

Construction a Diagnostic Test in the Form of Two-tier Multiple Choice on Calculus Material

Edy Nurfalaha^{1,*}, Irvana Arofah², Ika Yuniwati³, Andi Haslinah⁴, Dwi Retno Lestari⁵

¹Study Program of Mathematics Education, Universitas PGRI Ronggolawe Tuban, Indonesia

²Study Program of Mathematics, Universitas Pamulang, Indonesia

³Study Program of Mechanical Engineering, Politeknik Negeri Banyuwangi, Indonesia

⁴Study Program of Mechanical Engineering, Universitas Islam Makassar, Indonesia

⁵Ministry of Research and Technology/ National Agency for Research Dan Innovation, Indonesia

Received July 5, 2020; Revised August 21, 2020; Accepted September 11, 2020

Cite This Paper in the following Citation Styles

(a): [1] Edy Nurfalaha, Irvana Arofah, Ika Yuniwati, Andi Haslinah, Dwi Retno Lestari, "Construction a Diagnostic Test in the Form of Two-tier Multiple Choice on Calculus Material," *Mathematics and Statistics*, Vol. 8, No. 5, pp. 577 - 582, 2020. DOI: 10.13189/ms.2020.080512.

(b): Edy Nurfalaha, Irvana Arofah, Ika Yuniwati, Andi Haslinah, Dwi Retno Lestari (2020). *Construction a Diagnostic Test in the Form of Two-tier Multiple Choice on Calculus Material*. *Mathematics and Statistics*, 8(5), 577 - 582. DOI: 10.13189/ms.2020.080512.

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This work is a research development of two-tier multiples choice diagnostic test instruments on calculus material. The purpose of this study is; 1) Obtaining the construction of a two-tier multiples choice diagnostic test based on the validity of the contents and Constable, 2) obtaining the quality of two-tier multiples choice diagnostic tests based on the reliability value. The method used is focused on the construction of diagnostic tests. The development research was adapted from the Retnawati development model. The research generated: 1) Construction of a two-tier multiples choice diagnostic test based on the validity of the contents and the construction obtained that the two-tier multiples choice diagnostic test is proven valid. 2) The quality of two-tier multiples choice diagnostic tests based on the reliability value gained that the compiled two-tier diagnostic test instruments. The validity of the content is evidenced by the average validity index (V), for the two-tier multiples choice diagnostic test instrument obtained an average validity index (V) of 0.9333 and for an interview guideline instrument acquired the validity index (V) 0.7556 in which both the validity index (V) approaches the value 1. Whereas for the validity of the construction acquired three dominant factors based on the scree-plot and corresponds to many factors on the calculus material examined in this study. The quality of two-tier multiples choice diagnostic tests is compiled of two-tier diagnostic test instruments based on the reliability

value gained.

Keywords Two-tier Multiple Choice, Calculus Material, Retnawati Development Model, Reliability Value, Validity

1. Introduction

Mathematics education has a very important role, because mathematics is a fundamental science that is used widely in various areas of life. Good education is capable of producing output or achievement and quality and has the ability that can be beneficial for others [1]. Chambers [2] mentions that mathematics is a science of abstract patterns that have characteristics as a tool to solve problems, as a foundation of scientific and technological studies, and can provide ways to model the situation in real life.

In addition, as the students learn mathematics, students will learn about the power of mathematics that will later develop the skills of learning to learn. The student's reasoned ability through the mathematical learning process will increase students' readiness to become human beings who have a lifetime learner or a lifelong study. Mathematics also plays a very important role in the development of science, because mathematics is the basis

of science and technology and also mathematics is one of the knowledges that has an important role in thinking, namely as a tool to solve problems in everyday life.

Based on the explanation, the importance of mathematics is important so that mathematics needs to be learned, understood, and mastered by students. But the results of the mathematics learning performed in schools are still not optimal, it can be seen from the number of students who are still experiencing difficulties in learning mathematics. The difficulties experienced by the students are explained in the research conducted by Yeo [3], that from the results of the interview difficulties experienced by students in the understanding of mathematics is a lack of understanding of the problem caused, lack of knowledge to do strategies in resolving a problem, inability to translate problems into mathematics, and the inability of students to use true mathematics.

Wijaya, Heuvel-Panhuizen, Doormana, and Robitzsch [4] in his research on students' difficulties in resolving context-based problems focusing on student error analysis. In the study, students' difficulty analysis was seen from four things: (1) comprehension, (2) Transformation, (3) Mathematical Processing, and (4) encoding. Based on the research it has been obtained that the most students experiencing difficulty is at the stage of transformation (inability to transform context-based problems into mathematical models) and comprehension stage (inability to understand the meaning of the problem). Based on the previous description, it can be obtained that students are still experiencing difficulties in resolving mathematical problems so that the students make mistakes in the process of solving the mathematical problems.

Based on the study of literature, there are some mistakes that students have done in solving math problems. Research conducted by Herutomo and Saputro [5] showed that the problem of algebraic material that occurred in one of the junior high school in Semarang is that students still make mistakes in solving the problems of algebraic operations, then the student is also wrong in interpreting the meaning of 'scribble' that denominator and divisible. These things show that students do not use their knowledge of integer and fractional operations in working with algebraic material. Students are still struggling and many make mistakes in resolving algebraic matter. The most basic difficulties experienced by students is translating the story into a mathematical model. The result is obvious, if the mathematical model form is wrong then the next process will also be wrong.

The two-tier multiples choice diagnostic test is a form of test instrument. Before an instrument is used then the instrument must be analyzed first which will then show the quality of the instrument and show that the instrument is appropriate and feasible for use. The quality of an instrument can be seen from two main criteria: validity and reliability.

Suryabrata [6] defines the validity of the instrument as

the extent to which the instrument measures what is supposed to be measured. As for the reliability of an instrument shows the consistency of the data measurement results, it means if the instrument is tested to the same person or group of people but at different times, or if the instrument is tested to a person or group of people at the same time or can also be at different times. In addition to the validity and reliability there are several other things that also need to be analyzed from the instrument such as difficulty level, differentiation power, and the effectiveness of distractors for multiple choice questions.

Based on what has been described in the above exposure, the authors are interested in conducting research by constructing a diagnostic test in the form of two-tier multiple choice to analyze or diagnose the mistakes students have made in complete mathematical problems. Based on this, the researcher titled this study was "development of Two-Tier Multiples Choice Diagnostic test instruments on calculus material".

2. Methodological Research

The research methods used in this research are development research focused on the construction of diagnostic tests. The product developed in this development research is a test instrument constructed based on a two-tier multiples-choice format used to uncover the mathematical errors that occur in high school students. The procedure on this research is divided into two phases, namely the first stage of product development and the second is the application stage of the product [7].

The development was adapted from the Retnawati development model [8], there were nine steps in it, namely (1) Determining the purpose of the instrument preparation, (2) Looking for relevant theories or material coverage, (3) Drafting instrument item indicator, (4) Arranging instrument items, (5) Validation of contents, (6) revisions based on the input validator, (7) Conduct a trial to the respondent to obtain the participant's response data, (8) Perform analysis (reliability, difficulty level, and differentiation power), and (9) assemble the instrument. Once the product development stage is completed, it will proceed to the product application stage. At the stage of this product application includes two things that carry out tests using the resulting product and then interpret the test results.

Aiken formulates the Aiken's formula V to count content-validity coefficient based on assessment result from the expert panel as much as n people toward an item from the terms of how far the item represents the measured contract. The submitted formula by Aiken can be shown below [10,11]:

$$V = \frac{\sum s}{n(c-1)}$$

Where,

V = validity index item

S = score applied, each rater reduced low score in category used ($s=r-l_0$, $\rightarrow r$ = rater score choice and l_0 = low score in score categorizing)

N = number of raters

C = number of criterion/rating

Reliability

The instrument's reliability is intended to see the consistency of the tests made if the observation is repeated. The level of instrument's reliability empirically proven by the amount of the reliability coefficient which is in the range of 0 to 1 [10,11]. The higher the coefficient value means the higher the reliability, and vice versa. The coefficient formula of *Alpha Cronbach's* used to estimate the test reliability and calculate it using *IteMan 4.3* computer program. The reliability estimation is based on the index of instrument reliability that is good if > 0,7 [10,11].

3. Results

3.1. Compile the Test Specification

The initial step in preparation of diagnostic tests on this research is determining test objectives. The test objectives developed in this study are to know the mistakes students often have in solving math problems. The tested material is a customized calculus material with the low percentage of the UN's absorption in the 2015/2016 school year and the 2016/2017 school year.

The next step is to compile the test grid. The test grid contains the material, indicator, and item number of the question. The selected calculus material is a limit of algebraic functions, derivatives of algebraic functions, and an integral indefinite algebraic function. The chosen

indicator is an indicator that is adjusted to the standard competency and basic competencies of mathematical subjects on the calculus material. The diagnostic test is developed in the form of two-tier multiple choice, i.e. the first level is a question with five answers, while the second level contains the reason or the student's calculation of the answers.

Based on the grids specified in the previous step, writing the question is done by adjusting the indicator on the created grid. Based on the grids that have been created, the 11 indicators are lowered to 15 items with each item having five alternate answers. Fifteen items were developed consisting of four grains of the domain limit of algebraic function, six grains of the domain of algebraic function, and five of the domains integral not necessarily the function of algebraic.

Study of the contents or validation is done by three experts who are of the Faculty of the Mathematics Education Studies Program. The research is analyzed quantitatively and qualitatively. Quantitatively each test item is score and analyzed with a formula Aiken. The score of each test item on the validation sheet is between 1 and 5. While the question is qualitative in the form of summaries of the opinions of each expert for the improvement of grain items. In addition to test instruments, instrument interview guidelines are also validated in the same way. The score of each question on the interview guidelines is between 1 and 4. For the assessment of each item, a validation sheet is given to each member. The result of the expert validation sheet fill is then analyzed with the Aiken formula that will result in the validity index (V).

The range of V digits that may be obtained is between 0 and 1. The higher the V number or the closer the value is 1 then the validity of an item/grain is also higher, and if the V number approaches 0 then the eligibility an item/item is also getting lower. The following is the result of the validity index (V) calculation.

Table 1. Index validity of Two-Tier diagnostic test Instruments

Items	Validator 1	Validator 2	Validator 3	Index validity (V)	Description
1	4	5	5	0,9167	Valid
2	5	4	5	0,9167	Valid
3	5	5	4	0,9167	Valid
4	3	5	5	0,8333	Valid
5	5	5	4	0,9167	Valid
6	4	5	5	0,9167	Valid
7	4	4	5	0,8333	Valid
8	4	5	5	0,9167	Valid
9	5	5	5	1	Valid
10	5	4	5	0,9167	Valid
11	5	5	5	1	Valid
12	5	5	5	1	Valid
13	5	5	5	1	Valid
14	5	4	5	0,9167	Valid
15	5	5	5	1	Valid
Average				0,9333	Valid

Table 2. Validity Index

Items	Validator 1	Validator 2	Validator 3	Index validity (V)	Description
1	4	4	3	0,8889	Valid
2	3	4	3	0,7778	Valid
3	2	4	3	0,6667	Valid
4	3	3	3	0,6667	Valid
5	2	4	3	0,6667	Valid
6	3	3	3	0,6667	Valid
7	3	3	3	0,6667	Valid
8	3	4	4	0,8889	Valid
9	3	4	3	0,7778	Valid
10	4	4	3	0,8889	Valid
Average				0,7556	Valid

Based on the validation result of the diagnostic test instrument two-tier multiples choice calculus material as in Table 1 it is obtained that each validator provides an assessment with the final result of its validity index of more than 0.8 which means high validity. So in general it can be concluded that the diagnostic test instrument two-tier multiples choice calculus material in this study is valid, which means the diagnostic test instrument two-tier multiples choice calculus material has fulfilled each indicator of the problem and is valid for analyzing the mistakes of students.

Non-test instruments in the form of interview guidelines are also validated by experts. Based on the validation results it is obtained that the interview guidelines are provided with valid categories. Guidelines validation results along with an interview instrument for each criterion are met. It indicates that the guidelines and the interview instruments are valid for use.

3.2 Validity of Construction of Diagnostic Test Instruments

The product trials were carried out at SMA Yogyakarta in 35 students of the grade XII IPA. Based on the results of the test product data obtained will be used to analyze the validity of the construction and reliability,

Table 3. Results of KMO and Bartlet test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	Approx. Chi-Square	.753
		349.916
Bartlett's Test of Sphericity	df	105
	Sig.	.000

The validity of the construction is evidenced by the Exploratory Factor Analysis (EFA) using SPSS. The result of the analysis of the factors on the adequacy of the samples showed Khi-squared value in the Bartlet test of 349.916 with a degree of freedom 105 and a P-value of less than 0.01. Also acquired Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) of 0.753. These two points indicate that the sample size used in the analysis of this factor has been adequate. More results can be seen in the

Table 3.

Based on the value of Eigen and component variance analysis result factors can be obtained that the student's response data to the diagnostic test of two-tier multiples choice material calculus SMA contains 3 Eigen values greater than 1, so it can be said that the two-tier multiples choice diagnostic test contains 3 factors. It is also strengthened by the results of the scree-plot of Eigen value, which is derived graph from three components while the other shows the ramps graph. These results indicate that there are 3 dominant factors measured in the diagnostic test instrument of two-tier multiples choice calculus material.

Scree Plot

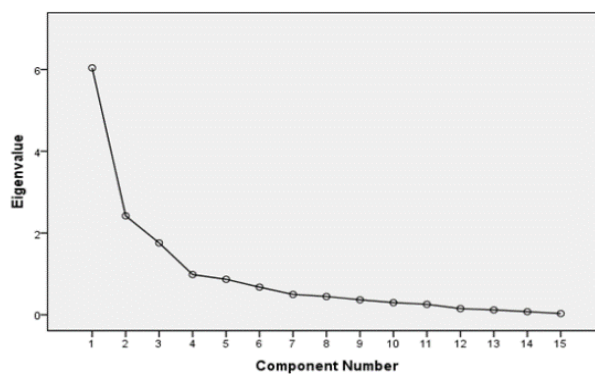


Figure 1. Scree plot

The number of factors contained in the instrument can be known from the scree-plot as in Figure 1. Many factors are characterized by the pouring of the chart of Eigen value acquisition. Figure 1 shows that there are 3 factors measured in the diagnostic instrument of two-tier multiples choice calculus material.

The next Eigen value can be presented with a scree plot on Figure 1. Based on the results of the plot scree can be seen that the value of Eigen began to rise in the 1st factor. Meanwhile starting from the 4th Factor until 15th factor show that the value of Eigen is stably decrease. It indicates that a diagnostic test device of two-tier multiples choice calculus material measures 3 dominant factors. Here are given a list table of Eigen values.

Table 4. Eigen value and component variance result factor analysis

Component	Eigen value	Difference values of Eigen	Proportion (%)	Cumulatif (%)
1	6,040	3,62	40,268	40,268
2	2,420	0,661	16,135	56,403
3	1,759	0,775	11,728	68,130
4	0,984	0,116	6,559	74,689
5	0,868	0,191	5,789	80,478
6	0,677	0,178	4,526	84,993
7	0,499	0,05	3,326	88,320
8	0,449	0,082	2,990	91,310
9	0,367	0,068	2,448	93,758
10	0,299	0,041	1,993	95,751
11	0,258	0,106	1,717	97,468
12	0,152	0,032	1,017	98,485
13	0,120	0,043	0,801	99,286
14	0,077	0,047	0,511	99,797
15	0,030		0,203	100,000

After determining the number of factors contained, the next will be the naming factor. The naming factor is done based on the load factor after rotation, taking into account the magnitude of the payload of the most factors on each component or item. The naming factor contained in the instrument test of two-tier multiples choice calculus material is carried out by researchers based on the indicators and the arrangement of the grid instrument. The load of unrotated factors is presented in table 4 and the payload of the rotated factor is presented in the following table 5.

Table 5. Charge factor after rotation

Item	Integral	Limit	Derivative
Item_1	.076	.880	.213
Item_2	.188	.934	.211
Item_3	.108	.903	.273
Item_4	.370	.620	.079
Item_5	.100	.019	.559
Item_6	-.123	.159	.752
Item_7	.276	.307	.566
Item_8	-.013	.241	.852
Item_9	.246	.268	.714
Item_10	.809	.393	-.129
Item_11	.861	.270	.099
Item_12	.855	.098	.062
Item_13	.877	.034	.194
Item_14	.442	.009	.516
Item_15	.578	.116	.399

Based on the exploratory factor analysis can be

concluded that the diagnostic test instrument two-tier multiple choice of calculus High school material is valid to measure the student's mathematical skills in calculus SMA material.

3.3 Reliability of Diagnostic Test Instruments

Reliability refers to the consistency of the test score or other measurement results of a measurement to another measurement. In other words, a test is said to be reliable if the results of its measurements approach the actual state of the student or is able to distinguish between students who are clever and not. According to Ebel and Frisbie [9] reliability of the instrument is fulfilled if Cronbach's value is $\alpha \geq 0.65$. The result of the reliability of the two-tier diagnostic test instruments using SPSS was obtained by Cronbach's Alpha 0.780, making it larger than 0.65. Based on this, there is a conclusion that the prepared two-tier diagnostic test instruments are reliable. The results of reliability estimation using SPSS can be seen in the following table.

Table 6. Reliability Estimation Results

Cronbach's Alpha	N of Items
0,803	15

4. Discussion

Diagnostic test Instruments Two-tier multiples choice calculus material is made as many as 15 questions. Diagnostic test Instruments Two-tier multiples choice calculus material is validated at 35 students in one of the state high school in Yogyakarta City. In addition to being

validated on students, such test instruments are also validated by experts and validation results by experts mentioning that two-tier multiples choice diagnostic instrument calculus material is worth using to analyze students' mistakes on calculus material.

Furthermore, to prove the validity of the construct proved with exploratory factor analysis using the help of SPSS program. The analysis of the factors conducted using Bartlett test resulted in a KMO value of 0.753. The KMO value is already more than 0.5 which means the samples used in this study were sufficient. Moreover, to see the Eigen value that is above 1. There are 3 factors that have an Eigen value above 1 and the difference between the three factors is also quite a lot. As for the other factors the difference is not more than 0.2. The same thing is also noticeable when noticing the scree plot in Figure 4 which indicates there are 3 dominant factors measured on this test. It is thus evident that this two-tier multiple-choice diagnostic test device is valid to measure students' mistakes in calculus high school material.

Furthermore, after the two-tier multiple-choice diagnostic test instrument on the calculus material proved to be valid, researchers tested the test instrument to analyze students' mistakes on calculus material. The student's fault in solving the diagnostic test problem of two-tier multiple-choice calculus material is seen based on the results of the diagnostic test provided. The test was given to 5 SMA Negeri in Yogyakarta with 551 students as the subject of research. After accumulated all the students' answer sheets, the researcher then corrected to see how many students answered correctly and answered wrong in each item. Once corrected for the wrong student answers it will be analyzed deeper to see the types of mistakes students are doing.

5. Conclusions

Based on the results of research and discussion above it can be concluded that the construction of a two-tier multiples choice diagnostic test based on the validity of the contents and the construct was obtained that the two-tier multiples choice diagnostic test is proven valid. The validity of the content is evidenced by the average validity index (V), for the two-tier multiples choice diagnostic test instrument obtained an average validity index (V) of 0.9333 and for an interview guideline instrument acquired the validity index (V) 0.7556 in which both the validity index (V) approaches the value 1. Whereas for the validity of the construct acquired three dominant factors based on

the scree-plot and corresponds to many factors on the calculus material examined in this study. The quality of two-tier multiples choice diagnostic tests is compiled of two-tier diagnostic test instruments based on the reliability value gained. It is shown with the obtained value of Cronbach's alpha 0.780 which is greater than 0.65.

REFERENCES

- [1] Santrock, J. W. (2011). *Educational Psychology*. (5th ed). New York: McGraw-Hill Company.
- [2] Chambers, P. (2008). *Teaching Mathematician, Developing as A Reflective Secondary*. London: SAGE
- [3] Yeo, K.K.J. (2009). Secondary Students' Difficulties in Solving Non-Routine Problems. Research in Mathematics Education in Singapore. Retrieved from <https://eric.ed.gov/?id=EJ904874>
- [4] Wijaya, A., Heuvel-Panhuizen, M., Doorman, M., Robitzsch, A. (2014). Difficulties in Solving Context-based PISA Mathematics Tasks: An Analysis of Students' Errors. *The Mathematics Enthusiast*, 11(3). hlm: 555-584
- [5] Herutomo, R.A. & Saputro, T.E.M. (2014). Analisis Kesalahan dan Miskonsepsi Siswa Kelas VIII pada Materi Aljabar. *Jurnal Ilmu Pendidikan dan Pengajaran*, 1(2), 134-145. <https://doi.org/10.17509/edusentris.v1i2.140>.
- [6] Suryabrata, S. (2008). *Metodologi Penelitian*. Jakarta: Rineka Cipta.
- [7] Non Syafriaedi, Ahmad Fauzan, I Made Arnawa, Syafrri Anwar, Wahyu Widada, (2019). The Tools of Mathematics Learning Based on Realistic Mathematics Education Approach in Elementary School to Improve Math Abilities. *Universal Journal of Educational Research*, 7(7), 1532 - 1536. DOI: 10.13189/ujer.2019.070707.
- [8] Retnawati, H. (2016). *Validitas reliabilitas & karakteristik butir*. Yogyakarta: Parama Publishing.
- [9] Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of Education Measurement*. New jersey: prentice hall
- [10] Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The Development of an Instrument to Measure the Higher Order Thinking Skill in Physics. *European Journal of Educational Research*, 8(3), 743-751. doi: 10.12973/eu-jer.8.3.743
- [11] Ramadhan, S., Sumiharsono, R., Mardapi, D., & Prasetyo, Z. K. (2020). The Quality of Test Instruments Constructed by Teachers in Bima Regency, Indonesia: Document Analysis. *International Journal of Instruction*, 13(2). doi: 10.29333/IJI.2020.13235A