ITEM ANALYSIS OF HIGH SCHOOL SPECIALIZATION MATHEMATICS EXAM QUESTIONS WITH ITEM RESPONSE THEORY APPROACH

Lovieanta Arriza 1*, Heri Retnawati 2, Rizki Tika Ayuni 3

1,2,3 Educational Research and Evaluation Program, Graduate School, Universitas Negeri Yogyakarta Karangmalang Campus, Colombo Street No.1, Yogyakarta, 55281, Indonesia

Corresponding author’s e-mail: * lovieanta0341pasca.2022@student.uny.ac.id

ABSTRACT

Analysis of item characteristics on test instruments is carried out to determine high-quality items. This study aims to describe the parameters of specialized high school mathematics test items using the IRT approach. It is an exploratory, descriptive study employing a quantitative approach. The research subjects were 36 students of grade XI high school who took the specialization mathematics subject. Response data with dichotomous scoring were analyzed using the IRT approach with the R program to obtain information about item parameters and student ability. The results of the model fit test showed that most of the specialization mathematics exam items fit the Rasch model. The results showed that all items met the criteria of good quality because they had good difficulty parameters. Relatively, the test items were suitable for students with abilities between -2.6 and 2.8 logits. This estimation is also supported by the TIF with a maximum value of 3.049 at 0.08 logit ability and SEM of 0.541. Test items that have been proven to be of high quality can be used as examples in both teaching and diagnostic assessments. Further research could consider the discrimination parameter when analyzing the characteristics of the questions.

Keywords:
Item Analysis;
Item Response Theory;
Rasch Model;
Dichotomus.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.
1. INTRODUCTION

One of the main and integral components of the learning process is assessment [1], [2]. Improving the quality of education can be achieved through improving the quality of learning and the quality of the assessment system. The assessment process will provide information on the learning outcomes that have been carried out [3], [4]. In addition, assessment also provides feedback on student learning progress [5], [6]. In educational assessment, tools are needed in the form of assessment instruments [7]. There are two types of assessment instruments, namely tests to measure student learning achievement, intelligence, aptitude, skills, and nontests. Most instruments contain one or a set of items that aim to measure the knowledge and skill domain or task domain that represents the student’s ability [8].

The form of test instrument commonly used in educational assessment and measurement is multiple choice. A multiple-choice test is a set of tests that contain questions with two or more alternative answers, but only one answer is correct, and the test takers have to choose one correct answer [9]. Multiple-choice tests are widely used because of their ease and effectiveness in measuring students’ knowledge or skill domains [10], [11], [12]. Multiple-choice questions also allow teachers to quickly measure a wide range of knowledge, skills, and competencies across disciplines and fields including student’s ability to understand concepts and principles, make judgments, draw conclusions, give reasons, complete statements, interpret data, and apply information [13], [14].

The test instruments that have been used still need to be tested empirically for quality. Data from the test results will be analyzed to obtain evidence about the characteristics of the items concerned. Then from the results of the analysis of the characteristics of these items will be obtained the basis for making the necessary revisions. The quality of the instrument requires validity, reliability, and good item parameters [15]. Conclusions about the characteristics of the items on a test will lead to decisions about whether or not the item should be used, whether it should be discarded, whether it can still be revised, or whether it has indeed met the requirements of a good item. The systematic work procedure for evaluating all test items based on empirical data is called item analysis [16]. Item analysis is a key step in evaluating tests in the field of educational measurement [17].

In educational measurement, there are two types of approaches that can be used to analyze instrument quality, namely classical test theory and modern theory [18]. Although it is called classical test theory, it is still used today. The purpose of both test theories is the same, which is to find the most appropriate way to obtain a pure score that reflects the actual ability of the participants. However, classical test theory is considered to have weaknesses. The main weakness of classical test theory is that the characteristics of items and examinees cannot be separated. Item parameters are highly dependent on the ability of participants, and vice versa [19], [20]. In other words, the analysis approach with classical test theory cannot be used because the assessment results are highly dependent on the group of test takers.

Item response theory (IRT) has emerged as a solution to current analytical approaches in educational measurement [21], [22]. Item response theory is a modern psychometric framework that provides several beneficial traits compared to classical test theory [23]. Using IRT, the difficulty, discrimination and efficiency of the item information function can be evaluated [24]. In IRT, the probability of a subject answering an item correctly depends on the subjects’ ability and item characteristics [25]. This means that test takers with high ability will have a greater probability of answering correctly when compared to participants who have low ability [22], [26].

The application of IRT in analyzing the quality of instrument items has been widely published. As revealed in Malaspina and Arias [27], which describes the characteristics of items in an early mathematics test instrument using the Item Response Theory (IRT) approach. The results indicate that the item characteristics align with the student sample and do not contain bias. Another study conducted by Ramadhani et al. [28] analyzed items related to statistical reasoning in the context of ethnomathematics for junior high school students. This finding further strengthens the notion that IRT analysis provides consistent information about item characteristics that remains unchanged regardless of students’ varying abilities. In addition to these two studies, there is also another research examining the characteristics of mathematics items, specifically focusing on the topic of rational numbers. Khairani and Shamsuddin [29] conducted this study, concluding that the formulated items have good quality and serve as a guide for teachers to use high-quality items in classroom assessments.
The specialization mathematics subject taught in high schools today is sufficiently intriguing to be chosen as the object of analysis. Based on previous research, no research has examined the quality of specialization mathematics exam questions used in schools. Specialization mathematics is a subject taken by students with a particular interest in mathematics, especially those who intend to continue their studies in mathematics-related fields, thereby increasing their chances of working in mathematics-related fields. This underscores the importance of analyzing the characteristics of questions that measure students’ abilities in specialization mathematics in secondary schools.

Based on the discussions with the high school mathematics teacher, the information obtained is that so far, no analysis of test items has been carried out by the teacher, either by applying classical test theory or IRT. The test questions developed by teachers are directly used in the assessment process without determining the characteristics of each item. This also applies to specialization mathematics exam questions. In fact, it is important for teachers to know the characteristics of test items so that they are in accordance with the measurement objectives [30].

Based on preliminary studies and relevant research publications that have been stated previously, the application of IRT in analyzing item parameters will be more profitable than classical test theory, both in determining the characteristics of test items and estimating student abilities. Thus, this research aims to describe the characteristics of specialization mathematics exam items using the IRT approach developed by teachers. This study hopes to determine the quality of mathematics specialization exam questions used in schools. In addition to being used for final assessment, high-quality exam questions can be a reference in preparing other questions and are beneficial for teachers in providing examples of problems in classroom learning. This is very important considering that specialization mathematics is a subject attended by students who have a particular interest in exploring mathematics.

2. RESEARCH METHODS

This research is exploratory, descriptive research with a quantitative approach that aims to describe the characteristics of specialization mathematics exam questions developed by teachers. The research subjects were 36 class XI high school students consisting of 10 male students and 26 female students. The sample was chosen because they were students taking specialization mathematics subjects. The test instrument used is specialization mathematics questions consisting of 15 multiple choice (dichotomy) questions with five response categories. Based on the test instrument, student answer results are obtained and collected through the documentation method. The instrument validation aspects analyzed include: a) prerequisite tests for the IRT model, namely unidimensionality and local independence tests, b) model fit testing, c) estimation of item difficulty levels, and d) measurement information function.

Before the data is analyzed with the IRT approach, there are two assumptions that should be fulfilled. Ahmad and Mokshein stated that the assumptions of unidimensionality and local independence should be tested before conducting IRT-based item analysis [31]. Conducting a unidimensionality test in IRT is very important to prove that the question instrument only measures one dimension of ability [32], [33]. The unidimensionality assumption test can be proven through factor analysis by reviewing the eigenvalues in the inter-item covariance variance matrix or can be seen in the eigenvalue scree plot, which shows one dominant component [34], [35]. The unidimensionality test in this case is proven through factor analysis using the R program. Factor analysis is carried out by first conducting a feasibility test analysis, namely the KMO-MSA test (Kaiser-Meyer-Olkin measure of sampling adequacy) and Bartlett’s test. The KMO-MSA test aims to see the adequacy of the sample, while Bartlett’s Test serves to prove the homogeneity of the data. If the (KMO)-MSA value is > 0.5 and Bartlett’s significant test is < 0.05, the unidimensionality test can continue [36].

The unidimensional assumption is proven through principal component analysis (PCA) by counting eigenvalue ($\lambda$m) of covariance matrix using Equation (1) as follows [37]:

$$C_x v_m = \lambda_m v_m$$  \hspace{1cm} (1)

The next assumption test is local independence. According to Hambleton et al [19], local independence is mathematically expressed by the Equation (2):

$$P(u_1, u_2, ..., u_n|\theta) = P(u_1|\theta)P(u_2|\theta) ... P(u_n|\theta)$$
\[ = \prod_{i=1}^{n} P(u_i|\theta) \]  \hspace{2cm} \hspace{2cm} (2)

Description:
- \(i\) : 1, 2, 3, ... n
- \(n\) : Number of test item
- \(P(u_i|\theta)\) : The probability of test takers who have the ability \(\theta\) can answer item number-\(i\) correctly
- \(P(u_1, u_2, ..., u_n|\theta)\) : The probability of test takers who have the ability \(\theta\) can answer item number-\(i\) to number-\(n\) correctly

After the IRT assumptions are met, the next analysis is to test model suitability. *Chi-square* \((\chi^2)\) method is one of statistical test of model suitability which used to test latent \([38]\). Equation (3) below is *Chi-square* \((\chi^2)\).

\[ X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \]  \hspace{2cm} \hspace{2cm} (3)

Description:
- \(X^2\) : *Chi-square* distribution
- \(O_i\) : Observation value number-\(i\)
- \(E_i\) : Expectation value number-\(i\)

Item parameter estimation is carried out after determining a suitable model. The formula used in this analysis refers to the Rasch model which contains item difficulty level parameters. Below is Equation (4) for estimating item parameters with the Rasch model \([39]\).

\[ P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}} \]  \hspace{2cm} \hspace{2cm} (4)

Description:
- \(P_i(\theta)\) : The probability of test takers who have the ability \(\theta\) are randomly selected can answer item order-\(i\) correctly
- \(\theta\) : Ability level
- \(b_i\) : Difficulty level item number-\(i\)
- \(e\) : Natural numbers that are close in value 2,718

The results of estimating item parameters using the Rasch model above are then strengthened by the Item Information Function (IIF). Mathematically, the item information function satisfies Equation (5) as follows \([26]\):

\[ I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \]  \hspace{2cm} \hspace{2cm} (5)

Description:
- \(i\) : 1, 2, 3, ..., n
- \(I_i(\theta)\) : Information function number-\(i\)
- \(P_i(\theta)\) : The probability that a participant with ability \(\theta\) will answer item number-\(i\) correctly
- \(P_i'(\theta)\) : Derivative function \(P_i(\theta)\) to \(\theta\)
- \(Q_i(\theta)\) : The probability that a participant with ability \(\theta\) will answer item number-\(i\) incorrectly

The criteria for a valid instrument based on several aspects can be seen in Table 1.
Table 1. Valid Test Criteria are Seen from Various Aspects and Criteria [40]

<table>
<thead>
<tr>
<th>The validity aspect of the item</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unidimensional test</td>
<td>There is only one dominant factor in the specialization mathematics exam questions through scree plot factor analysis</td>
</tr>
<tr>
<td>Local independence</td>
<td>Can be known by proving the unidimensional assumption</td>
</tr>
</tbody>
</table>
| Model fit testing              | $\chi^2 > \chi^2_{tabel}$  
                           | $\text{sig} > \alpha (0.05)$ |
| Index difficulty item          | Good: $-2 \leq b_i \leq 2$  
                           | Not good: $b_i > 2$  :  $b_i < -2$ |
| Person ability (wright map)    | All levels of difficulty of the questions are within the student's ability domain |
| Test Information Function      | The information function test has the maximum value in the student's ability domain |

3. RESULTS AND DISCUSSION

Based on the preliminary analysis conducted, the KMO-SMA and Barlett's values were obtained as a measure of sample adequacy, with the output presented in Figure 1 below.

![Figure 1. KMO and Bartlett's test](image)

Based on Figure 1, it could be seen that the KMO-MSA value was 0.508 and the significant Bartlett's test was 0.000. This indicates that the 36 samples used for IRT analysis have met the requirements of sample adequacy and the data are homogeneous so that factor analysis can be carried out. The unidimensionality assumption test can be proven through factor analysis by reviewing the eigenvalue in the inter-item covariance variance matrix or can be seen in the eigenvalue scree plot which shows one dominant component. The results of data processing for principal component analysis with the R program can be seen in the eigenvalue section in Figure 2 below.

![Figure 2. Test Results of The Main Components](image)

Based on Figure 2, it can be found that the response data contains 5 components. From the five components, component 1 was the most dominant component because it has the largest eigenvalue of 4.615. The difference between the eigenvalues of components 1 and 2 is 2.7669341. Meanwhile, the difference between components 2 and 3 is 0.1855432. The eigenvalue differences for the subsequent components remain less than the differences between the eigenvalues of components 1 and 2. This further indicates a significant decrease in the eigenvalue differences, especially between component 1 and the other components. This factor value can then be presented in the scree plot in Figure 3.
Figure 3 above is the scree plot result of the factor analysis. The dots in the figure show the number of dimensional factors in the math specialization exam questions. The connecting lines between the points show the distance and slope of the differences between factors. The scree plot of the factor analysis shows a very sharp drop between factor 1 and factor 2, and the eigenvalues then start to skew at factor 3 so that the scree plot almost forms a right angle. This indicates that there was only one dominant factor in the math specialization exam questions. The results of this analysis are in accordance with the assumptions in the IRT approach where a set of questions or tests only measure one latent trait so that the assumption of unidimensionality is met [19].

The next assumption test is local independence. In its proof, this local independence can be fulfilled if the participant’s answer to an item does not affect the participant's answer to another item [41], [22]. Local independence can be known by proving the unidimensionality assumption [19]. In this analysis, it can be seen in Figure 3 that the assumption of unidimensionality has been met so that the local independence test has also been met.

Furthermore, the model fit was assessed using various indicators. One of them was to compare the output of several different model fit indices, Akaike information criterion (AIC), Bayesian information criterion (BIC), loglikelihood which can be seen in Table 2.

### Table 2. Comparison of Model Fit Tests

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>loglikelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rasch</td>
<td>666.99</td>
<td>692.33.00</td>
<td>-317.50</td>
</tr>
<tr>
<td>2PLM</td>
<td>675.71</td>
<td>723.21.00</td>
<td>-307.85</td>
</tr>
<tr>
<td>3PLM</td>
<td>698.95</td>
<td>770.21.00</td>
<td>-304.48</td>
</tr>
</tbody>
</table>

Table 2 provides information on the values obtained from the calculation of AIC, BIC, and loglikelihood for each model. In order to determine which model best fits the test data, the selection criteria are seen from the smallest value in each model. The smaller the fit index value, the better the model fits the data [42]. Table 3 shows that the Rasch model provides the smallest value among other models in each AIC, BIC, and loglikelihood.

Model fit testing can also be done using statistical methods by determining the chi-square of each item on each IRT logistic parameters. This was done by calculating the chi-square ($\chi^2$) value and then comparing it with the chi-square ($\chi^2$) value from the table, or by reviewing the probability (significance) value. Items are said to fit the model if the calculated chi-square value was smaller than the chi-square ($\chi^2$) table or sig value > $\alpha$ (0.05) [43]. The results of the fit test for each item in the Rasch, 2 PL and 3 PL models were presented in Table 3.
Based on Table 3, it can be seen that all items were suitable for the Rasch model, 13 items were suitable for the 2PL model and the suitability for the 3PL model was also 13 items. Based on the percentage, the suitability with the Rasch model is the greatest compared to 2PL and 3PL. Therefore, it can be concluded based on this analysis that the analysis of the instrument of specialization mathematics exam questions was suitable for the Rasch model.

Rasch is a measurement model developed by mathematician George Rasch from Denmark. Analysis with the Rasch model is determined based on the level of item difficulty and participant ability simultaneously. The probability of answering an item correctly is differentiated based on item difficulty and individual ability. People with low ability should not be able to correctly answer items that have a high level of difficulty [44]. In this case, the scoring used to be analyzed with the Rasch model was dichotomous data on the results of the specialization mathematics exam, namely the wrong answer was given a score of 0 and the correct answer was given a score of 1.

The next analysis was the estimation of item parameter values with reference to the Rasch model, namely the item difficulty parameter. In the analysis of item characteristics using the Rasch model, an item is said to have good quality if it has a difficulty index (b) ranging from -2 ≤ b ≤ 2 [19]. The overall results of item parameter estimation using the Rasch model were presented in Table 4.

<table>
<thead>
<tr>
<th>Item</th>
<th>Rasch Model</th>
<th>2PL</th>
<th>3PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_1</td>
<td>0.149</td>
<td>0.187</td>
<td>0.700</td>
</tr>
<tr>
<td>B_2</td>
<td>0.717</td>
<td>0.375</td>
<td>0.219</td>
</tr>
<tr>
<td>B_3</td>
<td>0.369</td>
<td>0.379</td>
<td>0.018</td>
</tr>
<tr>
<td>B_4</td>
<td>0.497</td>
<td>0.490</td>
<td>0.215</td>
</tr>
<tr>
<td>B_5</td>
<td>0.513</td>
<td>0.880</td>
<td>0.467</td>
</tr>
<tr>
<td>B_6</td>
<td>0.512</td>
<td>0.437</td>
<td>0.677</td>
</tr>
<tr>
<td>B_7</td>
<td>0.418</td>
<td>0.645</td>
<td>0.697</td>
</tr>
<tr>
<td>B_8</td>
<td>0.262</td>
<td>0.390</td>
<td>0.073</td>
</tr>
<tr>
<td>B_9</td>
<td>0.417</td>
<td>0.494</td>
<td>0.366</td>
</tr>
<tr>
<td>B_10</td>
<td>0.282</td>
<td>0.462</td>
<td>0.427</td>
</tr>
<tr>
<td>B_11</td>
<td>0.056</td>
<td>0.046</td>
<td>Not Fit</td>
</tr>
<tr>
<td>B_12</td>
<td>0.116</td>
<td>NaN</td>
<td>NaN</td>
</tr>
<tr>
<td>B_13</td>
<td>0.674</td>
<td>0.706</td>
<td>0.696</td>
</tr>
<tr>
<td>B_14</td>
<td>0.212</td>
<td>0.120</td>
<td>0.037</td>
</tr>
<tr>
<td>B_15</td>
<td>0.188</td>
<td>0.208</td>
<td>0.126</td>
</tr>
</tbody>
</table>

Table 4. Results of Item Parameter Estimation with The Rasch Model

<table>
<thead>
<tr>
<th>Item</th>
<th>Index Difficulty</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.159</td>
<td>Good</td>
</tr>
<tr>
<td>2</td>
<td>-0.137</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>0.308</td>
<td>Good</td>
</tr>
<tr>
<td>4</td>
<td>0.611</td>
<td>Good</td>
</tr>
<tr>
<td>5</td>
<td>0.159</td>
<td>Good</td>
</tr>
<tr>
<td>6</td>
<td>0.308</td>
<td>Good</td>
</tr>
<tr>
<td>7</td>
<td>-0.438</td>
<td>Good</td>
</tr>
<tr>
<td>8</td>
<td>-0.137</td>
<td>Good</td>
</tr>
<tr>
<td>9</td>
<td>-1.253</td>
<td>Good</td>
</tr>
<tr>
<td>10</td>
<td>1.445</td>
<td>Good</td>
</tr>
<tr>
<td>11</td>
<td>-0.591</td>
<td>Good</td>
</tr>
<tr>
<td>12</td>
<td>-0.438</td>
<td>Good</td>
</tr>
</tbody>
</table>

Based on Table 3, it can be seen that all items were suitable for the Rasch model, 13 items were suitable for the 2PL model and the suitability for the 3PL model was also 13 items. Based on the percentage, the suitability with the Rasch model is the greatest compared to 2PL and 3PL. Therefore, it can be concluded based on this analysis that the analysis of the instrument of specialization mathematics exam questions was suitable for the Rasch model.

Rasch is a measurement model developed by mathematician George Rasch from Denmark. Analysis with the Rasch model is determined based on the level of item difficulty and participant ability simultaneously. The probability of answering an item correctly is differentiated based on item difficulty and individual ability. People with low ability should not be able to correctly answer items that have a high level of difficulty [44]. In this case, the scoring used to be analyzed with the Rasch model was dichotomous data on the results of the specialization mathematics exam, namely the wrong answer was given a score of 0 and the correct answer was given a score of 1.

The next analysis was the estimation of item parameter values with reference to the Rasch model, namely the item difficulty parameter. In the analysis of item characteristics using the Rasch model, an item is said to have good quality if it has a difficulty index (b) ranging from -2 ≤ b ≤ 2 [19]. The overall results of item parameter estimation using the Rasch model were presented in Table 4.

Table 4. Results of Item Parameter Estimation with The Rasch Model

<table>
<thead>
<tr>
<th>Item</th>
<th>Index Difficulty</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.159</td>
<td>Good</td>
</tr>
<tr>
<td>2</td>
<td>-0.137</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>0.308</td>
<td>Good</td>
</tr>
<tr>
<td>4</td>
<td>0.611</td>
<td>Good</td>
</tr>
<tr>
<td>5</td>
<td>0.159</td>
<td>Good</td>
</tr>
<tr>
<td>6</td>
<td>0.308</td>
<td>Good</td>
</tr>
<tr>
<td>7</td>
<td>-0.438</td>
<td>Good</td>
</tr>
<tr>
<td>8</td>
<td>-0.137</td>
<td>Good</td>
</tr>
<tr>
<td>9</td>
<td>-1.253</td>
<td>Good</td>
</tr>
<tr>
<td>10</td>
<td>1.445</td>
<td>Good</td>
</tr>
<tr>
<td>11</td>
<td>-0.591</td>
<td>Good</td>
</tr>
<tr>
<td>12</td>
<td>-0.438</td>
<td>Good</td>
</tr>
</tbody>
</table>
Based on Table 4, it was obtained that all items of the specialization mathematics exam with dichotomous scoring had good item difficulty parameters. This information proved that 100% of all test items were able to describe the function of students’ abilities. Where students who have high abilities will find it easy to work on the items, conversely students who have low abilities will find it difficult to answer the items.

The next analysis showed information that the ability of students who answered the specialization math exam items was in theta (θ) -1.955 to 2.093 logit with an average theta of 0.00. Based on this data, it could be concluded that it was natural that many items were classified as moderate for students because their abilities were relatively moderate or average abilities. The estimation of participants’ abilities can also be seen in Figure 4.

![Student ability estimation graph](image1)

Figure 4. Student ability estimation graph

The subsequent analysis of the information function yields values that are inversely related to the Standard Error of Measurement (SEM), as presented in Figure 5.

![Information Function Graph of Test and SEM](image2)

Figure 5. Information Function Graph of Test and SEM

Figure 5 above is the result of the Test Information Function (TIF) and SEM analysis. The blue colored curve line is the TIF graph that has a high point of 3.049 at theta (θ) 0.08 logit. Meanwhile, the red line which is inversely proportional to the TIF is the SEM graph with the lowest point of 0.541. This meant that the specialization mathematics exam items produced optimal information when used by students with 0.08 logit ability. The TIF and SEM curves intersected at theta -2.6 and 2.8 which means that the test was overall suitable for students with abilities between -2.6 and 2.8 logits. This range also indicated that the specialization math exam questions were able to measure the ability of students with a fairly wide range [45]. This is an
important note considering that the abilities of students in all schools are not at the same level. Therefore, test questions prepared by teachers must be able to facilitate various levels of student ability [46], [47].

4. CONCLUSIONS

Based on the results of the study, it can be concluded that the test sample data has met the IRT assumptions. The results of the model fit test analysis show that all items of the specialization mathematics exam are suitable for estimation with the Rasch model compared to the 2PL and 3PL models. From the results of estimating item parameters with the Rasch model, all items meet the criteria for a good difficulty index because the dominant ability of the participants measured is classified as moderate. The ability of students who answered the math specialization exam items was in the theta (θ) -1.955 to 2.093 logit with an average theta of 0.00. The results of estimating item parameters and student ability are supported by the Test Information Function (TIF) with a maximum value of 3.049 at 0.08 logit ability. This is reinforced by the SEM curve, which is inversely proportional to the TIF with the lowest point of 0.541. This provides information that the specialization math exam items produce optimal information when used by students with a logit ability of 0.08.

REFERENCES


