

## The utilization of junior high school mathematics national examination data: A conceptual error diagnosis

\*<sup>1</sup>Kartianom; <sup>2</sup>Djemari Mardapi

\*Graduate School of Universitas Negeri Yogyakarta  
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\*Email: [kartianom@gmail.com](mailto:kartianom@gmail.com)

*Submitted: 23 January 2018 | Revised: 22 February 2018 | Accepted: 26 February 2018*

### Abstract

The goal of the research is to gain insights into the characteristics of the items in the mathematics national examination, the attributes on which the items were formulated and the result of a conceptual error diagnosis of the mathematics materials based on the result of the junior high school mathematics national examination. This is quantitative descriptive research. The data were collected from 3,079 grade-nine students of junior high schools who took the National Examination in the academic year of 2015/2016. The sample was established randomly based on the package code of the examination which is P0C5520 with 574 students as the examinees. Documentation method was applied in collecting the data. The result of the research shows that – upon the implementation of the classical test theory – there are 16 items in ‘difficult’ category, 24 in ‘intermediate’ category, and no items in ‘easy’ category. Furthermore, upon the implementation of the item response theory, the result shows that 28 items are in ‘good’ category and 12 items are in ‘poor’ category. In addition, there are 50 attributes on which the Junior High School Mathematics National Examination test (package P0C520) is formulated. Four attributes are content attributes and the rest (46) are process skill attributes. The result of the diagnosis shows that there are 11 types of errors made by the students when trying to complete the content items. Most of the errors are conceptual errors related to the geometric materials especially in the sub-materials of polyhedron, triangles, and quadrangles.

**Keywords:** *conceptual error, attributes, junior high school mathematics national examination*

### How to cite item:

Kartianom, K., & Mardapi, D. (2017). The utilization of junior high school mathematics national examination data: A conceptual error diagnosis. *REiD (Research and Evaluation in Education)*, 3(2), 163-173. doi:<http://dx.doi.org/10.21831/reid.v3i2.18120>

### Introduction

In the education system, evaluation is an urgent thing to perform. Evaluation is a medium to put students in the context of what they understand and what they are able to perform, while describing what they do not understand and what they are not able to perform (Sumintono & Widhiarso, 2015, pp. 2–3). The goal of the evaluation on the result of the study as conducted by the government is

to measure the competence level of the graduates on certain subjects as formulated in National Examination (or *Ujian Nasional* – UN). The items in National Examination are formulated based on the competence standards of the graduates, basic competence and achievement indicator.

Most of the education practitioners utilize the reports on the result of the National Examination as the supporting data in the process of policy-making, as a medium in

comparing the achievement of the examinees in the national level and as a medium in mapping the quality of national education. For example, the report of the Junior High School National Examination result for Mathematics in Baubau Municipality in the academic year of 2014/2015 shows that the average score on Mathematics is ( $\bar{X}$ ) 42.62 with 15.0 as the lowest score and 97.5 as the highest score (Ministry of Education and Culture, 2015). The result indicates that some examinees gave incorrect responses to some of the items of the Mathematics National Examination. The mistakes might be caused by the level of the items in the examination and the examinees' lack of conceptual knowledge or because they made a conceptual errors.

A good examination item must go through a calibration process, so the information on the items can be gained from the applied test. This information is commonly called characteristics of the items, which can be estimated by using two approaches, namely: Classical Test Theory (CTT) and Item Response Theory (IRT). A good item can be reviewed from its difficulty level, discrimination index, and distractor effectiveness. In the CTT approach, the index of the difficulty level of a good item must be 0.3 – 0.8, while the discrimination index must be  $\geq 0.3$  and the option of each item at least has to be selected by 5% of the examinees (Mardapi, 2012, p. 128). In the IRT approach, the index of the difficulty level of a good item must be (*ai*) -2.0 – +2.0 (Hambleton, Swaminathan, & Rogers, 1991, p. 13), while the discrimination index must be (*bi*) 0 - +2.0 (Hambleton et al., 1991, p. 15), and pseudo guessing index must be (*ci*)  $0 - 1/k$  (Hambleton et al., 1991, p. 17).

Items with very low or very high facility index cannot be categorized as good items because they cannot differentiate the level of ability of the examinees. The error indication of the examinees can be caused by the difficulty level. It might not be caused by the lack of competence. Items with negative discrimination index indicate that the correctness of the answer is questionable. The correctness of the answer is also questionable if the distracting items are only selected by  $<5\%$  of the examinees. The examinees with the pseudo

guessing index  $>1/k$  show that the distracting items are not able to attract those with low capability (Abadyo & Bastari, 2015).

A conceptual error is an error in understanding the concept in which the understanding is not in accordance with the scientific definition as agreed generally by the experts in that field. In mathematics, this error happens when students fail to relate the initial concept with the newly-given one (Russell, O'Dwyer, & Miranda, 2009, p. 416). In fact, a conceptual error is closely related to the conceptual knowledge of the examinees. Mathematics conceptual knowledge is the examinees' understanding of the scope of the field of mathematics. The scope of mathematics subject include: (1) number, (2) algebra, (3) geometry and measurement, and (4) statistics and probability. Therefore, in mathematics, a conceptual error can be defined as an incorrect use of the concepts which do not follow the scientific definition in the scope of mathematics field (numbers, algebra, geometry, and measurement and statistics and probability).

In order to learn about the error indication related to a conceptual error, there should be diagnosis process. The goal of the diagnosis activity is to understand the strength and weakness of the examinees (Leighton & Gierl, 2007, p. 242). The cognitive diagnosis model (CDMs) can be utilized in two ways, (a) retrofitting (post-hoc analysis) from non-diagnostic examination to gain richer or wider information and (b) designing or constructing a set of items for diagnostic purposes (Ravand & Robitzsch, 2015, p. 3). In the approach of retrofitting (post-hoc analysis), non-diagnostic examination instruments are reconstructed in a way that they can be used to identify the strength and weakness of the examinees in defining the attributes based on which the test items are formulated.

Attributes are the description of knowledge in completing examination contents in a certain domain (Wang & Gierl, 2011, p. 166) and the basis of cognitive or skill process crucial to completing the test items (Gierl, Cui, & Zhou, 2009, p. 5; Gierl, Zheng, & Cui, 2008, pp. 66–67; Yamtinah & Budiyo, 2015, p. 71). In mathematics, attributes consist of three categories: content attributes (common

materials), process attributes (expected capability after learning the materials in the content attributes) and skill attributes (specific mathematical skills critical in certain materials) (Tatsuoka, 2009, p. 2). Attributes utilized in this research are content attributes and process skill attributes.

There are already many studies taking advantages of diagnosis activities in Indonesia. However, most of them focus on the development of the diagnostic instruments. Secondary data such as national examination, PISA and TIMSS are rarely used in diagnostic activities. If we take a look at the studies in the last six years (2011-2017), secondary data have been a fresh medium to gain information on the influential factors in the academic achievement of examinees (Kartianom & Ndayizeye, 2017, p. 200) and the difficulty of the examinees in completing the mathematics test items of the National Examination (Isgiyanto, 2011, p. 308; Retnawati, 2017, p. 33). Even though National Examination is neither the main factor in determining the passing of the examinees, nor the main requirement in continuing to higher education level, the result of the National Examination is valuable data for diagnostic purposes.

To be more specific, the poor result of the Junior High School National Examination in Baubau Municipality was driven by the lack of comprehensive diagnosis on the result of the National Examination, especially on the subject of Mathematics. Both of the academia and the municipality administrator do not seem to see diagnostic activities as an urgent matter. The data of the National Examination are left untouched and have not yet been transformed into insightful information. The objective of this research is to gain insights into the characteristics of the test items and see the result of the diagnosis on the conceptual error in mathematics materials based on the result of the Junior High School Mathematics National Examination in Baubau Municipality.

## Method

This research is quantitative descriptive research which applies content analysis in drawing conclusion by identifying various

characteristics specifically in a message – in the test items and the responses of the examinees - objectively, systematically and generally. The research was conducted in Baubau Municipality. The data were collected from the Center for Education Evaluation (commonly known as PUSPENDIK) in Jakarta, in the form of National Examination sheets and the response sheets.

The data source is the ninth graders of junior high schools in the academic year of 2015/2016 in Baubau Municipality. The total number of the examinees is 3,079. The sample was established randomly (random sampling) based on the package code of the examination content. The researchers selected the package code of P0C5520 with 574 examinees in total. The object of the research is 40 test items and 22,960 responses of the examinees.

The *expost facto* data in the form of the the examinees' responses and the items in the Junior High School Mathematics National Examination were collected using documentation technique. The data were analyzed for diagnostic information. The items in the National Examination were selected to be the data because they had been standardized. Therefore, the bias has been minimized. Moreover, they had been calibrated, which allowed the researchers to compare the existing series and the packages from each year.

A good examination instrument must be valid and reliable. In this research, the instruments chosen are the instruments of the National Examination which have been tested in large and small scales. Therefore, it is safe to assume that the validity and reliability of the instruments are fulfilled. The validity implemented in this research is closely related to the attribute formation. The validity of the content of the attributes on which the test items are formulated was proven based on the judgment of the experts. In order to produce the content validity index of the attributes formation, the result of the judgment was then calculated using Aiken formulation. Based on the Aiken index, the researchers formulated criteria in order to show the content validity of the attributes formation (see Table 1) (Kartianom, 2017, p. 153).

Table 1. Content validity index criteria

Aiken Index	Content Validity Criteria
> 0.4	Low
0.4 – 0.8	Medium
> 0.8	High

In order to understand the characteristics of the items using CTT approach, the data were analyzed using TAP software version 14.7.4. Table 2 shows the criteria of good items based on CTT approach (Mardapi, 2012, p. 128).

Table 2. Item characteristic criteria using CTT

Parameter	Criteria
$a_i$	More than or equal with 0.3
$b_i$	0.3 to 0.8
$c_i$	The answer choice is chosen by at least 5% of the examinees

Description:

- $a_i$  = Items differentiators index
- $b_i$  = Items difficulty level index
- $c_i$  = Distractor effectiveness index

Using IRT approach, the data were analyzed with the help of Bilog-Mg software. Prior to the analysis, the sample was tested for its adequacy using SPSS11.5 software. The sample is considered adequate when the value of *Kaiser Mayer Olkin Measure* (KMO) > 0.5 with significance value (Sig.) of < 0.05. After that, the assumption test was conducted on the item parameter estimation using IRT approach. The assumption to be fulfilled was local unidimension and independency. Unidimension assumption was conducted with the support of SPSS 11.5 software based on the formation of the dominant factor. The formulated factor was with the Eigen value > 1.0. The dominant factor has large Eigen value discrepancy with the next factor and it has at least 20% cumulative frequency (Retnawati, Munadi, & Al-Zuhdy, 2015). The local independency assumption will be automatically fulfilled when the unidimensional assumption is fulfilled (Retnawati, 2014, p. 141).

When the assumption in IRT approach has been fulfilled, the next one is goodness of fit test. There are three models in IRT approach: model 1-PL, model 2-PL and model 3-PL. The goodness of fit test is conducted with the support from Bilog-Mg software by

comparing the significant value of  $\chi^2$  with  $\alpha = 0.05$  and also ICC curve. If the value of sig.  $\chi^2 > \alpha = 0.05$ , the items can be categorized as fit with the model. For ICC curve, the data are considered fit when the distribution of the data matches the model (Figure 1).

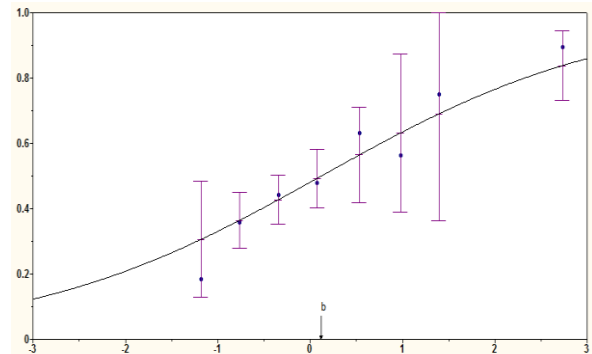


Figure 1. ICC curve

In each model, the criteria of good items in the IRT approach are presented in Table 3 (Hambleton et al., 1991, pp. 13–17).

Table 3. IRT criteria of items characteristics

Model	Parameter Criteria		
	$a_i$	$b_i$	$c_i$
1-PL	0 up to +2	-	-
2-PL	0 up to +2	-2 up to +2	-
3-PL	0 up to +2	-2 up to +2	0 up to 1/k

Description:

- $a_i$  = Item discrimination index
- $b_i$  = Items difficulty level index
- $c_i$  = Pseudo guessing index

In this research, the error made by the examinees was analyzed through the response of the Mathematics examination contents (answer sheets of the examinees) of the National Examination in the academic year of 2015/2016. The analysis was conducted by formulating the probable description of the alternative response to the test items. At this point, the researchers did not use the description of the examinees' answers and the responses to determine the achievement of the students, but to understand the type and the area of the error.

In order to conduct the diagnosis on the a conceptual error made by the examinees, the researchers: (1) identified the attributes of the examination content by defining the op-

tions of responses to each item using the content analysis; (2) named the type of the error in each response option based on the attributes on which the items were formulated; (3) analyzed the response option using TAP software version 14.7.4 to measure the percentage of each type of error in each material. There was a follow up for the most dominant type of error in order to understand the area of the error.

### Findings and Discussion

#### The Characteristics of the Test Items

##### Classical Test Theory

To understand the difficulty level, differentiator, and distractor effectiveness of the examination content, the researchers applied the classical test theory when analyzing the items. The data were in the form of answer sheets - multiple choices with the answer key. Table 4 shows the result of the recapitulation of the characteristics of the test items based on the difficulty level of the items in each material.

Table 4. The difficulty level of the items in each material

Materials	Category			Total
	Easy	Medium	Difficult	
Numbers	0	7	4	11
Algebra	0	4	6	10
Geometry	0	9	4	13
Statistics	0	3	1	4
Probability	0	1	1	2
<b>Total</b>	<b>0</b>	<b>24</b>	<b>16</b>	<b>40</b>

Table 4 shows that: (1) the materials on number have seven items in ‘medium’ category and four items in ‘difficult’ category; (2) the materials on algebra have four items in ‘medium’ category and six items in ‘difficult’ category; (3) the materials on geometry have nine items in ‘medium’ category and four items in ‘difficult’ category; (4) the materials on statistics have three items in ‘medium’ category and one item in ‘difficult’ category; and (5) the materials on probability have one item in ‘medium’ category and one item in ‘difficult’ category.

Table 5 shows the result of the recapitulation of the characteristics of the test items based on the differentiators of the items in each material.

Table 5. The differentiators of the items in each materials

Materials	Category		Total
	Good	Not Good	
Numbers	9	2	11
Algebra	6	4	10
Geometry	8	5	13
Statistics	1	3	4
Probability	2	0	2
<b>Total</b>	<b>26</b>	<b>14</b>	<b>40</b>

Table 5 shows that overall the discrimination index of the test items in the content of the Mathematics National Examination in Baubau Municipality has 26 items in ‘good’ category and 14 items in ‘not good’ category. If we take a closer look at the materials: (1) the materials on numbers have nine items in ‘good’ category and two items in ‘not good’ category, (2) the materials on algebra have six items in ‘good’ category and four items in ‘not good’ category, (3) the materials on geometry have eight items in ‘good’ category and five items in ‘not good’ category; (4) the materials on statistics have one item in ‘good’ category and three items in ‘not good’ category; and (5) the materials on probability have two items in ‘good’ category and no item is in ‘not good’ category.

Other critical information in the classical test theory is distractors effectiveness. The distribution of the response choice can be considered as effective or acceptable when each option in the test items is chosen by at least 5% of the examinees (Mardapi, 2012, p. 129). Figure 2 presents the functionality percentage of the distracting items.

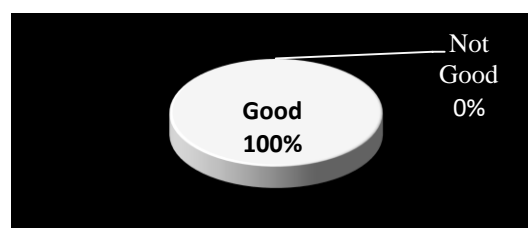


Figure 2. The functionality percentage of the distractors

Figure 2 shows that 100% of the items have effective distractors. This means the distractors in the items of the Junior High School Mathematics National Examination in Baubau Municipality are well-functioned distractors. In other words, they are able to attract the examinees.

#### Item Response Theory

Principally, the item response theory uses the probabilistic model. There are three analytic models: 1PL, 2PL and 3PL. In order to correctly select analytic model, the goodness of fit test is a crucial process. However, before that, the sample adequacy and assumption test has to be conducted. Table 6 shows the result of the sample adequacy test.

Table 6. The result of the *KMO* and *Bartlett* KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.810
Bartlett's Test of Sphericity	Approx. Chi-Square	2425.233
	df	780
	Sig.	0.000

Table 6 shows that the KMO value is at 0.810 or 0.5 higher. This means that the sample used in this research is adequate. Next, unidimensional assumption test was conducted while considering the scree plot (Figure 3).

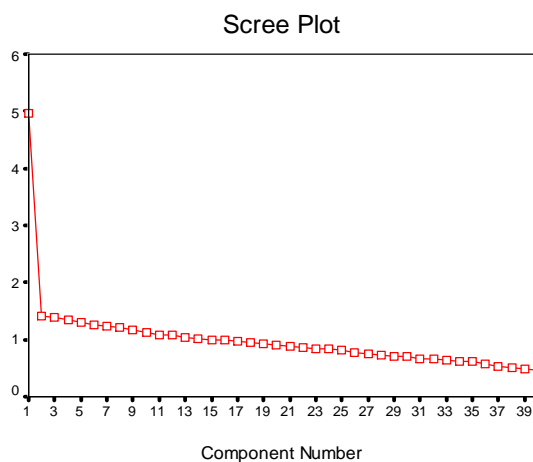


Figure 3. The scree plot of the result of the exploratory factor analysis

The scree plot in Figure 3 shows that there is one dominant factor in the Junior High School Mathematics National Exami-

nation in the academic year of 2015/2016 in Baubau Municipality. This can be seen from the shift in the Eigen value of the first factor up to the second factor. In the second factor and beyond, the shift of the Eigen value is not too high. Therefore, it is safe to conclude that the unidimensional assumption test on the contents of the Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality has been fulfilled. When the unidimensional assumption test has been fulfilled, the local independency assumption is automatically fulfilled. This also means that there is a correlation among the factors in the Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality, so the goodness of fit test can be conducted. The goodness of fit test for models 1-PL, 2-PL and 3-PL is conducted by comparing the significant value of  $\chi^2$  with  $\alpha = 0.05$  and ICC curve. Table 7 shows the result of the goodness of fit test for 1-PL, 2-PL and 3-PL.

Table 7. The result of the goodness of fit between the items and the model

Fitting Model	Fitting Items		
	Model 1-PL	Model 2-PL	Model 3-PL
Sig. <i>Chi-Square</i> Value	24	35	13
Using ICC curve	5	12	2

Table 7 shows that based on the goodness of fit test, 24 items fit with model 1-PL, 35 items fit with model 2-PL and 13 items fit with model 3-PL. When the goodness of fit test with ICC curve is applied, five items fit with model 1-PL, 12 items fit with model 2-PL and two items fit with model 3-PL. This makes model 2-PL the fittest analytic model.

The parameter used in model 2-PL is the difficulty level (*bi*) and differentiators (*ai*), whereas guessing (*ci*) for the item is considered zero. The items which fit with model 2-PL are brought to the next analytic step. The items are as follows, items 1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 and 40. In model 2-PL, the items that do not fit with model 2-PL are not included



in the next analytic steps even though they have difficulty and differentiators as the parameter. These excluded items are items 6, 11, 18, 23 and 28.

Table 8 shows the result of the characteristics analysis on the test items based on model 2-PL with the support from Bilog-MG program.

Table 8. The characteristics of the test items based on the parameter of difficulty level and differentiators

Category	Parameter Frequency		Desc.
	a	b	
Good	35	28	28
Not Good	0	7	7
<b>Total</b>	<b>35</b>	<b>35</b>	<b>35</b>

Table 8 shows that based on the criteria of model 2-PL, there are 28 items in ‘good’ category and 7 items in ‘not good’ category. In fact, those 7 items in ‘not good’ category possess good differentiators but have bad difficulty level. Those items are items 33, 9, 15, 29, 19, 21, and 35. Respectively, their difficulty level parameters are 4.463, 4.027, 3.870, 2.747, 2.644, 2.100, and 2.028. These items have very high difficulty level with item 33 having the highest difficulty level. In terms of the differentiator’s parameter, 40 items fall in ‘good’ category. This strengthens the indication that the error in the examinees responses – specifically while trying to complete items 33, 9, 15, 29, 19, 21 and 35 – is not caused by the difficulty level. In addition to items parameter, the researchers also gain insights into the test information function as shown in Figure 4.

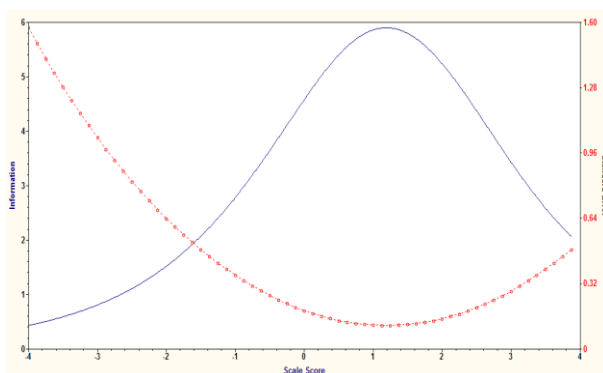


Figure 4. Information functions and test measurement error

Figure 4 shows that the content of Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality has higher information than the error in measurement with the ability range from -1.6 to +4.0. If the examination was delivered to the examinees with the ability range lower than -1.6 and higher than +4.0, the error in the measurement would be a lot higher than the information function.

### Subject-Matter Mastery in the Mathematics National Examination

The subject-matter mastery of the test takers of the National Examination of Mathematics of the academic year 2015/2016 can be seen from the proportion of true answers of the test takers on the number, algebra, geometry, statistics, and probability materials as presented in Figure 5.

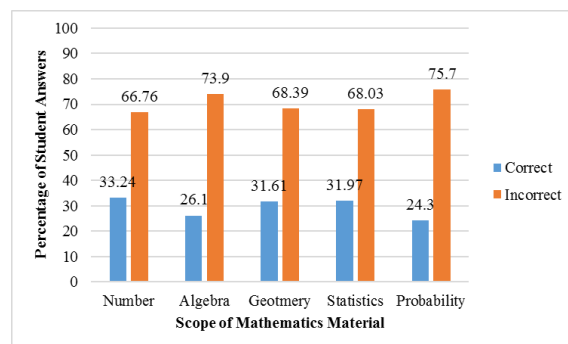


Figure 5. Percentage of student's answers to each material

Figure 5 shows that all materials tested on the Mathematics National Examination of the academic year 2015/2016 in Baubau Municipality are considered difficult by the test takers. This can be seen from the percentage of the wrong answers that are greater than the percentage of the correct answers of the test takers on each material.

### Attributes on which Test Items are Formulated

The attributes, on which the items are formulated, are developed and validated by five experts (expert judgment), three of whom are mathematics teachers of state junior high schools in Yogyakarta who previously had in-

volved in the development of the examination, and two are Mathematics lecturers. Generally, all of the attributes of the items of the Junior High School Mathematics National Examination in the academic year of 2015/2016 in Baubau Municipality consist of four content attributes and 46 process skill attributes. The content validity index of the attributes of those 40 items is at 0.888 which falls in 'high' category. Table 9 shows the distribution of the attributes of the items in each material.

Table 9. The distribution of the test items attributes

No	Material	Content Attributes	Process Skill Attributes
1	Numbers	1	13
2	Algebra	1	13
3	Geometry	1	14
4	Statistics and Probability	1	6
	<b>Total</b>	4	46

Table 9 shows the distribution of the attributes on which the test items are formulated. Each material competence has several attributes. Some of the attributes are alike and some are different. Thus, the material competence has to be divided into groups along with all of the attributes.

## Diagnosis of the Examinees' Errors

### Error Type

The identification of the error focuses on the attributes which are not mastered and applied correctly by the examinees when they are trying to complete the items in the Mathematics National Examination. Based on the content analysis, the errors can be categorized into 11 types, which consist of: (1) conceptual errors, (2) language-related interpretative errors, (3) procedural errors, (4) calculation errors, (5) representation errors, (6) conceptual and language-related interpretative errors, (7) conceptual and calculation errors, (8) conceptual and calculation errors, (9) language-related interpretative and procedural errors, (10) representation and procedural errors, and (11) representation and calculation errors. Figure 5 shows the percentage of each type of error.

Furthermore, in general, Table 10 shows the frequency of each type of errors. Table 10 shows that most of the errors are conceptual errors. They are in the area of basic concept of numbers, algebra, geometry (plane figure and solid figure) and probability. Most of them are found in geometric materials.

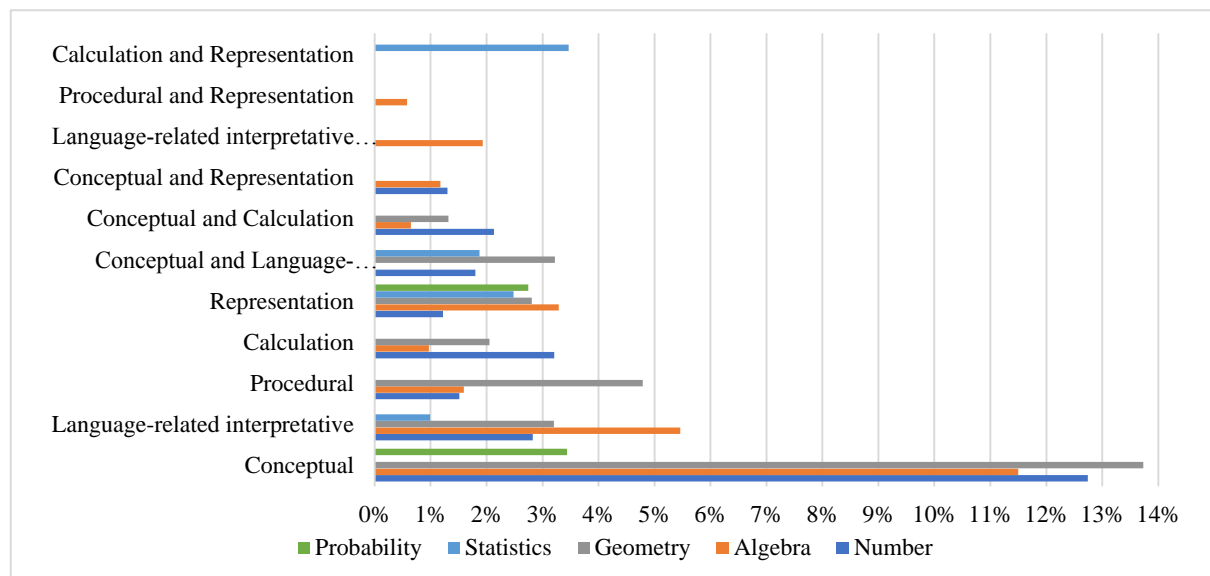


Figure 5. The percentage of each type of error in each material



Table 10. Types of errors made by the examinees

Types of Errors	Frequency	Percentage (%)
Conceptual	5804	41.41
Language-related interpretative	1749	12.48
Procedural	1106	7.89
Calculation	873	6.23
Representation	1759	12.55
Conceptual and Language-related interpretative	966	6.89
Conceptual and Calculation	575	4.10
Conceptual and Representation	347	2.48
Language-related interpretative and Procedural	271	1.93
Procedural and Representation	81	0.58
Calculation and Representation	486	3.47
<b>Total</b>	<b>14017</b>	<b>100</b>

#### *The Area of the Conceptual Errors*

The most dominant conceptual errors are: (1) the basic concept of integers in the materials of numbers, root form (irrational) and comparison; (2) the concept of relation and function, basic concept of algebraic operation, basic concept of integers and straight line equation in the materials of algebra; (3) the basic concept of geometry, polyhedron, triangles and quadrangles in the materials of

geometry; (4) the basic concept of probability in the materials of statistics. These all are shown in details in Figure 6.

#### Discussion

By using CTT and IRT, there are five items with a very high level of difficulty (Items 9,15,19,21, and 33). Item 9 is related to number; items 9, 15 and 21 are about algebra, while item 33 is related to geometry. The high percentage of students answering those items wrongly is due the very high level of item difficulty. Besides, the very high level of item difficulty indicates that there are a lot of students with incomplete attributes of those materials.

Based on the content analysis, there are 11 types of students' errors. The conceptual error is the dominant type of errors mostly occurred in geometry-related items. In line with the result of this research, Isgiyanto (2011) also found that, in Indonesia, the junior high school students are weak at geometry and measurement with the low level of attributes of content/concept completeness.

The conceptual errors made by the students are indicated by the conceptual errors occurring in number and algebra materials. The testees' understanding of numbers is the key to understanding the material of algebra. The understanding of numbers and algebra is the requirement for the understanding of the geometrical materials. Further, in their study,

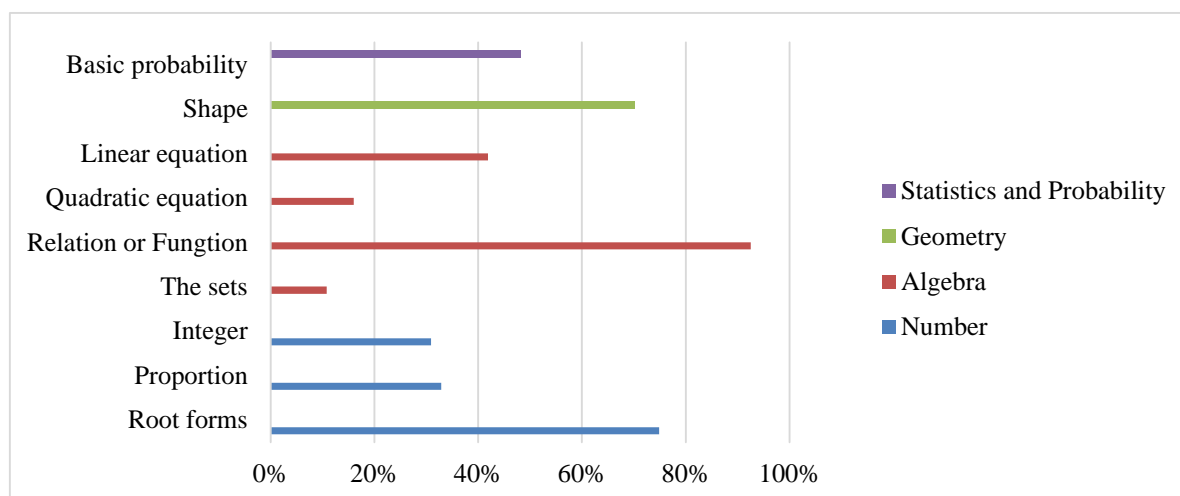


Figure 6. The area of error in each material

Russell et al. (2009, p. 416) mention that a conceptual error occurs because of the failure in connecting new concept with the earlier concept. Specifically, the conceptual error made by the students is located in the basic concept of integers, irrationals, comparisons, association and function, algebra operation, linear equation, polyhedron geometry, triangle, square, and probability.

The findings of this research are supported by the findings of a research conducted by Retnawati (2017, p. 33), which found that junior high school students in Yogyakarta, Indonesia found it difficult to finish the National Examination questions due to their disability to understand the concept of fraction, rationing fraction with square-root denominator, linear equation with one or two variables, determining the members of a sets, determining the gradient a linear equation, also the concept of area.

## Conclusion and Recommendations

### Conclusion

Based on the result of the analysis and description, it can be concluded that, *first*, based on the classical test theory, 16 test items are in 'difficult' category, 24 are in 'medium' category, and no item is in 'easy' category. Based on item response theory, 28 items are in 'good' category and 12 items are in 'not good' category. *Second*, there are 50 attributes – 4 content attributes and 46 process skill attributes - on which the Junior High School Mathematics National Examination content (package P0C5520) are formulated. *Third*, there are 11 types of errors made by the examinees when they tried to complete the examination. Most of the errors are conceptual errors in the materials of geometry especially in the sub materials of polyhedron, triangles and quadrangles.

### Recommendation

Based on the conclusion, the recommendations are: (1) for users of the diagnostic information. The result of the research can be used as the materials for training on the process of conducting diagnostic information. It is expected that this type of training can be

used to improve the quality of learning process in the schools with low result in the Mathematics National Examination. (2) For researchers, this research focuses only on diagnosis the types and areas of error made by the examinees when trying to complete Junior High School Mathematics National Test items based on the attributes of the items. Therefore, this research can be deepened by diagnosing the errors or difficulties faced by the examinees with the help of R packages CDM program while using model DINA.

## References

- Abadyo, A., & Bastari, B. (2015). Estimation of ability and item parameters in mathematics testing by using the combination of 3PLM/GRM and MCM/GPCM scoring model. *REiD (Research and Evaluation in Education)*, 1(1), 55–72.
- Gierl, M. J., Cui, Y., & Zhou, J. (2009). Reliability and attribute-based scoring in cognitive diagnostic assessment. *Journal of Educational Measurement*, 46(3), 293–313. <https://doi.org/10.1111/j.1745-3984.2009.00082.x>
- Gierl, M. J., Zheng, Y., & Cui, Y. (2008). Using the attribute hierarchy method to identify and interpret cognitive skills that produce group differences. *Journal of Educational Measurement Spring*, 45(1), 65–89. Retrieved from <https://pdfs.semanticscholar.org/0a0b/180342ee51f6121dd4e3199c9cc4df3bc377.pdf>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. New Delhi: Sage Publications.
- Isgiyanto, A. (2011). Diagnosis kesalahan siswa berbasis penskoran politomus model partial credit pada matematika. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(2), 308–325. Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/1099/1151>
- Kartianom, K. (2017). *Diagnosis kesalahan konsep materi matematika SMP berdasarkan hasil ujian nasional di kota Baubau*. Master Thesis, Universitas Negeri Yogyakarta,

- Indonesia.
- Kartianom, K., & Ndayizeye, O. (2017). What 's wrong with the Asian and African students' mathematics learning achievement? The multilevel PISA 2015 data analysis for Indonesia, Japan, and Algeria. *Jurnal Riset Pendidikan Matematika*, 4(2), 200–210. <https://doi.org/10.21831/jrpm.v4i2.16931>
- Leighton, J. P., & Gierl, M. J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, 26(2), 3–16. <https://doi.org/10.1111/j.1745-3992.2007.00090.x>
- Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Ministry of Education and Culture. (2015). *Laporan hasil ujian nasional*. Jakarta: Balitbang.
- Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research & Evaluation*, 20(11). Retrieved from <http://pareonline.net/getvn.asp?v=20&n=11>
- Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.
- Retnawati, H. (2017). Diagnosing the junior high school students' difficulties in learning mathematics. *International Journal on New Trends in Education and Their Implications*, 8(1), 33–50. Retrieved from [http://www.ijonte.org/FileUpload/ks63207/File/04.heri\\_retnawati.pdf](http://www.ijonte.org/FileUpload/ks63207/File/04.heri_retnawati.pdf)
- Retnawati, H., Munadi, S., & Al-Zuhdy, Y. A. (2015). Factor analysis to identify the dimension of Test of English Proficiency (TOEP) in the listening section. *REiD (Research and Evaluation in Education)*, 1(1), 45–54. <https://doi.org/10.21831/reid.v1i1.4897>
- Russell, M., O'Dwyer, L. M., & Miranda, H. (2009). Diagnosing students' misconceptions in algebra: Results from an experimental pilot study. *Behavior Research Methods*, 41(2), 414–424. <https://doi.org/10.3758/BRM.41.2.414>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan*. Bandung: Trim Komunikata.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. New York, NY: Routledge/Taylor & Francis.
- Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement*, 48(2), 165–187. <https://doi.org/10.1111/j.1745-3984.2011.00142.x>
- Yamtinah, S., & Budiyo, B. (2015). Pengembangan instrumen diagnosis kesulitan belajar pada pembelajaran kimia di SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(1), 69–81. <https://doi.org/10.21831/pep.v19i1.4557>