	-.04	.07	.90	-1.7	.90	-1.7	.39	.44	52.3	51.0	Q12
13	957	413	.10	.07	1.07	1.1	1.07	1.1	.44	.44	47.5	51.0	Q13
14	1003	413	-.11	.07	1.01	.2	1.01	.3	.42	.44	48.7	50.9	Q14
15	1000	413	-.09	.07	1.10	1.5	1.10	1.6	.49	.44	42.9	50.9	Q15
16	994	413	-.07	.07	1.08	1.2	1.08	1.2	.44	.44	46.5	51.0	Q16
17	964	413	.07	.07	.96	-.7	.96	-.6	.48	.44	52.8	51.0	Q17
18	972	413	.03	.07	1.12	1.8	1.12	1.8	.35	.44	43.1	51.1	Q18
19	974	413	.02	.07	1.04	.7	1.05	.8	.26	.44	47.7	51.1	Q19
20	997	413	-.08	.07	.97	-.4	.97	-.4	.47	.44	50.8	50.9	Q20
21	1010	413	-.14	.07	1.01	.3	1.02	.3	.44	.44	47.5	50.8	Q21
22	1050	413	-.32	.07	1.08	1.2	1.08	1.2	.56	.44	43.6	50.0	Q22
23	1037	413	-.26	.07	1.05	.8	1.05	.8	.53	.44	47.2	50.3	Q23
24	974	413	.02	.07	1.06	1.0	1.07	1.1	.37	.44	46.7	51.1	Q24
25	968	413	.05	.07	1.00	.0	1.00	.0	.41	.44	47.9	51.1	Q25
26	943	413	.16	.07	.97	-.4	.97	-.4	.31	.44	50.8	50.9	Q26
27	936	413	.20	.07	1.02	.3	1.02	.4	.32	.44	46.0	50.8	Q27
28	975	413	.02	.07	.89	-1.8	.88	-1.9	.40	.44	54.5	51.1	Q28
29	981	413	-.01	.07	.97	-.5	.97	-.5	.33	.44	49.2	51.0	Q29
30	949	413	.14	.07	1.06	.9	1.05	.9	.30	.44	45.5	50.9	Q30
31	955	413	.11	.07	1.07	1.1	1.07	1.1	.23	.44	44.8	51.0	Q31
32	974	413	.02	.07	1.12	1.8	1.12	1.8	.35	.44	43.3	51.1	Q32
33	981	413	-.01	.07	1.04	.6	1.04	.6	.41	.44	48.2	51.0	Q33
34	959	413	.09	.07	1.06	1.0	1.07	1.1	.33	.44	46.5	51.0	Q34
35	955	413	.11	.07	1.04	.7	1.04	.7	.40	.44	47.2	51.0	Q35
36	961	413	.08	.07	1.04	.6	1.05	.8	.34	.44	47.7	51.0	Q36
37	955	413	.11	.07	.96	-.6	.96	-.6	.40	.44	51.1	51.0	Q37
38	921	413	.27	.07	.89	-1.7	.89	-1.8	.43	.44	52.8	50.6	Q38
39	825	413	.72	.07	.65	-6.3	.66	-6.2	.59	.43	58.1	48.6	Q39
40	801	413	.84	.07	.74	-4.5	.74	-4.6	.54	.43	55.4	48.2	Q40
MEAN	979.5	413.0	.00	.07	1.00	-.1	1.00	-.1			47.8	50.6	
S.D.	56.8	.0	.26	.00	.11	1.8	.11	1.8			4.3	.8	

Figure 8. Item Discrimination

Comparative Analysis Between Classical and Modern Theory

The results of the analysis of classical and modern theories obtained the index of difficulty level and item discrimination as follows.

Table 11. Comparison of Classical and Modern Theories

Parameter	Classical Theory		Modern Theory	
	Many Items with Good Categories	Percentage	Many Items with Good Categories	Percentage
Difficulty level	40	100	40	100
Item discrimination	2	0.05	27	67.5

Based on Table 11, the level of item difficulty analyzed by classical and modern theory has the same results (good category). However, item discrimination using modern theory has more items in the good category than classical theory. If we compare the index of discriminatory items between classical (Table 7) and modern (Figures 6 and 8), it can be seen that there is a match between the categories of item discrimination. It means that if item discrimination is not good with the classical theory, then item discrimination is also not good with the modern theory (13 items correspond to the item discrimination index, and 27 do not match).

Information and Measurement Error (SEM) Function

The function of information to reveal latent ability was measured by using a test that was expressed through item donation. The test information function is also the sum of the functions of each item. The information function is inversely proportional to measurement error, or standard error measurement (SEM). The value of the information function of the test device will be high if the items that make up the test have a high information function. The following is a picture of the curve of the relationship between the information function and SEM.

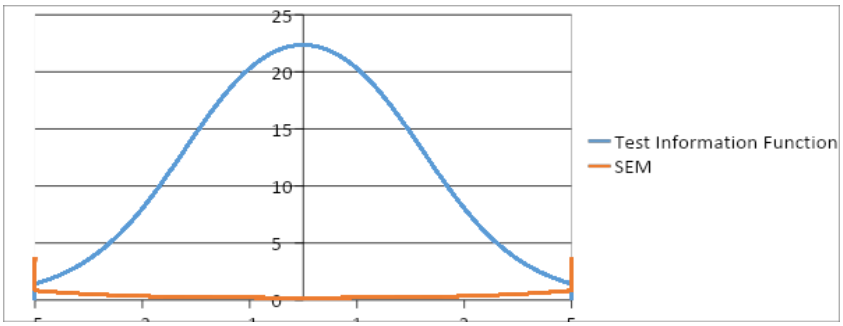


Figure 9. Graph of Information Function and Measurement Error

Figure 9 shows that this instrument provides information of 22.36 (maximum) and has a measurement error of 0.21 (smallest) for medium-ability students. The lower and upper limits of the interval are the ability scores where the graph of the information function and the SEM graph intersect at that interval. This graph states that the greater the value of the information function, the smaller the measurement error (SEM), and the item information function expresses the strength or contribution of the test items in revealing the latent trait as measured by the test. This information function provides a description of the item according to the model (which helps with item selection) (Retnawati, 2014). These results conclude that this test instrument is suitable for students with medium abilities.

Based on the test results, it can also be seen that the actual competency information of students in the vocational school was based on the test answers. Instrument analysis is based on two patterns of student answers that have the same tendency based on Bloom's Taxonomy (Bloom, 1956). A total of six student answers were selected as samples with different abilities (high, medium, and low).

Item 1: The cognitive domain to be achieved is C2 (Understanding). Based on students' answers, the two dominant patterns of student answers are: (1) students have understood the general form of arithmetic sequences and know the first term and other terms, and (2) students have not been able to formulate general forms in arithmetic sequences and perform algebraic operations on general forms arithmetic sequence.

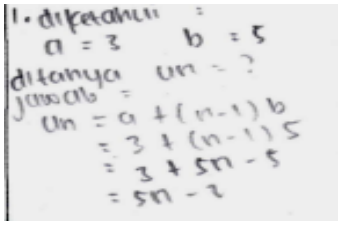
Question 1:

Given an arithmetic sequence: 3, 8, 13, 18, ...
The formula for the nth term of the sequence is ....

A. $U_n = 5n - 3$
B. $U_n = 5n - 2$
C. $U_n = 2n + 1$
D. $U_n = 4n - 1$
E. $U_n = 3n + 2$

Reason: .....

Pattern 1:



Pattern 2:

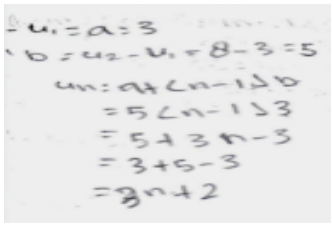


Figure 10. An Example of Student Answers in Item 1

Item 2: The cognitive domain to be achieved is C2 (understanding). Based on students' answers, the two dominant patterns of student answers are (1) being able to understand the number of terms in a sequence by using the general formula for an arithmetic sequence, or determining the number of arithmetic sequences without using a general formula (only writing down all the terms from the first term until the last term) and (2) students who can already use the general formula for arithmetic sequences but have not been able to determine the number of arithmetic sequences because of errors in performing algebraic operations on general arithmetic sequences.

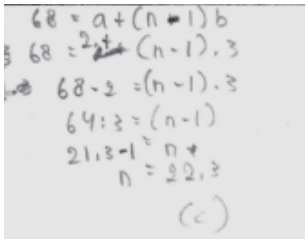
Question2:

Given an arithmetic sequence: 2, 5, 8, 11, ..., 68.
The number of terms in the sequence is…

A. 12
B. 13
C. 22
D. 23
E. 24

Reason: .....

Pattern 1



Pattern 2

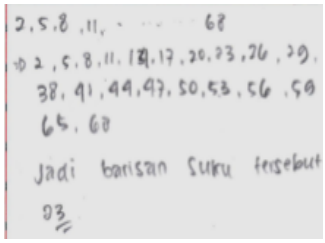


Figure 11. An Example of Student Answers in Item 2

Item 3: The cognitive domain to be achieved is C3 (applying). Based on students' answers, the two dominant patterns of student answers are: (1) students have understood and determined the difference between two non-adjacent arithmetic sequences using the general formula for arithmetic sequences, and (2) other students can determine the number of arithmetic sequences even though they are not using a general formula or by writing the terms of the known terms and inserting several terms.

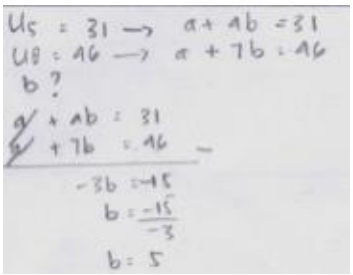
Question 3:

An arithmetic sequence, the 5th term is 31 and the 8th term is 46. The difference between the sequences is...

A. 5
B. 6
C. 7
D. 8
E. 11

Reason: .....

Pattern 1



Pattern 2

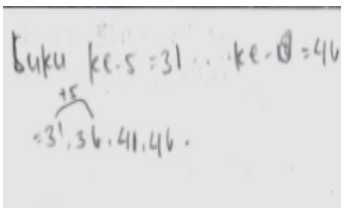


Figure 12. An Example of Student Answers in Item 3

Item 4: The cognitive domain to be achieved is C3 (applying). Based on students' answers, the two dominant patterns of student answers are: (1) students have understood and can determine the nth term in an arithmetic sequence that is known to be two non-adjacent terms using the general formula for arithmetic sequences, and (2) students cannot determine the number of arithmetic sequences because it does not use a general formula but by writing the terms of the known terms and inserting several terms and continuing until the nth term.

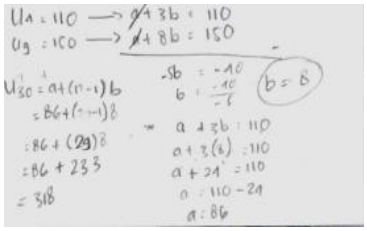
Question 4:

The 4th and 9th terms of an arithmetic sequence are 110 and 150. The 30th terms of the arithmetic sequence are...

A. 308
B. 318
C. 326
D. 344
E. 354

Reason: .....

Pattern 1



Pattern 2

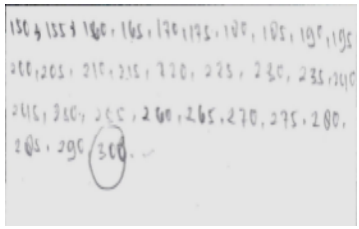


Figure 13. An Example of Student Answers in Item 4

Item 5: The cognitive domain to be achieved is C1 (remembering). Based on students' answers, the two dominant patterns of student answers are: (1) students have understood and determined the middle term of an arithmetic sequence, and (2) students can determine the middle term of an arithmetic sequence but do not use general formulas, but rather by writing the terms from known terms and inserting several terms and then defining them.

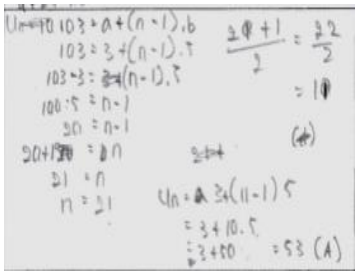
Question 5:

An arithmetic sequence is: 3, 8, 13, 18, ..., 103. Then the middle term of the sequence is ...308

A. 53
B. 52
C. 20
D. 11
E. 10

Reason: .....

Answer Pattern 1



Answer Pattern 2

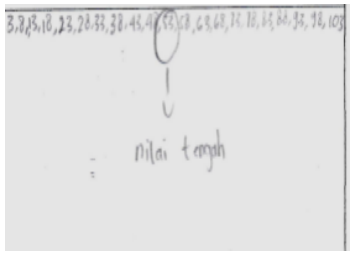


Figure 14. An Example of Student Answers in Item 5

Description of Instrument Development and Student Ability

The instrument developed from this research is the open polytomous response test, and all parameters have been accepted. This instrument is a combination of a multiple-choice test and an essay. Multiple choice test is easier to check students' answers, but their mathematical thinking processes cannot be known in depth. While the essay test has the advantage of being able to find deeper mathematical thinking processes, it takes a long time to check the answers.

Analysis of learning instruments is an important source of composite scores in the final report. In the final report, the student's ability score was first changed to a score of 0–10 (previously 0–100). The conversion is done through a linear transformation by dividing the student's score by the ideal score. Then, the result is multiplied by 10 to get a value range of 0–10 or multiplied by 100 to get a value range of 0–100. In the range of 0–10, the scores of students' mathematical ability were 8.56 (highest) and 4.31 (lowest), or in the range of 0–100, students' math ability scores were 85.6 (highest) and 43.1 (lowest).

The results of the analysis of student abilities are presented in the form of predicates ranging from very low to very high according to the specified category. The results of this analysis show that most students have very low to medium abilities, as much as 62% (253 students), and the remaining 38% (160 students) have high and very high abilities. Other analysis results found that high and very high-ability students tend to work according to concepts with more creative completion steps (different from the teacher's example), but students who have very low to medium abilities can solve problems according to concepts with less creative completion steps (e.g., routine or according to the teacher's example).

Other results show that the assessment with an open response polytomous makes it easier for teachers to explore students' difficulties with a material. Then, from this exploration, the teacher can continue with improvements for students who have learning difficulties. An important finding of this study is that information about students' learning difficulties through the open polytomous response test is more secondhand and complete than other assessment instruments, such as multiple-choice tests (Gierl et al., 2017) or essay tests (Putri et al., 2020).

Discussion

This research is development research aimed at developing a good mathematical assessment instrument using polytomous responses according to classical and modern theories. The results of the data analysis found that there were differences in the results of the analysis between classical theory and modern theory, namely on item discrimination. Classical theory analysis obtained 38 items with bad criteria and only 2 items with good criteria. In contrast, modern

theory analysis obtained 27 items with good criteria and 13 items with bad criteria. According to evaluation theory, modern theory aims to cover the weaknesses of classical theory. The results of the classical theory analysis are often categorized as poor, but modern analysis results are categorized as good, and vice versa (Retnawati, 2014). That is, an item that is not in a good category with classical theory should be analyzed according to the modern theory before revising or replacing the item.

Research on learning assessment with open response polytomous was carried out by Yang et al. (2017). This research aims to diagnose student errors in completing calculus material at university. The instrument compares two types of test namely the two-tier test and the open polytomous response test. The research findings suggest that the open polytomous response test provides more detailed information on student error than the two-level test (see Figure 15). The research findings are in line with the results of this study, which states that the open polytomous response test provides more detailed information about students' abilities.

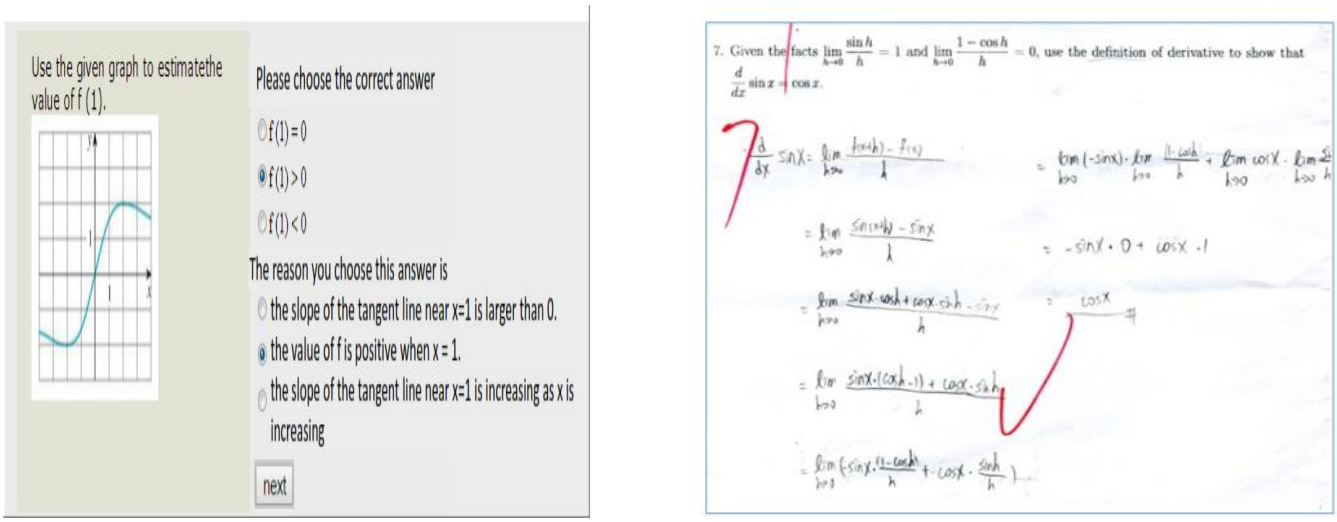


Figure 15. An Example of Student Answers on The Two-Tier Test and The Open Polytomous

Response Test

Another study on learning assessment with open response polychromous was conducted by Ayanwale (2021). Ayanwale's research compares two polytomous response test analysis methods namely the Parallel and Partial Credit Model. The results of his research stated that the Partial Credit Model analysis had a first factor Eigenvalue of 20.5% (ideal Eigenvalue of more than 20%) compared to the Parallel analysis, which had the first factor Eigenvalue of 11.7%. The results of Ayanwale's research are in line with this study, which obtained the first factor Eigenvalue of 20.220%, and the author can state that the instrument developed is suitable to be used to assess vocational students in Indonesia, maybe even outside Indonesia.

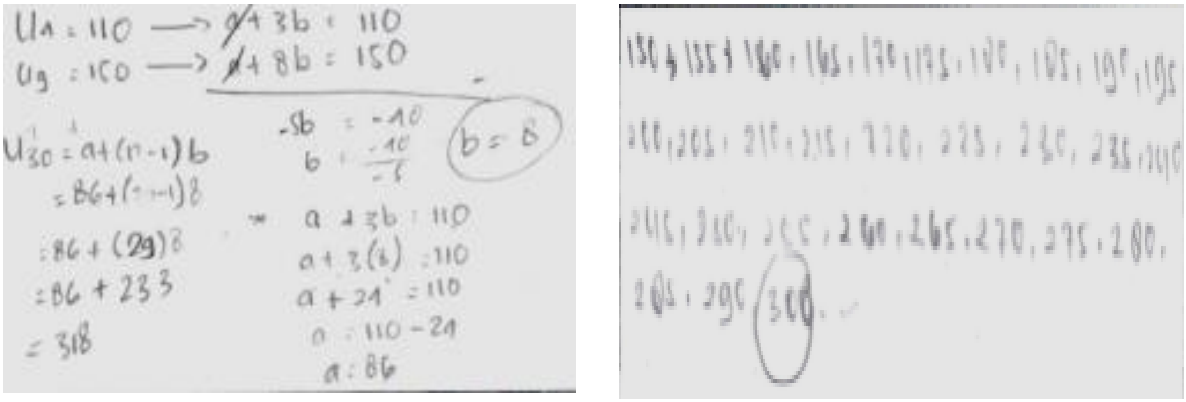
Other studies related to classical and modern theory were conducted by Sarea (2018) and Saepuzaman et al. (2021). Sarea's research states that the response polytomous test has good criteria (classical and modern theory), and the classical theory analysis has more items than modern theory. Meanwhile, Saepuzaman's research found that the response polytomous test had good criteria (classical and modern theory), and the modern theory analysis had more items than classical theory.

The results of the analysis of student answers obtained information that there were two patterns of students' solving questions: (1) formulas and (2) trial and error. Students who use formula patterns in solving problems tend to be carried out by high-ability students, and students who use trial and error patterns tend to be carried out by students with medium and low abilities. Students who use both patterns can answer the questions correctly as shown in Figure 16 below.

The 4th and 9th terms of an arithmetic sequence are 110 and 150. The 30th terms of the arithmetic sequence are…

A. 308
B. 318
C. 326
D. 344
E. 354

Reason: .....



(i) (ii)

Figure 16. An Example of Student Answer Patterns with (i) Formulas, (ii) Trial and Error

Both patterns of solving math problems are common ways. This is in line with the opinion of Mason et al. (2010) that there are several ways that can be used in solving math problems, namely trial and error, using a drawing or model, analogy, and formula. Syahlan (2017) states that two ways that students often use to solve math problems are (1) trial and error and (2) formulas. Usually, students use trial and error methods to solve easy problems and formulas methods to solve difficult questions. If combined with student answers and Syahlan's opinion, students with medium and low abilities in solving easy questions will use trial and error methods, and students with high abilities in solving difficult questions will use formulas.

The results of these student answers can be important information for teachers in teaching mathematics. This means that before teaching material, the teacher must know the students' initial abilities (high, medium, or low) and the level of difficulty of the material. Some of the benefits for teachers who know students' initial abilities are: teachers can develop professionally centered on students (Gonzalez, 2018), teachers can adjust the level of cognitive and increase student learning engagement (Dong et al., 2020), assist teachers in designing pedagogical practices and correcting students' misconceptions (Geofrey, 2021).In addition, the benefit for teachers who know the level of difficulty of a material is that they can design remedial learning plans (Muhson et al., 2017; Wulanningtyas et al., 2020).The description above shows that there are many benefits that the teacher will get if the teacher knows the students' actual abilities, and this can be known by the teacher if the teacher uses the open polytomous response test.

Conclusion

Based on the results of the research and discussion, it can be concluded that (1) the open polytomous response test has a good category according to classical and modern theory. Thus, the test instrument requirements are accountable (qualifies for a good test) is a good test instrument according to the analysis of the two theories (classical and modern theory), and (2) the open polytomous response test can provide information on the actual competence of students; this is observed in the students' arguments in giving reasons for their choices. This is observed in the students' arguments when giving reasons for their choices. Therefore, the open polytomous response test can be used as an alternative to learning assessment.

Recommendation

Based on the research results, there are several recommendations for teachers, schools, and other researchers. For teachers, they should familiarize students with giving a test in the form of a polytomous response before giving the test. In schools, principals or other leaders should encourage other teachers to take advantage of this test and develop other assessment instruments. For other researchers, they should conduct research by developing instruments with other polytomous responses (assessment of learning and assessment as learning) on other materials. In addition, for further research, it is suggested to conduct a study that develops an assessment instrument with a learning response polytomous (pretest). This is important so that students' prior knowledge can be known and learning can be effective.

Limitations

The research carried out has several limitations. Firstly, the selected schools have not met the researchers' expectations, for example, representing schools of high, medium, and low quality. In addition, during the collection of research data that begins with learning, it is not fully controlled by the researcher, so the students who are the research samples are less conditioned. The research is only limited to algebraic materials (sequences and series, matrices, equations, and quadratic functions), and does not represent other mathematical materials, such as geometry and statistics.

Authorship Contribution Statement

Sutiarso: Conceptualization, design, development of instruments, analysis, article writing, final approval. Rosidin: Design, development of instruments, analysis, editing. Sulistiawan: Development of instruments, collect data, analysis, editing.

References

Alagumalai, S., Curtis, D., & Hungi, N. (2005). *Applied Rasch measurement: A book of exemplars.* Springer-Kluwer Academic Publishers.

Anisa. (2013). Perbandingan penskoran dikotomi dan politomi dalam teori respon butir untuk pengembangan bank soal mata kuliah matematika dasar [Comparison of scoring dichotomies and polytomies in item response theory for the development of a question bank for basic mathematics courses]. *Jurnal Matematika, Statistika, & Komputasi, 9*(2), 95-113. https://bit.ly/39wv73P

Antara, A. A., Yasna, I. M., Dewi, N. W., & Maduriana, I. M. (2019). Karakteristik tes prestasi belajar model campuran dikotomus dan politomus generalized partial credit model [Characteristics of learning achievement test mixed dichotomous and polytomous generalized partial credit model]. *Jurnal Suluh Pendidikan, 17*(1), 32-37. https://bit.ly/3yFQ6eO

Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan* [Educational evaluation basics]. Bumi Aksara.

Ayanwale, M. A. (2021). Calibration of polytomous response mathematics achievement test using generalized partial credit model of item response theory. *Educatum: Journal of Science, Mathematics and Technology, 8*(1), 57-69. https://bit.ly/3Nob4mN

Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook: The cognitive domain.* David McKay.

Chandrasegaran, A. L., Treagust, D., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice, 8*(3), 293-307. https://doi.org/10.1039/B7RP90006F

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Harcourt.

DeMars, C. (2010). *Item response theory.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195377033.001.0001

Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). Four-tier diagnostic test with certainty of response index on the concepts of fluid. In C. Anwar (Ed.), *Young scholar symposium on transdisciplinary in education and environment (YSSTEE)* (pp. 641-650). UIN Raden Intan. https://bit.ly/3lfARkW

Dong, A., Jong, M. S., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontier Psychology Journal*, *11*(1), 1-10. https://doi.org/10.3389/fpsyg.2020.591203

Earl, L. M. (2013). *Assessment for learning; Assessment as learning: Changing practices means changing beliefs.* Assessment & Support Team. https://bit.ly/3MnA9hw

Falani, I., Akbar, M., & Naga, D. S. (2020). The Precision of students' ability estimation on combinations of item response theory models. *International Journal of Instruction, 13*(4), 545-558. https://doi.org/10.29333/iji.2020.13434a

Geofrey, M. (2021). Children's prior knowledge is very important in teaching and learning in this era of constructivism. Researh Gate. https://doi.org/10.13140/RG.2.2.28470.22083

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice test in education: A comprehensive review. *Review of Educational Research, 87*(6), 1082–1116. https://doi.org/10.3102/0034654317726529

Gonzalez, G. (2018). Understanding teacher noticing of students' prior knowledge: Challenges and possibilities. *The Mathematics Enthusiast, 15*(3), 483-528. https://doi.org/10.54870/1551-3440.1442

Gregory, R. J. (2015). *Psychological testing: History, principles, and applications.* Pearson.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38–47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x

Hambleton, R. K., & Linden, W. J. (1982). Advances in item response theory and applications: An introduction. *Applied Psychological Measurement, 6*(4), 373-378. https://doi.org/10.1177/014662168200600401

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Kluwer. https://doi.org/10.1007/978-94-017-1988-9

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Sage Publications.

Hilton, A., Hilton, G., Dole, S., & Goos, M. (2013). Development and application of a two-tier diagnostic instrument to assess middle-years students' proportional reasoning. *Mathematics Education Research Journal, 25*(1), 523–545. https://doi.org/10.1007/s13394-013-0083-6

Ikmawati. (2020). Pengaruh disiplin dan kreativitas belajar terhadap hasil belajar matematika di SMK negeri dan swasta [The effect of discipline and creativity in learning on math learning outcomes in public and private vocational schools]. *Primatika: Jurnal Pendidikan Matematika, 9*(1), 35-42. https://doi.org/10.30872/primatika.v9i1.250

Kartono. (2008). Penyetaraan tes model campuran butir dikotomus dan politomus pada tes prestasi belajar [Equalization of the dichotomous and polytomous mixed item model tests on the learning achievement test]. *Jurnal Penelitian dan Evaluasi Pendidikan, 12*(2), 302-320. https://doi.org/10.21831/pep.v12i2.1433

Khiyarunnisa, A., & Retnawati, H. (2018). A two-tier diagnostic test instrument on calculus material: What, why, and how? In A. W. Subiantoro (Ed.), *5th ICRIEMS Proceedings* (pp. 479-485). Faculty of Mathematics and Natural Sciences. https://bit.ly/3liFMS8

Khusnah, M. (2019). The development of two-tiers diagnostic test for identifying tenth-grade student's misconception about the categorization of hadith. [Master's thesis, Malang State Islamic University]. UIN Malang Digital Archive. https://bit.ly/3sEcNMA

Lestari, S. A., Zawawi, I., Khikmiyah, F., & Fauziyah, N. (2021). Development evaluation tool two-tier multiple choice using wondershare quiz creator to identify mathematical connection. *Journal of Mathematics Education, 6*(2), 133-148. https://bit.ly/3Psrc8s

Linacre, J. M. (2012). *Winstep: Rasch-model computer programs.* Winsteps.Com. https://bit.ly/3wljq95

Malhotra, N. K. (2006). *Riset pemasaran* [Marketing research]. Erlangga.

Mason, J., Burton, L., & Stacey, K. (2010). *Thinking mathematically* (2nd ed.). Pearson Educational Limited.

Muhson, A., Lestari, B., Supriyanto, & Baroroh, K. (2017). The development of practical item analysis program for Indonesian. *International Journal of Instruction, 10*(2), 199-210. https://doi.org/10.12973/iji.2017.10213a

Myanda, A. A., Riezky, M. P., & Maridi. (2020). Development of two-tier multiple-choice test to assess students' conceptual understanding on respiratory system material of 11th high school.*InternationalJournal of Science and Applied Science: Conference Series, 4*(1), 44-55. Sebelas Maret University. https://bit.ly/3yIePzb

Oktaria. (2016). *Development of mathematics teaching materials with ICT for students in vocational schools* [Unpublished master's thesis]. University of Lampung.

Permendikbud. (2016). *Standar isi pendidikan dasar dan menengah* [Standards of content for primary and secondary education]. Indonesian Government publication service. https://bit.ly/3FZHlhp

Plomp, J. (2013). *Educational design research: An introduction.* Institute for Curriculum Development (SLO). https://bit.ly/3MzqkNm

Putri, B. S., Kartono, & Supriyadi. (2020). Analysis of essay test instruments using higher order thinking skill (HOTS) at high school mathematics students using the Rasch model. *Journal of Research and Educational Research Evaluation, 9*(2), 58-69. https://bit.ly/37QCtyu

Putri, E. L., Dwijanto, D., & Sugiman. (2017). Analysis of mathematical communication skills and confidence of 10th grade of SMK in geometry material viewed from cognitive style. *Unnes Journal of Mathematics Education, 6*(1), 97-107. https://bit.ly/3yF4iF5

Retnawati, H. (2014). *Teori respons butir dan penerapannya* [Item response theory and its application]. Nuha Medika. https://bit.ly/39TFDlF

Rosidin, U. (2017). *Evaluasi dan asesmen pembelajaran* [Learning evaluation and assessment]. Media Akademi.

Rovita, C. A., Zawawi, I., & Huda, S. (2020). Pengembangan alat evaluasi pembelajaran matematika berbasis two-tier

multiple choice menggunakan Ispring Suite 9 [Development of an evaluation tool for learning mathematics based on two-tier multiple choice using Ispring Suite 9]. *Postulat: Jurnal Inovasi Pendidikan Matematika, 1*(2), 150-163. https://doi.org/10.30587/postulat.v1i2.2094

Saepuzaman, D., Istiyono, E., Haryanto, Retnawati, H., & Yustiandi. (2021). Analisis estimasi kemampuan siswa menjawab soal fisik dengan pendekatan item response theory [Analysis of the estimation of students' ability to answer physical questions with an item response theory approach]. *Karst: Journal of Physics Education and Its Application/Karst: Jurnal Pendidikan Fisika dan Terapannya, 4*(1), 8-13. https://doi.org/10.46918/karst.v4i1.948

Sarea, M. S. (2018). Karakteristik soal ujian akhir semester pendidikan agama islam dan budi pekerti tingkat sekolah dasar [Characteristics of the final exam for Islamic religious education and character at the elementary school level]. *An-Nahdhah, 11*(2), 303-318. https://bit.ly/3FToOU7

Sarea, M. S., & Ruslan, R. (2019). Karakteristik butir soal: Teori tes klasik and respon [Characteristics of items: Classical and response test theory]. *Didaktika: Jurnal Kependidikan, 13*(1), 1-16. https://doi.org/10.30863/didaktika.v13i1.296

Suhaini, M., Ahmad, A., & Bohari, N. M. (2021). Assessments on vocational knowledge and skills: A content validity analysis. *European Journal of Educational Research, 10*(3), 1529-1540. https://doi.org/10.12973/eu-jer.10.3.1529

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan* [Rasch modeling applications in educational assessment]. Trim Komunikata.

Sundari, Kahar, M. S., & Erwinda, E. (2021). Analisis kemampuan berpikir tingkat tinggi menggunakan instrumen HOTS berbasis two-tier diagnostic test [Analysis of higher order thinking skills using the HOTS instrument based on a two-tier diagnostic test]. *Aksioma: Jurnal Program Studi Pendidikan Matematika, 10*(4), 2726-2735. https://doi.org/10.24127/ajpm.v10i4.4260

Suwarto. (2012). *Pengembangan tes diagnostik dalam pembelajaran* [Development of diagnostic test in learning]. Graha Ilmu.

Syahlan. (2017). Sepuluh strategi dalam pemecahan masalah matematika [Ten strategies for solving math problems]. *Indonesian Digital Journal of Mathematics and Education, 4*(6), 358-369. https://doi.org/10.31227/osf.io/6qfpm

Syaifuddin, M. (2020). Implementation of authentic assessment on mathematics teaching: Study on junior high school teachers. *European Journal of Educational Research, 9*(4), 1491-1502. https://doi.org/10.12973/eu-jer.9.4.1491

Treagust, D. (1988). Development and use of diagnostic test to evaluate student's misconceptions in science. *International Journal of Science Education, 10*(2), 159-169. https://doi.org/10.1080/0950069880100204

Untary, H., Risdianto, E., & Kusen. (2020). *Analisis data penelitian dengan model rash dan Winstep*[Analysis of research data with rash and Winstep models]. Halaman Moeka Publishing. https://bit.ly/3NlnGLe

Vani, I. K., Paloloang, B., & Idris, M. (2019). Pengaruh persepsi dan minat belajar terhadap hasil belajar matematika siswa kelas X SMKN 6 Palu [The influence of perception and interest in learning on mathematics learning outcomes for students of class X Vocational School 6 Palu]. *Jurnal Elektronik Pendidikan MatematikaTadulako, 6*(4), 455-468. https://bit.ly/3wjuXp9

Wulan, A. (2018). *Menggunakan asesmen kinerja untuk pembelajaran sain dan penelitian* [Using performance assessment for science learning and research]. UPI Press.

Wulanningtyas, M. E., Suswanti, & Marhaeni, N. H. (2020). The student of error analysis and remedial program in working on the story of comparative material and turning value. *Jurnal Daya Matematis/Journal of Mathematical Power, 8*(2), 161-166. https://doi.org/10.26858/jdm.v8i2.14344

Yang, T. C., Fu, H. T., Hwang, G. J., & Yang, S. J. (2017). Development of an interactive mathematics learning system based on a two-tier test diagnostic and guiding strategy. *Australasian Journal of Educational Technology, 3*(1), 60-80. https://doi.org/10.14742/ajet.2154

Appendix

Instrument of the Open Polytomous Response Test

Name of Student : ……..

Class/Department : ……..

School : ……..

Instructions: Mark (x) one of the correct answer choices, and give the reason

(use another piece of paper to write down your reason)

1. Given an arithmetic sequence: 3, 8, 13, 18, ... The formula for the nth term of the sequence is ....
   A. $U_n = 5n - 3$ D. $U_n = 4n - 1$
   B. $U_n = 5n - 2$ E. $U_n = 3n + 2$
   C. $U_n = 4n - 1$

   Reason: .....

2. Given an arithmetic sequence: 2, 5, 8, 11, …, 68. The number of terms in the sequence is…
   A. 12 D. 23
   B. 13 E. 24
   C. 22

   Reason: .....

3. An arithmetic sequence, the 5th term is 31 and the 8th term is 46. The difference between the sequences is...
   A. 5 D. 8
   B. 6 E. 11
   C. 7

   Reason: .....

4. The 4th and 9th terms of an arithmetic sequence are 110 and 150. The 30th terms of the arithmetic sequence are...
   A. 308 D. 344
   B. 318 E. 354
   C. 326

   Reason: .....

5. An arithmetic sequence is: 3, 8, 13, 18, …, 103. Then the middle term of the sequence is …
   A. 53 D. 11
   B. 52 E. 10
   C. 20

   Reason: .....

6. Given the arithmetic sequence: 4, 10, 16, 22, …. If two numbers are inserted in every two consecutive terms then the 10th term of the sequence is...
   A. 18 D. 24
   B. 20 E. 26
   C. 22

   Reason: .....

7. The nth term of an arithmetic series is $U_n = 3n - 5$. The formula for the sum of the first n terms of the series is ...
   A. $S_n = \frac{n}{2}(3n - 7)$ D. $S_n = \frac{n}{2}(3n - 3)$
   B. $S_n = \frac{n}{2}(3n - 5)$ E. $S_n = \frac{n}{2}(3n - 3)$
   C. $S_n = \frac{n}{2}(3n - 4)$

   Reason: .....

8. The sum of the first n terms of an arithmetic series. $S_n = n^2 - 19n$. The formula for the nth term of the Series is.....
   A. $5n - 20$ D. $2n - 20$
   B. $5n - 10$ E. $2n - 10$
   C. $2n - 30$

   Reason: .....

9. The sum of all integers between 100 and 300 which are divisible by 5 is ...
   A. 8,200 D. 7,600
   B. 8,000 E. 7,400
   C. 7,800

   Reason: .....

10. PT. Angkasa Jaya in the first year produced 5,000 units. In the second year, production was reduced by 80 units per year. In what year did the company produce 3,000 units?
    A. 24 D. 27
    B. 25 E. 28
    C. 26

    Reason: .....

11. The middle term of an arithmetic sequence is 25. If the difference is 4 and the 5th term is 21. Then the sum of all the terms in the sequence is ...
    A. 175 D. 295
    B. 189 E. 375
    C. 275

    Reason: .....

12. A number of candies are distributed among five children according to the rules of an arithmetic sequence. The younger the child, the more candy he gets. If the second child receives 11 pieces of candy and the fourth child 19 pieces, then the total number of candies is... pieces
    A. 60 D. 75
    B. 65 E. 80
    C. 70

    Reason: .....

13. The sum of the first n terms of a series is $2n^2$-n. So the 12th term of the series is...
    A. 564 D. 45
    B. 276 E. 36
    C. 48

    Reason: .....

14. The number of terms in the geometric sequence: 3, 6, 12, ..., 3072 is ...
    A. 9 D. 12
    B. 10 E. 13
    C. 11

    Reason: .....

15. A geometric sequence has the 2nd term is 8, and the 5th term is 64. The 7th term of the sequence is ...
    A. 32 D. 256
    B. 64 E. 512
    C. 128

    Reason: .....

16. The value of the middle term of the geometric sequence:
    6, 3, ..., 3/512 is ...
    A. $\frac{1}{16}$ D. $\frac{4}{16}$
    B. $\frac{2}{16}$ E. $\frac{5}{16}$
    C. $\frac{3}{16}$

    Reason: .....

17. A piece of wire is cut with the first piece being 8 cm, and the next piece 1.5 times the previous cut. Then the 5th piece is... cm
    A. 18 D. 35
    B. 24 E. 40.5
    C. 27.5

    Reason: .....

18. The sum of an infinite geometric series is 12 and the first term is 9. The ratio is ...
    A. $\frac{3}{4}$ D. $-\frac{1}{2}$
    B. $\frac{1}{4}$ E. $-\frac{3}{4}$
    C. $\frac{1}{3}$

    Reason: .....

19. A ball falls from a height of 10 m and bounces back 3/4 times its previous height. The total number of paths until the ball stops is.... m
    A. 60 D. 90
    B. 70 E. 100
    C. 80

    Reason: .....

20. Given $\begin{bmatrix} 4 & 8 & 5 \\ 6 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 2a & b+a & 5 \\ 6 & 3 & 1 \end{bmatrix}$.

    The value of $3a + b$ is ...
    A. 8 D. 14
    B. 10 E. 20
    C. 12

    Reason: .....

21. Given $K = \begin{bmatrix} a & 2 & 3 \\ 5 & 4 & b \\ 8 & 3c & 11 \end{bmatrix}$ and $L = \begin{bmatrix} 6 & 2 & 3 \\ 5 & 4 & 2a \\ 8 & 4b & 11 \end{bmatrix}$. If K = L, then c is ...
    A. 12 D. 15
    B. 13 E. 16
    C. 14

    Reason: .....

22. Given $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 5 & 2 \\ -1 & 0 \end{bmatrix}$. Then (A + C) – (A + B) is ...
    A. $\begin{bmatrix} 5 & 4 \\ 5 & 4 \end{bmatrix}$ D. $\begin{bmatrix} 3 & -1 \\ -1 & -1 \end{bmatrix}$
    B. $\begin{bmatrix} 4 & 7 \\ 2 & 5 \end{bmatrix}$ E. $\begin{bmatrix} 7 & -1 \\ 1 & -1 \end{bmatrix}$
    C. $\begin{bmatrix} 4 & 0 \\ -4 & -4 \end{bmatrix}$

    Reason: .....

23. Given matrix $P = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 4 & 5 \\ 0 & 7 & 6 \end{bmatrix}$, then $(P^t)^t$ is ...

A. $\begin{bmatrix} 1 & 2 & 0 \\ 3 & 3 & 7 \\ 1 & 5 & 6 \end{bmatrix}$
B. $\begin{bmatrix} 1 & 3 & 1 \\ 3 & 3 & 7 \\ 1 & 5 & 6 \end{bmatrix}$
C. $\begin{bmatrix} 1 & 3 & 1 \\ 2 & 4 & 5 \\ 0 & 7 & 6 \end{bmatrix}$
D. $\begin{bmatrix} 0 & 7 & 6 \\ 2 & 4 & 5 \\ 1 & 3 & 1 \end{bmatrix}$
E. $\begin{bmatrix} 0 & 7 & 6 \\ 2 & 4 & 5 \\ 1 & 3 & 1 \end{bmatrix}$

Reason: .....

24. Given matrix $A = \begin{bmatrix} 2x & -5 \\ 3 & y \end{bmatrix}$, $B = \begin{bmatrix} y & 2 \\ 2 & 4 \end{bmatrix}$, and $C = \begin{bmatrix} 8 & -3 \\ 5 & 2x \end{bmatrix}$.

If A + B = C, then $x + y$ = ...
A. –5
B. –1
C. 1
D. 3
E. 5

Reason: .....

25. If $A = [3 \quad 5]$ and $B = \begin{bmatrix} 1 & 4 \\ 2 & 6 \end{bmatrix}$, then 2AB is...
A. [13 42]
B. [26 84]
C. [26 42]
D. [13 84]
E. [30 36]

Reason: .....

26. If $A = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix}$, $B = \begin{bmatrix} -2 & 0 \\ 1 & 3 \end{bmatrix}$, and $C = \begin{bmatrix} 3 & -1 \\ 1 & -2 \end{bmatrix}$ then A(B – C) = ...

A. $\begin{bmatrix} -5 & -14 \\ 10 & 18 \end{bmatrix}$
B. $\begin{bmatrix} -5 & -4 \\ 10 & 6 \end{bmatrix}$
C. $\begin{bmatrix} 1 & -16 \\ -2 & 22 \end{bmatrix}$
D. $\begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix}$
E. $\begin{bmatrix} -7 & 19 \\ -2 & 20 \end{bmatrix}$

Reason: .....

27. Matrix X that satisfies $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} X = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}$ is ...

A. $\begin{bmatrix} -6 & -5 \\ 5 & 4 \end{bmatrix}$
B. $\begin{bmatrix} 5 & -6 \\ 4 & 5 \end{bmatrix}$
C. $\begin{bmatrix} -6 & -5 \\ 4 & 5 \end{bmatrix}$
D. $\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$
E. $\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$

Reason: .....

28. If matrix $A = \begin{bmatrix} 3 & -x \\ 6 & 8 \end{bmatrix}$ is singular matrix. Value of x that satisfies is ....

A. – 5
B. – 4
C. – 3
D. 3
E. 4

Reason: .....

29. Given matrix A = $\begin{bmatrix} 0 & 2 & 3 \\ -2 & 0 & 4 \\ -3 & -4 & 0 \end{bmatrix}$, then matrix determinant A is ...

A. 0
B. 1
C. 2
D. 2
E. 4

Reason: .....

30. Transpose matrix P is P^T. If $P = \begin{bmatrix} 4 & 7 \\ 3 & 5 \end{bmatrix}$ then matrix $(P^T)^{-1}$ is ...

A. $\begin{bmatrix} -5 & 7 \\ 3 & -2 \end{bmatrix}$
B. $\begin{bmatrix} 3 & -4 \\ -5 & 7 \end{bmatrix}$
C. $\begin{bmatrix} -5 & 3 \\ 7 & -4 \end{bmatrix}$
D. $\begin{bmatrix} -3 & 5 \\ 4 & -7 \end{bmatrix}$
E. $\begin{bmatrix} 4 & -3 \\ -7 & 5 \end{bmatrix}$

Reason: .....

31. Given $A = \begin{bmatrix} 4 & 3 \\ 2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 2 \\ 3 & 1 \end{bmatrix}$. Inverse matrix $(AB)^{-1}$ = ...

A. $\frac{1}{2}\begin{bmatrix} -6 & 11 \\ 16 & -29 \end{bmatrix}$
B. $\frac{1}{2}\begin{bmatrix} 29 & 11 \\ 16 & 6 \end{bmatrix}$
C. $-\frac{1}{2}\begin{bmatrix} 29 & 11 \\ 16 & 6 \end{bmatrix}$
D. $-\frac{1}{2}\begin{bmatrix} 6 & 11 \\ 16 & 29 \end{bmatrix}$
E. $-\frac{1}{2}\begin{bmatrix} -6 & 11 \\ 16 & -29 \end{bmatrix}$

Reason: .....

32. The roots of the quadratic equation 3 and – 4 are ...
A. $x^2 + x - 12 = 0$
B. $x^2 - x - 12 = 0$
C. $x^2 - x + 12 = 0$
D. $x^2 - 3x + 4 = 0$
E. $x^2 - 4x + 3 = 0$

Reason: .....

33. The roots of the quadratic equation $5x^2 + 4x - 12 = 0$ ar e ...
A. -2 and $\frac{5}{6}$
B. 2 and $-\frac{5}{6}$
C. 2 and $\frac{6}{5}$
D. -2 and $-\frac{6}{5}$
E. -2 and $\frac{6}{5}$

Reason: .....

34. The roots of the equation $x^2 - 2x - 3 = 0$ are x_1 and x_2. If $x_1 > x_2$, then $x_1 - x_2$ is ...
A. -4
B. -2
C. 0
D. 2
E. 4

Reason: .....

35. The roots of the quadratic equation: $2x^2 - 3x - 9 = 0$ are x_1 and x_2. Value of $x_1^2 + x_2^2$ is ...
A. $11\frac{1}{4}$
B. $6\frac{3}{4}$
C. $2\frac{1}{4}$
D. $-6\frac{3}{4}$
E. $-11\frac{1}{4}$

Reason: .....

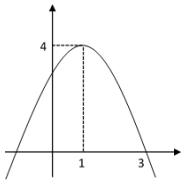
36. The roots of the quadratic equation: $x^2 + 2x + 3 = 0$ are α and β. The quadratic equation (α – 2) and (β – 2) is ...
A. $x^2 + 6x + 5 = 0$
B. $x^2 + 6x + 7 = 0$
C. $x^2 + 6x + 11 = 0$
D. $x^2 - 2x + 3 = 0$
E. $x^2 + 2x + 11 = 0$

Reason: .....

37. A quadratic function that has a minimum value 2 for x = 1 and has a value of 3 for x = 2 is ...
A. $y = x^2 - 2x + 1$
B. $y = x^2 - 2x + 3$
C. $y = x^2 + 2x - 1$
D. $y = x^2 + 2x + 1$
E. $y = x^2 + 2x + 3$

Reason: .....

38. The figure below is a graph of the quadratic equation? ...
A. $y = x^2 + 2x + 3$
B. $y = x^2 - 2x - 3$
C. $y = -x^2 + 2x - 3$
D. $y = -x^2 - 2x + 3$
E. $y = -x^2 + 2x + 3$



Reason: .....

39. If f is a quadratic function whose graph passes through the points (1,0), (4,0) and (0, –4) then the value of f(7) is ...
A. – 16
B. – 17
C. – 18
D. – 19
E. – 20

Reason: .....

40. The graph equation of a quadratic function has an extreme point (–1, 4) and through (0, 3) is ...
A. $y = -x^2 + 2x - 3$
B. $y = -x^2 + 2x + 3$
C. $y = -x^2 - 2x + 3$
D. $y = -x^2 - 2x - 5$
E. $y = -x^2 - 2x + 5$

Reason: .....