Characteristics of non-compulsory mathematics test items on nationally standardized school examination in Kaimana Regency, West Papua Indonesia

F Sampouw and H Retnawati

1 Graduate Program of Mathematics Education, Universitas Negeri Yogyakarta, Depok, Sleman, Yogyakarta, Indonesia
2 Department of Mathematics Education, Faculty of Mathematics and Natural Science, Universitas Negeri Yogyakarta, Depok, Sleman, Yogyakarta, Indonesia

E-mail: falenthinosampouw.2018@student.uny.ac.id

Abstract. Nationally Standardized School Examination result is used as a students’ graduation requirement. Therefore, it is absolutely necessary that its test items must be well constructed to assess. The aim of this study was to describe the characteristics of non-compulsory mathematics test items on Nationally Standardized School Examination of senior high school in Kaimana Regency, West Papua based on the items analysis. This study was a descriptive exploratory type of quantitative research. Data obtained from a set of Non-compulsory Mathematics test items including documentation of students’ answers of the following test items on Nationally Standardized School Examination of senior high school in Kaimana Regency, West Papua. A set of this test item was validated by 5 experts and analyzed by applying Classical Test Theory. Results showed that the content validity was valid for 0.9 as the average of index Aiken. The reliability of test items was in less reliable level. The items difficulty indicated that 29 items (85.3%) were hard and 5 items (14.7%) were moderate. The value of items discrimination index for 12 items (53.3%) were classified as good and 22 items (64.7%) were poor. For multiple-choice items, the distractor was made by the developer, worked effectively.

1. Introduction
Assessment is one of the most important components of education. In learning mathematics, assessment is an essential part to identify what students’ have successfully achieved [1]. Students’ assessment is mostly conducted through a test [2]. In addition, students’ achievement can be identified based on the test results [3].

In Indonesia, students’ assessment for senior high school students is conducted through National Examination and Nationally Standardized School Examination. National examination results would not only be an indicator to determine the quality of education but also an important factor to be considered on students’ admission for higher level of education [4]. In the other hand, the Nationally standardized school examination result is used as a students’ graduation requirement [5].

Nationally Standardized School Examination is a test to measure students’ achievement conducted by each educational level institution based on the graduate competency standard for the purpose of getting a learning achievement recognition [4]. Nationally Standardized School Examination result is an important determinant of students’ graduation. Therefore, it is absolutely necessary factor for students to focus on.
Nationally Standardized School Examination results as its function to determine students’ graduation needs to be valid to describe students’ abilities. Good quality of measurement tools is needed as an assessment to get a valid result to what needs to be measured [6]. Good quality of test items provides a well described description of students’ abilities. In addition, it should not be too easy that most students get the answer correct neither be too hard that most students get most answer incorrect. Thus, item test of Nationally Standardized School Examination needs to be analyzed. The question developers can recognize what items work or does not work to measure student performance through item analysis [7].

One of the methods to identify how good the quality of test item is by analyzing its characteristics [8]. There are two different approaches which can be applied to analyse a set of test item. They are classical and modern approach. An analysis needs to be applied to identify whether or not a set of test item is proper to be used as an assessment. Research therefore needs to be conducted to describe the characteristics of test items on Nationally Standardized School Examination. The result of item analyse study will describe students’ performance and help teachers to develop a good learning activity for the future. It also helps test item developers to know that test item qualified to assess student performance.

A classical approach is applied in this study to identify the characteristics of a set test item of Non-compulsory Mathematics on Nationally Standardized School Examination in Kaimana Regency, West Papua. There are several criteria need to be analyzed to recognize the quality of test items, they are items difficulty, items discrimination, items reliability and effectivity of distractor [9]. This analysis is necessary applied to identify how proper the test item of non-compulsory mathematics on Nationally Standardized School Examination is as an assessment.

2. Method
This study was a descriptive exploratory type of quantitative research. Data obtained from a set of Non-compulsory Mathematics test items including documentation of students’ answers of the following test items on Nationally Standardized School Examination year 2019 in Kaimana Regency, West Papua. There was only one school in Kaimana where an examination of non-compulsory mathematics on Nationally Standardized School Examination was first conducted as it was the only school where Curriculum of 2013 was implemented.

A set of test item of non-compulsory mathematics on Nationally Standardized School Examination consisted of item specification and its test item. Item specification included the context of material, cognitive level, learning objectives, items’ indicator, and test item. There were thirty multiple choice questions and five short answer questions.

Item specification of non-compulsory mathematics on Nationally Standardized School Examination was analyzed by five experts to identify how proper its components were based on its specifications. Index Aiken was used to determining test items’ validity. If the index is below 0.4 then validity classifies as poor, it is fair if its validity index is between 0.4 and 0.8, and good if its validity index is above 0.8 [8]. Documentation of students’ answer was a set of 106 students’ response that joined the non-compulsory mathematics examination on Nationally Standardized School Examination. It was analyzed by using Quest application to indicate its items’ reliability, items difficulty, and items discrimination. Items are said reliable if the value of internal consistency equal or greater than 0.85. Item difficulty is classified as hard if P value is lower than 0.3, classified as moderate if the P value in a range 0.3 and 0.7, and classified as easy if the P value greater than or equal to 0.7. The items discrimination is determined by categorizing the value of Point Biserial (Pt-Biserial). The item discrimination is classified as good if its Pt-Biserial value greater than or equal to 0.3.

3. Results
A set test item specification on Nationally Standardized School Examination was validated by five experts as validators to determine its validity, ensuring that each test item is relevant for each indicator based on what was written as students’ learning objectives. Validators used Likert scale between 1 and
5 to determine its validity indicating how relevant it was in meeting students’ learning objectives. The lower the number of Likert scale the less relevant test item was. It is not valid if its Likert scale is 1, slightly valid if it is 2, moderately valid if it is 3, valid if it is 4, and strongly valid if it is 5. Aiken index was calculated based on the result of validators’ analysis. There were 25.7% items classified as moderately valid and 74.3% indicated as strongly valid. The average of Aiken index obtained was valid for 0.9. To sum up, most of test items of non-compulsory mathematics on Nationally Standardized School Examination were valid as an assessment to measure students’ learning achievement. Table 1 below shows validity of each test item.

**Table 1. Test Items’ Validity**

| Validity Category | Items Number | Percentage |
|-------------------|--------------|------------|
| Strongly valid    | 1,2,5,6,7,8,9,10,11,12,13,14,15,16,17,19,21,22,24,25,26,31,32,33,34 | 74.3% |
| Moderately valid  | 3,4,18,20,23,27,28,30,35 | 25.7% |

A deeper analysis of test items was conducted to identify its reliability, items difficulty, and item discrimination. They were 35 items consisted of two different types of questions. They were 30 multiple choice questions and 5 short answer questions. However, there were only 29 multiple choice question were analyzed since one question from multiple choice question type, number 21 was not included to be analyzed as it has no answer. Therefore, only 34 items in total were analysed. Students were awarded 1 mark for every correct answer and 0 for every incorrect answer on short answer questions. Data obtained from students’ response to identify each items analysis. Data were analyzed by using Quest application. Table 2 below shows internal consistency of test items.

**Table 2. Result of Internal Consistency**

| N  | L  | SD | Internal Consistency |
|----|----|----|----------------------|
| 106| 29 | 2.71| 0.34 |
| 106| 5  | 1.14| 0.57 |

Table 2 above shows each internal consistency of two different types of test item. Internal consistency for 29 multiple choice questions was 0.34. However, internal consistency of 5 short answer questions was higher on 0.57. Test items’ reliability was then determined based on the lower bound of its internal consistency. A set test item on Nationally Standardized School Examination is reliable once its internal consistency calculation result is greater than or equal to 0.85 as its one of the assessments to determine each of students’ result on graduation [8]. In contrast, two different types of questions in this set test item both obtained a lower internal consistency calculation result than 0.85. Therefore, it was slightly reliable as an assessment.

Table 3 below shows the difficulty of each item which has been analyzed by using Quest application.

**Table 3. Result of item difficulty (P)**

| No. item | P   | Category | No. item | P   | Category | No. item | P   | Category |
|----------|-----|----------|----------|-----|----------|----------|-----|----------|
| 1        | 0.415 | moderate | 13       | 0.264 | Hard     | 25       | 0.245 | hard     |
| 2        | 0.132 | hard     | 14       | 0.208 | Hard     | 26       | 0.274 | hard     |
| 3        | 0.208 | hard     | 15       | 0.179 | Hard     | 27       | 0.179 | hard     |
| 4        | 0.330 | moderate | 16       | 0.189 | Hard     | 28       | 0.142 | hard     |
| 5        | 0.208 | hard     | 17       | 0.321 | Moderate | 29       | 0.104 | hard     |
| 6        | 0.198 | hard     | 18       | 0.217 | Hard     | 30       | 0.425 | moderate |
| 7        | 0.292 | hard     | 19       | 0.245 | Hard     | 31       | 0.434 | moderate |
Table 3 above shows that there are 5 questions classified as moderate and other 29 questions are classified as hard. It indicated that most items test of non-compulsory mathematics on Nationally Standardized School Examination were hard.

Table 4 below shows each of test item discrimination. Item discrimination is classified as good if its index is greater than or equal to 0.3 [8]. Table 4 shows that they are 12 items classified as good and other 22 items are classified as poor.

The distractor of multiple choice items needs to be analyzed. The indicator of good distractor of item is there are more than 5% of students choose the given choices of each item [10]. Based on the analyzed by Quest, it shows that there is no given choices of each item have percentage less than 5%. The smallest percentage is 6.6%. Thus, it can be concluded that the distractor works effectively.

4. Discussion

A set of test items were validated by five experts as validators with its Aiken Index 0.9, indicating that its test items of non-compulsory mathematics on Nationally Standardized School Examination is strongly valid. The percentage showed that 74.3% items were classified as strongly valid and 25.7% items were in moderately valid category. It shows that more items are strongly valid explaining that its test item is strongly relevant to indicator that students need to achieve.

In fact, the reliability of the test did not show the same trend as its validity. The reliability of the test was slightly reliable. One of the factors affecting test reliability is item difficulty [8]. The more items are classified as hard in item difficulty the lower its reliability is. Most of items in non-compulsory mathematics on Nationally Standardized School Examination are classified as hard. Item difficulty analysis by using Quest found that 29 items were classified as hard with its percentage 85.3% in total. There were on 5 items were in moderate category for its percentage was 14.7% only.
Lower reliability of test items in non-compulsory mathematics on Nationally Standardized School Examination is mainly affected by this factor. Assessor as a test maker needs to notice item difficulty on the test as an assessment. Most of test items in non-compulsory mathematics on Nationally Standardized School Examination are classified as hard for its item difficulty. There were 29 from 34 items in total were hard. It indicates that 85.3% of the questions are hard for students to answer. Additionally, very often item difficulty is determined differently by assessor compared to the item difficulty index obtained from an empirical analysis’ results. One of the main determinants is item difficulty is often assigned on the basis of an intuitive feel. A set test item of non-compulsory mathematics on Nationally Standardized School Examination for certain regency is designed by a group of non-compulsory mathematics teachers from different schools in that regency. Therefore, they may differently assign items difficulty as they have different intuitive feel about its difficulty.

There are two others factor effected higher level of difficulty in items difficulty analysis on students’ assessment. They are students’ ability and readiness. The higher the level of students’ ability and readiness, the more capable students to answer tests they take. The lowest item difficulty index was founded in item 32 which was classified as the hardest for its item difficulty. It was number 33 as one of short answer question type. Figure 1 below displays its item.

![Figure 1. Short essay question Item 32](image)

The test item above is designed to assess students’ ability in meeting an indicator to implement the derivative of function in solving problem in which students are asked to determine the area of trapezium. Students not only require a good understanding about derivative but also geometry to determine the answer for this question. However, derivative as one of the topic in calculus is hardly understood by students. Additionally, the percentage of its material absorption about calculus in Kaimana is low. It was found only 26.46% in 2017 and 29.21% in 2018 [11], [12]. Besides, they need visual and mathematical representation ability to identify what factors enlarge the area of trapezium. However, students’ ability about visual and mathematical representation is classified as low [13]. It shows that they have poor understanding in both abilities they need to determine the answer of this question. They find it absolutely harder to get its answer correct.

Another item which was classified as hard was item 29, a multiple choice questions type number 30. Item 29 designed to assess students’ ability to implement basic concept of Binomial Probability to solve problems. Figure 2 displays a multiple choice question number 30.
Most students answered incorrectly by choosing A. Several factors effecting students in answering questions incorrectly, those are students fail in understanding and implementing certain concepts, they are not careful in doing basic operation of mathematics [14]–[16]. In addition, most students apply formula incorrectly to determine the probability of an event [17]. Also, students’ ability of mathematical reasoning is poor in solving probability problems [18] while reasoning ability is essentially needed to solve mathematical problems [19].

A proper test items have its item discrimination classify as good indicating that its test item have the ability to differentiate between good and poor students. The result showed that there were only 12 from 34 items which classified as good for its item discrimination. It indicated that 35.3% items were classified as good. However, more items were classified as poor with its higher percentage which was 64.7%. Generally, the test items discrimination is classified as poor. The result found that item 16 has its item discrimination index negative which was -0.02. Figure 3 displays item 16.

Based on the analysis above, an assessor or teacher is absolutely need to improve quality of a set items test of non-compulsory mathematics which is going to be an assessment on Nationally Standardized School Examination in the future. It is important to have a good quality of proper test to
measure the ability of students. In addition, Nationally Standardized School Examination as one of students’ requirements to graduate needs to be proper so that it measures what needs to be measured. A set of test items consists of most items are classified as hard for its item difficulty and slightly reliable for its reliability does not have its ability to describe the ability of students.

The results showed that mean test score and standard deviation of this set of items was different for two different types of questions. Mean test score for multiple choice questions type was 6.52 with its standard deviation 2.71. However, mean test score for short answer questions was 1.10 with its standard deviation 1.14. It indicates that mean score for both types of item is classified as poor. It shows that the achievement of students is low. Evaluation is therefore necessary done by teachers to improve the learning process.

Evaluation is needed in order to improve student Nationally Standardized School Examination results. Firstly, needs to be done by the teacher. According to the result, that item difficulty is about 85.3% of the non-compulsory mathematics item test used in Nationally Standardized School Examination is in difficult category. It indicates that in general students did not understand the materials or topics being tested. To improve students' understanding and outcomes, teachers need to change the way of teaching [20]. Mastery of mathematical concepts is needed so that students are able to solve math problems. The learning process is not only focused on the use of formulas and procedures in finding the correct answer, but it requires students understanding of the basic concepts. Teaching that only emphasizes procedural does not equip students with the mathematical abilities that students need in the future [21]. Consequently, students are not able to work on mathematical problems in the future.

Teacher must improve the learning process and put more emphasis on students’ concepts understanding. This can be started by identifying materials that are not yet mastered by students. Then, teacher can use a learning trajectory to teach concept and give intensive enrichment if needed. In addition, in teaching difficult mathematical problems, teachers are advised to provide a sequence of logical steps and concepts needed to solve the problem. By providing various problems as exercises, students will find ideas on how to use the concepts they have in solving a mathematical problem.

Secondly, the evaluation needs to be done by the items' developers. The non-compulsory mathematics item tests used in Nationally Standardized School Examination were developed by subject teacher deliberations especially mathematics teacher deliberations at regency level. Different teachers probably will conduct different standard or level of item test. It depends on teachers' experience about the ability of their students. Thus it probably can be bias with standards on the other schools. The solution is to develop item tests through a group discussion forum. Through this discussion group forum, all teachers have opportunity to describe the abilities of their students. In addition, the teacher can also review items that are considered too difficult or even too easy. Thus the item tests were developed to have the same standard in every school.

5. Conclusion
Based on the analysis of items validity indicated that a set item test of non-compulsory mathematics on Nationally Standardized School Examination was strongly relevant to the indicator in assessing students’ learning achievement. It was a set of proper test to measure students’ ability.

The result of items analysis by applying Classical Test Theory framework and Quest application showed that this set item test was slightly reliable. There were 29 from 34 items were classified as hard for its item difficulty and 5 items only were classified as moderate. The higher percentage founded to be 85.3% item were classified as hard. Most of items test were therefore hard to be answered correctly by students. There were 12 from 34 items were classified as good for its item discrimination and 22 items were classified as poor. The lower percentage founded to be 35.3% item were classified as good for its item discrimination indicated that this set of test item was not accurate in selecting between good and poor students. However, the distractor of multiple choice items works effectively.
Lower average score indicated that students have not tried their best on doing the test they take. This was mainly affected by poor understanding of students. Not only teachers are highly suggested to evaluate their teaching practices but also students need to be more prepared to improve the learning process. Further research is therefore important to be conducted to determine students’ difficulties in answering test items.

6. References
[1] Retnawati H, Kartowagiran B, Arlinwibowo J, and Sulistyaningsih E 2017 Int. J. Instr. 10 pp 257–276.
[2] Santoso A 2018 J. Pendidik. Mat. dan Sains VI pp 158–168.
[3] Popham W J 2009 Instruction that measures up (Virginia: ASCD).
[4] Kemendikbud 2018 Permen Pendidik No. 4 Tahun 2018 tentang penilaian hasil belajar oleh satuan pendidikan dan penilaian hasil belajar oleh pemerintah (Jakarta: Kemendikbud) p 4.
[5] BSNP 2018 Tanya jawab pelaksanaan Ujian Sekolah Berstandar Nasional (USBN) tahun pelajaran 2017/2018 (Jakarta: BSNP) p 4.
[6] Kartowagiran B, Munadi S, Retnawati H, and Apino E 2018 SHS Web Conf. 42 pp 1–6.
[7] Crocker L and Alglna J 2008 Introduction to classical and modern test theory (Ohio: CENGAGE Learning).
[8] Retnawati H 2016 Analisis kuantitatif instrumen penelitian (Yogyakarta: Parama Publishing).
[9] Reynolds C R, Livingston R B, and Willson V 2010 Measurement and assessment in education (New Jersey: Pearson Education Inc).
[10] Haladyna T M and Downing S M 1993 How many options is enough for a multiple choice test item? Educ. Psychol. Meas. 53 pp 999–1010.
[11] Kemendikbud 2017 Rekap hasil Ujian Nasional 2017 (Jakarta: Puspendik).
[12] Kemendikbud 2018 Rekap hasil Ujian Nasional 2018 (Jakarta: Puspendik).
[13] Retnawati H, Arlinwibowo J, and Sulistyaningsih E 2017 The students’ difficulties in completing geometry items of national examination Int. J. New Trends Educ. Their Implic. 8 pp 1309–6249.
[14] Wasida M R and Hartono H 2018 Analisis kesulitan menyelesaikan soal model ujian nasional matematika dan self-efficacy siswa SMA J. Ris. Pendidik. Mat. 5 pp 82–95.
[15] Tias A A W and Wutsqa D U 2015 Analisis kesulitan siswa sma dalam pemecahan masalah matematika kelas XII IPA di kota Yogyakarta 2015 J. Ris. Pendidik. Mat. 2 pp 28–39.
[16] Rumasoreng M I and Sugiman S 2014 Analisis kesulitan matematika siswa SMA/MA dalam menyelesaikan soal setara UN di Kabupaten Maluku Tengah J. Ris. Pendidik. Mat. 1 pp 22–34.
[17] Jamal F 2014 Analisis kesulitan belajar siswa pada materi peluang kelas XI IPA SMA Muhammadiyah Melulahboh Johan Pahlawan J. MAJU 1 pp 18–36.
[18] Akbar G A M, Diniyah A N, Akbar P, Nurjaman A, and Bernard M 2018 Analisis kemampuan penalaran dan self confidence siswa SMA dalam materi peluang J. Educ. 1 pp 14–21.
[19] Ayal C S, Kusuma Y S, Sabandar J, and Dahlan J A 2016 J. Educ. Pract., 7 pp 50–58.
[20] Brown G T L 2018 Assessment of student achievement (New York: Routledge).
[21] Sarwadi H R H and Shahrill M 2014 Math. Educ. Trends Res. 2014 pp 1–10.