

## An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school

\*<sup>1</sup>Mutiara Kusumawati; <sup>2</sup>Samsul Hadi

<sup>1</sup>Graduate School of Universitas Negeri Yogyakarta  
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

<sup>2</sup>Faculty of Engineering, Universitas Negeri Yogyakarta  
Karangmalang, Depok, Sleman 55281, Yogyakarta, Indonesia

\*Corresponding Author. E-mail: mutia21@gmail.com

*Submitted: 04 July 2018 | Revised: 31 October 2018 | Accepted: 06 November 2018*

### Abstract

The multiple-choice test is a common test format used in education. One of the purposes of this test is to evaluate the success of the learning process in a particular subject. Therefore, the efficiency of the evaluation depends on the quality of the test items used. This research was conducted in order to reveal the quality of the final mathematics examination items statistically. It was descriptive quantitative research employing two-parameter logistic (2pl) model of Item Response Theory (IRT). The data were obtained from the sample of 353 students established using the purposive sampling technique. This finding shows that 40% of the 35 items tested are very difficult, 60% are in the medium level, and there is no easy item. The most difficult material is the trigonometric calculation. The percentage of the item discrimination index is described as follows: 8.57% of the items are categorized as very low, 51.43% are categorized as low, 31.43% of the items have a medium item discrimination index value, 5.71% have a high item discrimination index value, and 2.86% of the items are categorized as very high. Moreover, the research found that all distractors functioned well. The highest information on ability  $\theta = 0.4$  with information function value of 5.38 and SEM = 0.6. This test is suitable for students with the ability of  $-1.42 < \theta < 2.65$ .

**Keywords:** *difficulty index, discrimination index, efficiency distractor, mathematics examination*

### Introduction

Evaluation refers to a systematic process to determine which instructional goals are achieved by students (Gronlund, 1982, pp. 5–6). The accomplishment of instructional objectives is done by a measurement. By measuring and evaluating, teachers can diagnose the strengths and weaknesses of their students and take action for their progress and improvement. If it is effective, measurement and evaluation can improve the learning situation. Without evaluation and measurement, it is impossible to know the needs and abilities of the students (Tshabalala & Ncube, 2014, p. 141).

Thus, the final examination of mathematics subject for class X is conducted in order to provide general information and illus-

tration of student learning outcomes in the last semester. This is as a consideration of the teachers in particular and the schools in general that determines whether the students should keep on learning in the next grade or not. In addition, the results of these tests were used as an evaluation for educators in the implementation of the learning process. Therefore, the test items of mathematics final examination are one of the most important instruments in the learning process and must be well structured. A good final test item will give a good measurement result (Mardapi, 2012, p. 27). According to information from the drafting team of the Board of Muhammadiyah-School Principals Cooperation (or *Badan Kerjasama Kepala Sekolah Muhammadiyah*), gained through interviews during the survey,

the materials of the mathematics subject test items for the 10<sup>th</sup> grade students used so far have not gone through a good stage in the preparation. Sulistiawan (2016, p. 10) finds that in the final examination of mathematics subject matter, there are 10% invalid items. This analysis was conducted to give some suggestions to the test developers of the mathematics final examination for the tenth grade students.

The test items are presented in the objective test form. The objective test is defined as a structured test that asks the participants to fill in one or two words, or to choose the correct answer from several options. An objective test consists of the problem/test item and list of alternative solutions. A list of alternative solutions can be in the form of words, numbers, symbols or phrases, and called key answers. Participants of the test are usually asked to read the problem/test items and alternative solutions list, and choose one right or best alternative (Gronlund, 1982, p. 135). The right option to each item is called an answer key, while the other options are called distractors. On the other hand, an essay test provides an opportunity for the test participants to organize, arrange, or answer freely from the questions given. For some instructional objectives, objective tests are considered to be more efficient in order to measure learners' skills at both low and high levels (Gronlund, 1982, pp. 5–6).

The selection of the appropriate test form is determined by the purpose of the test, the number of test participants, the range of test materials, and the characteristics of the subjects tested. The multiple-choice test and True or False test are particularly appropriate when the number of the test takers is large, the time for test correction is short, and the coverage of the tested material is numerous. The advantage of an objective multiple-choice test is that an answer sheet can be checked using the computer, so scoring objectivity can be assured (Saudi Commission for Health Specialties, 2011, p. 68).

## Method

This research is a descriptive quantitative study conducted at two Muhammadiyah

high schools in Yogyakarta, Indonesia. A population sampling technique in which the entire population was used as the sample (Sugiyono, 2001, p. 61) was used in this study. The participants of the study were 353 grade X students. The data of this study were the students' responses to the final examination consisting of 35 items of multiple-choice questions which have five options for each item. The data were analyzed quantitatively using Bilog.

The quality of the items on the 10<sup>th</sup> grade students' mathematics final examination was analyzed using modern Item Response Theory (IRT). This is a theory which employs the mathematical function to connect the opportunities of the correct answers to the students' ability. The IRT has a mathematical formula that connects the participants' characteristics and the item features in the model (Hambleton, Swaminathan, & Rogers, 1991, p. 12). The advantages of IRT include: the item statistics is not dependent on the group; the test scores obtained can illustrate individual capabilities; it does not require parallel tests to calculate the reliability coefficients; and it can provide the right measurement for each ability score.

There are three logistic models in the IRT, namely one-parameter logistic model (1pl), two-parameter logistic model (2pl), and three-parameter logistic model (3pl). These three models are suitable to respond to dichotomous forms (Hambleton et al., 1991, pp. 12–17). The three models are distinguished by the number of parameters which are used to describe the item characteristics of each logistic model or item parameters. The item parameters are item difficulty index (b), item discrimination index (a), and pseudo guessing (c). These three elements are so interrelated that it causes a function or response curve which is called the Item Characteristic Curve (ICC).

IRT can provide good results if the data used were in accordance with the selected logistic model. The selection of the logistic model is determined based on the p-value, which means that if the p-value is more than 0.05, then the item is said to fit the model (Retnawati, 2014, p. 25). The chosen logistic

model is determined from the logistic model that produces the most suitable items. The 2pl model produces the most suitable items so that this study employs the 2pl model. The 2pl model formula is as follows.

$$P_i(\theta) = \frac{e^{D_i a_i (\theta - b_i)}}{1 + e^{D_i a_i (\theta - b_i)}} \dots\dots\dots i=1,2,3,\dots,n$$

Notes:

- $P_i(\theta)$ : the chance that test participants can answer test item  $i$  correctly
- $a_i$  : item  $i$  discrimination index
- $b_i$  : item difficulty index, a point of ability where the possibility to answer correctly is 0.5.
- $\theta$  : parameter of ability of test participants

The difficulty index is an opportunity to answer each item correctly at a certain level of ability. The percentage of the difficulty level used is elaborated as follows: an item of problem with a high difficulty level of 20%, 60% of items with medium difficulty level, and another 20% are items with low difficulty level (Arikunto, 1999, p. 210; Gajjar, Sharma, Kumar, & Rana, 2014, p. 19). The good index of difficulty levels is spread from -2.00 to +2.00 (Hambleton et al., 1991, p. 13). The closer the  $b$ , the easier the item is. The more the value of  $b$  approaching +2.00, the more difficult the item is. A good item is an item that is not too difficult or too easy. The overly easy question does not stimulate the students to increase the effort to solve the problem. Conversely, too difficult items will discourage the students from trying again because they are out of range (Daryanto, 2012, p. 197; Miller, Linn, & Gronlund, 2009, p. 21). In preparing the test item, the percentage of item difficulty level needs to be considered.

Discrimination Index (DI) is the effectiveness of an item measurement to distinguish learners with high ability from those with low ability. The discrimination index spreading from 0.01 to 0.34 is considered to be very low, 0.35 to 0.64 is low, 0.65 to 1.34 is moderate, 1.35 to 1.69 is high, and higher than 1.70 is very high (Baker, 2001, p. 34). The higher the DI, the more effective the item to distinguish learners with high ability from those with low ability.

The spread of alternative options is commonly used as the basis for the study of discrimination index. It is intended to find out whether the option is working or not. An option that is not the correct answer is called a distractor (Allen & Yen, 1979, p. 2). A distractor can be said to work well if it is at least selected by 5% of the test takers (Kolte, 2015, p. 321). If the distractor is selected by less than 5% of the respondents, then it is considered as a non-functioning distractor (NFD). The NFD must then be repaired or deleted, and replaced with another deceptive option (Haladyna & Downing, 1989, p. 55; Tarrant, Ware, & Mohammed, 2009, p. 3). The distractor efficiency is an indicator of whether the distractor on the item has been properly made or whether it has failed to perform its function as a distractor.

The Item Response Theory has several assumptions that need to be confirmed before modeling. These assumptions include: (1) unidimensional data, to show whether the model measures a single construct or not; and also (2) local independence, to show whether the response to each item is influenced by the response to another item (Hambleton et al., 1991, p. 19).

The unidimensionality assumption test of the data was conducted by employing SPSS application. The value of KMO shows 0.513 with the value of sig. = 0.000. Thus, the first assumption has been fulfilled. The local independence assumption is evident when the unidimensionality assumption of the participants' response data has been evident (Retnawati, 2014, p. 7).

## Findings and Discussion

### Findings

The number of multiple choice test items which are analyzed in this study is 35 items in 353 students. The average score achieved is 32.88 and the standard deviation is 13.81. The mean of discrimination index, difficulty level, and distractor efficiency are 0.71, 1.90, and 17.03 respectively. The standard deviations on each parameter are 0.40, 1.56, and 6.63, respectively (see Table 1).

Table 1. Mean and standard deviation of item parameters

Parameter	Mean	Standard Deviation (SD)
Discrimination index (DI)	0.71	0.40
Difficulty index ( <i>p</i> value)	1.90	1.56
Distractor efficiency (DE)	17.03	6.63

The discrimination index in general is low; 51.43% (18 items) is low and 8.57% (3 items) is very low. Only 2.86% (1 item) has a very high discrimination index and 5.71% (2 items) have a high discrimination index value. Moreover, 31.43% (11 items) have a moderate discrimination value.

The distribution of the item difficulty level is generally acceptable. The acceptance threshold of problem difficulty level is between -2.00 to 2.00 (Hambleton et al., 1991, p. 13). A number of 21 items (60.00%) can be accepted with the threshold value (*b*) ranging from -0.43 to 1.94, while the other 14 items (40.00%) have the difficulty level of more than 2.00, meaning that these items are very difficult. The difficult item has a threshold value (*b*) ranging from 2.08 to 4.95.

There are 140 distractors in these tests, in which each item has 4 distractors. All distractors (100%) functioned well because each of them was chosen by more than 5% of the test participants. This means that there is no Non-Functional Distractor (NFD) (Table 2).

## Discussion

A multiple choice test is one of the efficient evaluation tools. However, this effici-

ency is highly dependent on the quality of multiple choices that can be judged on the basis of the item analysis. The index of difficulty and discrimination levels is one of the steps to check whether multiple choice tests are well established or not. Another step used for further analysis is the functionality of the distractors.

### Difficulty Index

It should be noted that the item difficulty level in a test is divided into the following percentages: a problem item with a high difficulty level of 20%, 60% items with medium difficulty level, and another 20% are items with low difficulty level (Arikunto, 1999, p. 210). However, the percentage of difficult, medium, and easy items is not balanced. A total of 21 items (60.00%) can be accepted with the threshold value (*b*) from -0.43 to 1.94. The other 13 items (37.14%) have a difficulty level of more than 2.00, meaning that these items are difficult. Difficult items have a threshold value (*b*) ranging from 2.08 to 4.95. In this case, there is no easy item. The percentage distribution of the difficulty level of the item is illustrated in Figure 1. Thus, this percentage does not reflect a good distribution of test item difficulty level. A good test item must have a balanced percentage of test items with high level of difficulty and low level of difficulty, at 20% each. Of 35 items, 14 items are categorized as having a high level of difficulty and there is no item with low level of difficulty.

Table 2. Difficulty index interpretation, discrimination index, and distractor efficiency

Parameter	Item statistic	Interpretation	Total item (%)
Difficulty index ( <i>p</i> value)	<-2.00	Very easy	0 (0%)
	-2.00 to 2.00	Moderate	21 (60.00%)
	>2.00	Very difficult	14 (40.00%)
Discrimination index (DI)	>1.7	Very high	1 (2.86%)
	1.35 – 1.69	High	2 (5.71%)
	0.65 – 1.34	Moderate	8 (22.86%)
	0.35 – 0.64	Low	18 (51.43%)
	0.01 – 0.34	Very low	3 (8.57%)
Distractor Efficiency (DE)	>0.05	Well functioned	140 (100%)
	<0.05	Unwell functioned	0 (0.00%)

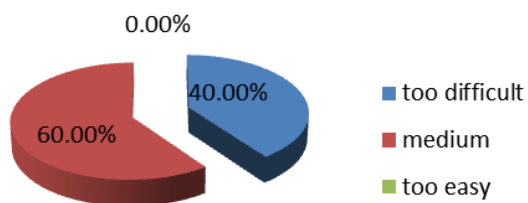


Figure 1. Distribution percentage of difficulty index

The test items should be sequenced from the easiest to the most difficult. Thus, the most difficult items should be placed at the last part of the test. However, when the

subject matter or subject changes, the item's difficulty index begins with the easiest. Figure 2 shows the difficulty index of each item. The simplest item is of the threshold value of -0.78 and is placed in the earliest part of the test. Students' lack of success in taking the test can also be caused by the wrong order of items. The unfavorable placement of difficult items affects the students' result (Debeer & Janssen, 2013, p. 177). Therefore, it should be placed at the end of the test with regard to the materials being tested. This should be a concern for the test developers.

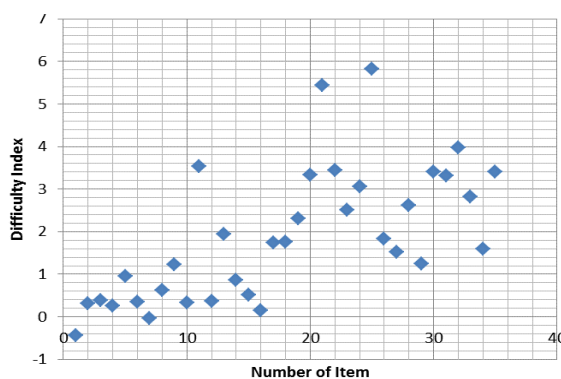


Figure 2. Scatter plot of difficulty index

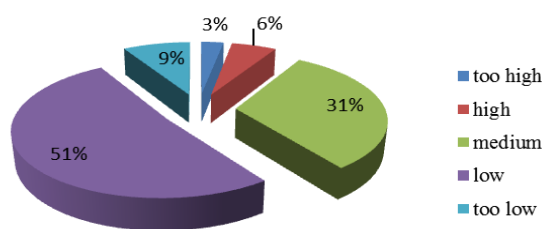


Figure 3. Distribution percentage of discrimination index

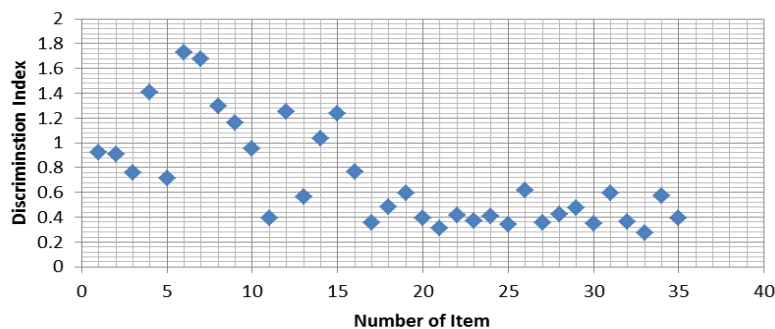


Figure 4. Scatter plot of discrimination index

### *Discrimination Index*

The discrimination index on this test item has been distributed evenly although the majority of the discrimination index is low (see Figure 3). The discrimination index is important to know the difference between high and low ability groups. This instrument needs to be improved so that it has a better discrimination index. Figure 4 shows the discrimination index of each item.

### *Distractor Efficiency*

Analyzing the distractor is done to determine the usefulness of each individual distractor on each item. In this study, 100% of the distractors (140 distractors) functioned well. This means that there is no non-functional distractor. If many students do not choose a particular distractor simultaneously, it is likely that these distractors do not make sense. Thus, the distractor does not effectively distract students. The non-functional distractor (NFD) will reduce the functionality of the distractor itself. The more NFD in the test, the easier the test items will be. Conversely, the fewer NFD in the test, the higher the items' level of difficulty will be. On the other hand, the functioning of the distractors themselves will get better (Allen & Yen, 1979, p. 2; Kolte, 2015, p. 321). The non-functionality of distractors is important in the preparation of a good multiple-choice test.

### *Information Function & Standard Error Measurement (SEM)*

The value of the information function of test items denotes the strength or contribution of each item to revealing the measured latent trait. The information function with the two-parameter model depends on the level of difficulty and also the discrimination index of the item. The greater the value of information, the lower the value of SEM. Based on the results of the analysis, the highest information on the ability  $\theta = 0.4$  with the information function value of 5.38 and SEM = 0.6. This is the maximum value of the information function. The two-point intersection between the information function and SEM is at  $\theta = -1.42$  and  $\theta = 2.65$ . This shows that the test is

suitable for students with the ability of  $-1.42 < \theta < 2.65$ .

There are many things which need to be considered in preparing objective form tests, including: (1) each question item must contain only one correct answer, (2) all distractors must be reasonable, (3) the length of the alternative answers should not give a clue to the correct answer, and (4) the correct answer should appear in each alternative position roughly in the same amount, but randomly (Gronlund, 1982, pp. 189–199). On the other hand, teachers must have the skills in preparing a test so that it is prepared to be of good quality. Educators should be aware that: (1) they should master the subject they teach, (2) they should have the skills to analyze test items, and (3) they should be able to help students make use of the information in the context of formulating educational policies. Thus, teachers' competence is not focused on the mastery of the material only. Their skills in analyzing the test results are also very important (Brookhart, 2011, p. 3). In addition, educators should also be able to arrange items that really matter in accordance with the subject matter to be tested.

In terms of the item's level of difficulty, the distribution of the items is in the levels of easy, medium, or difficult. This indicates the mastery of the material by the students. In terms of students' skill in the material being tested, there is no easy material. The material which has a medium level of difficulty includes determining the result of the subtraction operation on the function, determining the inverse of the linear function, determining the trigonometric ratio on the right triangle, determining the function of the composition consisting of two functions, solving the problem involving the addition operation of the function, determining the composition function of inverse function, determining the value of a composition function consisting of two functions, determining the function if the composition function is identified, determining the result of the mapping on the function, determining the inverse value of the composition function, determining the inverse of the fractional function, determining the function value if the composition function is identified,



determining the composition of the three functions, determining the inverse value of the fractional function, determining the diagonal length of the parallelogram, determining the angle of the triangle if two sides and one angle are known, determining the area of the hexagon if the radius of the outer circle is known, expressing the angle into degrees, determining the sides of a triangle if two angles and one side are identified, and determining the function if the composition is known.

On the other hand, the items which have a high-level of difficulty in the tenth grade Final Examination include determining trigonometric values in various quadrants and related angles, determining the pilot's visibility of the cruise ship if the plane's height and depth angle are identified, determining the cos angle if it is known to the three sides, determining the area of the triangle if two sides and an angle are known, analyzing the identity of trigonometry, determining the angle if both sides and extent are known, determining trigonometric values in various quadrants and related angles, determining the circumference of the octagon if the radius of the outer circle is found, determining the graphic equation, drawing the distance between the lower end of the staircase and the wall if the length of the ladder and the angle formed between the stairs and the floor is known, determining the inverse of the composition function, determining the area of the triangle if the length of the three sides is known, determining the trigonometric value in the various quadrants and correlation angles, and analyzing the identity of trigonometry.

In the tenth grade mathematics final examination test items, the easiest material is determining the result of the subtraction operation on the function. This is indicated by the smallest threshold value of -0.43. The difficult materials in the test include determining trigonometric values in various quadrants and related angles, determining the cos angle if the three sides are known, determining the angle if both side and the extent are known, and determining the trigonometric value in various quadrants and angle-related corners.

One of the difficult items found is determining the value from  $\sec \frac{2\pi}{3} \times \tan \frac{5\pi}{6} \times \sin \frac{3\pi}{2}$

with the answer choice A.  $-2\sqrt{3}$ , B.  $-\frac{2\sqrt{3}}{3}$ , C.  $\frac{2\sqrt{3}}{3}$ , D.  $\sqrt{3}$ , and E.  $2\sqrt{3}$ . This problem requires several stages of completion. Each factor must be determined in advance.  $\sec \frac{2\pi}{3}$  are -2,  $\tan \frac{5\pi}{6}$  are  $-\frac{\sqrt{3}}{3}$ , and  $\sin \frac{3\pi}{2}$  are -1 so that the result of these values is  $-\frac{2\sqrt{3}}{3}$  (B). This answer option was only chosen by 78 students (22%). The highest number of students' answers was in choice C, i.e. 125 students (35%). Alternatives A and E were chosen by 43 students (12%), 63 students (18%) chose Alternative D and the other students did not choose any option.

This finding is consistent with the finding of the research conducted in Riau Province, that calculating trigonometric ratio with sinus, cosine, and tangent formulas is a difficult subject in high school maths (Aisyah, 2013, p. 153). These results provide an illustration that the basic competence of high school mathematics has not been achieved in terms of learning indicators. Therefore, the need for evaluation and improvement in the learning process of mathematics is urgent to improve the students' learning achievement. The results of this study can also be an input to improve the teachers' competence, especially in the teaching of these difficult materials.

The same findings are also expressed by Wongapiwatkul, Laosinchai, and Panijpan (2011, p. 54) that studying trigonometry is difficult for students and the difficulties are caused by many interconnected things. Students may first learn trigonometric function, or learn it because they have difficulty in reasoning in trigonometry. In addition, manipulating trigonometric calculations is not the same as manipulating the algebra operation.

## Conclusion

The level of difficulty of the tenth grade mathematics final examination test items is in the medium category. Overall, the test items have a very good distractor efficiency. Of all given distractors, they were selected by over 5% of the test takers. The discrimination index in this test is not good because it is only at medium, low, and very low levels. In this test, it is known that the difficult materials in this test are: (1) determining the distance of an object using the concept of depression angle,

(2) determining the area of a triangle if two sides and one slant angle are known, (3) analyzing the trigonometric identity, (4) determining an angle if the length of the two sides of the triangle and its width are known, (5) determining the equation of a trigonometric graph if the graph is known, (6) determining the trigonometric values in various quadrants and related angles, (7) determining the circumference of an octagon if the diameter of the outer circle is known using the trigonometric formula, and (8) determining the area of a triangle if the length of the three sides is measured.

This finding is consistent with Aisyah's research finding that trigonometric material is difficult for learners (Aisyah, 2013, p. 153). She states that determining the equation for trigonometric charts if the graph images are known and determining trigonometric values in various quadrants and related angles are difficult material in mathematics. Educators must be concerned in this area to explore the material, and improve their teaching methods and strategies. Likewise, students should pay attention to the materials better. According to Keoviphone and Wibowo (2015, p. 8), the more systematic educators plan their learning materials, the more likely they will succeed.

## References

- Aisyah, U. (2013). *Pengembangan perangkat pembelajaran kompetensi sulit matematika SMA di Riau. Master Thesis*. Universitas Negeri Yogyakarta, Yogyakarta.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Cole Publishing.
- Arikunto, S. (1999). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Daryanto, M. (2012). *Evaluasi pendidikan*. Jakarta: Rineka Cipta.
- Debeer, D., & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Educational Measurement*, 50(2), 164–185. <https://doi.org/10.1111/jedm.12009>
- Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality Multiple Choice Questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine*, 39(1), 17–20. <https://doi.org/10.4103/0970-0218.126347>
- Gronlund, N. E. (1982). *Measurement and evaluation in teaching* (4th ed.). Cliffs, NY: Macmillan.
- Haladyna, T. M., & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 51–78. [https://doi.org/10.1207/s15324818ame0201\\_4](https://doi.org/10.1207/s15324818ame0201_4)
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Keoviphone, C., & Wibowo, U. B. (2015). Factors discouraging students from schooling: A case study at Junior Secondary School in Laos. *REiD (Research and Evaluation in Education)*, 1(1), 1–12. <https://doi.org/10.21831/reid.v1i1.4894>
- Kolte, V. (2015). Item analysis of multiple choice questions in physiology examination. *Indian Journal of Basic and Applied Medical Research*, 4(4), 320–326.
- Mardapi, D. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in*



- teaching* (10th ed.). Upper Saddle River, NJ: Pearson.
- Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.
- Saudi Commission for Health Specialties. (2011). *Item writing manual for multiple-choice questions*.
- Sugiyono. (2001). *Metode penelitian bisnis*. Bandung: Alfabeta.
- Sulistiawan, C. H. (2016). Kualitas soal ujian sekolah matematika program IPA dan kontribusinya terhadap hasil ujian nasional. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(1), 1–10. <https://doi.org/10.21831/pep.v20i1.7516>
- Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education*, 9(40), 1–8. <https://doi.org/10.1186/1472-6920-9-40>
- Tshabalala, T., & Ncube, A. C. (2014). The effectiveness of measurement and evaluation in Zimbabwean primary schools: Teachers and heads' perceptions. *International Journal of Innovation and Applied Studies*, 8(1), 141–148.
- Wongapiwatkul, P., Laosinchai, P., & Panijpan, B. (2011). Enhancing conceptual understanding of trigonometry using earth geometry and the great circle. *Australian Senior Mathematics Journal*, 25(1), 54–63.