



Item Analysis of Multiple-Choice Questions in Junior Secondary School Mathematics

Diandra Zivana

Universitas Negeri Semarang, Indonesia

Corresponding Author: zivana@unnes.ac.id

ABSTRACT

The quality of test items plays a crucial role in ensuring the validity and reliability of learning assessments in mathematics education. This study aims to analyze the quality of multiple-choice questions used in mathematics assessments at the junior secondary school level. The analysis focuses on several important indicators of item quality, including the difficulty index, discrimination index, and distractor effectiveness. The research employs a quantitative descriptive approach using item analysis techniques. The data were obtained from students' responses to a mathematics test consisting of multiple-choice questions administered to junior secondary school students. The collected data were analyzed using classical test theory to determine the level of difficulty, the ability of items to differentiate between high- and low-performing students, and the effectiveness of distractors in each question. The results indicate that a proportion of the test items fall into acceptable categories in terms of difficulty and discrimination, while several items require revision due to low discrimination power and ineffective distractors. These findings highlight the importance of systematic item analysis to improve the quality of assessment instruments in mathematics education. The study suggests that teachers should regularly conduct item analysis to ensure that test items accurately measure students' mathematical understanding and support effective learning evaluation.

Keywords: Mathematics Assessment, Difficulty Index, Discrimination Index

Received: 01.01.2026	Revised: 01.02.2026	Accepted: 20.02.2026	Available online: 08.03.2026
-------------------------	------------------------	-------------------------	---------------------------------

INTRODUCTION

Two important components in the world of education are teachers and students. Teachers who direct students in learning and students who are the objects of that learning (Inawati et al., 2022; Putri Andini & Mukhlis, 2023; Verawati et al., 2023). Pradani & Efendi, (2023) say that teacher competency standards include 4 main competencies, namely pedagogical competency, personality competency, social competency and professional competency. Pedagogical competency is a teacher's ability to manage student learning which at least includes the following things: (a) understanding the insight or foundation of education, (b) understanding students, (c) curriculum/syllabus development, (d) learning planning, (e) implementation of educational and dialogical learning, (f) use of learning technology; g) evaluation of learning outcomes (EHB), and (h) development of students to actualize the various potentials they have.

In the world of education, evaluation is carried out to determine the quality of the educational components. For example, a teacher conducts an evaluation of students, the aim is to find out the extent of the students' abilities. Evaluation or assessment is an umbrella concept that covers several underlying concepts, for example measurement and tests. Evaluation refers to or contains the meaning: an action or a process to determine the value of something. Assessment is the process of giving or determining value to certain objects based on certain criteria. Assessment is carried out with the aim of assessing student learning processes and outcomes at school, diagnosing students' learning difficulties and determining grade promotion or graduation. The function of assessment is as a tool to determine whether or not instructional goals have been achieved, provide feedback for improving the teaching and learning process, increase student learning motivation, and as a basis for compiling reports on student learning progress to their parents (Hidayatunnisa'i et al., 2023; Nur Cahyo et al., 2022; Prawiki & Helendra, 2022).

In general, the evaluation tools used can be classified into two types, namely tests and non-tests. Test techniques can be carried out both written and unwritten. Non-test techniques are used to assess attitudes, skills, behavior, and so on. Tests in education are systematic, valid, reliable and objective assessment tools or assessment methods to determine students' abilities, skills and level of knowledge regarding

teaching materials, in the form of a task or problem that must be solved by a student or group of students (Hanan et al., 2023; Kusumayanti & Jannah, 2022; Rismaulhijjah & Kuswanti, 2022). Based on the objectivity of scoring, tests are divided into objective tests and subjective tests. A multiple-choice test is a form of text that has the correct or most appropriate answer. In this test, the correctness of the answers is absolute. The answer only has four possibilities, namely A, B, C, and D.

The Mathematics subject test in connection with or as part of Mathematics learning is a tool used to try to measure how much students have mastered the Mathematics they have studied. The Mathematics subject test is a tool to measure student abilities. The activity of analyzing question items (item analysis) is an activity that teachers must carry out in order to improve the quality of the questions that have been prepared. And the aim is to identify good, bad and bad questions. By analyzing questions, information can be obtained about the shortcomings of a question and "hints" for making improvements. In this case the author will discuss the study of test content, namely analyzing the level of validity, reliability, level of difficulty, and discrimination. Mira et al., (2022); R. P. Sari, (2022); V. N. I. Sari et al., (2022), As the name suggests, teacher-made tests are tests created by the teacher himself. The test is intended to measure the level of success of students in achieving goals after the teaching process is managed by the teacher concerned.

METHOD

The research method used in this research is descriptive research method using a quantitative approach. According to (Sugiyono, 2017, 2019) quantitative descriptive research methods aim to describe phenomena, events, symptoms and events that occur factually, systematically and accurately. Phenomena can take the form of forms, activities, relationships, characteristics as well as similarities and differences between phenomena. The research was conducted in Class VIII at SMP Muhammadiyah 13 Surabaya in the odd semester of the 2024/2025 academic year using multiple choice questions designed to determine students' understanding of the Class VIII Mathematics subject at SMP Muhammadiyah 13 Surabaya.

Table 1. Indicators of understanding of material regarding Class VIII Mathematics

No	Learning objectives
1.	Identify the Religious and Character Education section and describe its function
2.	Describe the process of Religious and Character Education and relate the importance of this process to living creatures
3.	Identify and create simulations using simple charts or tools regarding the Religious and Character Education cycle.

The measurement of the level of difficulty of a question item is produced based on a comparison of the number of incorrect and correct responses (Jumini et al., 2023; Marambaawang et al., 2023; Riau & Mukhlis, 2023) and then the results are compared with the question item difficulty index criteria in table 2.

Table 2. Question item difficulty index criteria

Difficulty Index Value	Interpretation
0.00-0.15	very difficult
0.16-0.30	difficult
0.31-0.70	keep
0.71-0.85	easy
0.86-1.00	It's easy

The discriminating power index for questions is classified based on the quality of the questions which can be referred to from the final score obtained by each respondent (Hanan et al., 2023). This value is used as a reference in the Pearson product moment test to classify the calculated r value for each question item to obtain a distinguishing power index. There are 6 classifications of the differentiating power index categories as shown in table 3.

Table 3. Distinctive power index categories for test items

Differentiating Power Index Value	Criterion
Negative	Broken
0 - 0.19	Bad
0.20 - 0.29	Keep
0.30 - 0.39	Good
= 0.4	Very good
> 0.4	Superior

RESULTS

The validity test is carried out in two ways, namely (1) comparing the Pearson correlation value or rcount value with the rtable value, the rcount value must be greater than the rtable value ($\text{rcount} > \text{rtable}$). 2) Combining the Pearson correlation value with the significance obtained, the value obtained must be positive and supported by a significance value < 0.05 . Test results using Pearson correlation can be seen in table 4.

Table 4. Validity test results

No Item	Result R Calculate	Significant Value	Validity
item01	.631**	0.000	valid
item02	0.318	0.087	Invalid
item03	.635**	0.000	valid
item04	.696**	0.000	valid
item05	.486**	0.007	valid
item06	.512**	0.004	valid
item07	.528**	0.003	valid
item08	.488**	0.006	valid
item09	.474**	0.008	valid
item10	.440*	0.015	valid
item11	0.231	0.220	Invalid
item12	.671**	0.000	valid
item13	.393*	0.032	valid
item14	.789**	0.000	valid
item15	.538**	0.002	valid
item16	.465**	0.010	valid
item17	.482**	0.007	valid
item18	.520**	0.003	valid
item19	.696**	0.000	valid
item20	.577**	0.001	valid
item21	.465**	0.010	valid
item22	.482**	0.007	valid
item23	.520**	0.003	valid
item24	.696**	0.000	valid
item25	.577**	0.001	valid

In this research, based on the data above, it can be stated that the validity test questions are categorized as valid with a high value of 789 and the highest significant value is 0.000, so this instrument is suitable for use as a measuring tool in evaluating Class VIII Mathematics. The reliability test based on Chronbach alpha compares the test result values with a magnitude of 0.06. If the test result value is greater than 0.06 then the question item is declared reliable. The data shown is stated in table 5.

Table 5. Reliability Test

Reliability Statistics	
Cronbach's Alpha	N of Items
.901	23

The suitability parameters of an instrument are not only determined by validity and reliability but must also be able to describe differences in students' knowledge and abilities, therefore a difficulty index instrument and the ability to differentiate questions are needed to be able to differentiate the ability level of each student (Ismiyati et al., 2023; Rahayu et al., 2023; R. P. Sari, 2022). The item difficulty index is carried out to identify the level of difficulty of each item by comparing the test results and the item difficulty index in table 2. The results of the difficulty index test show that the item items are in the very easy category of 20%, the item items are in the medium category. 72% and the questions are in the easy category 20%. More detailed results are presented in table 6.

Table 6. Question item difficulty index

No Item	R Value Calculate	Decision Making Criteria	Criterion
1	0.60	Difficulty index 0.00-0.15 = very difficult 0.16-0.30 = difficult 0.31-0.70 = medium 0.71-0.85 = easy 0.86-1.00= very easy	medium
2			
3	0.66		medium
4	0.50		medium
5	0.70		medium
6	0.60		medium
7	0.76		easy
8	0.66		medium
9	0.73		easy
10	0.76		easy

No Item	R Value Calculate	Decision Making Criteria	Criterion
11			
12	0.53		medium
13	0.70		medium
14	0.53		medium
15	0.63		medium
16	0.76		easy
17	0.63		medium
18	0.70		medium
19	0.60		medium
20	0.60		medium
21	0.76		easy
22	0.63		medium
23	0.70		medium
24	0.60		medium
25	0.60		medium

The question discriminating power test functions to describe the differences in abilities possessed by each respondent and compare them with the question discrimination power category. The test results show that the questions are at a very good level with a presentation of 4%, the questions are in the good category with a presentation of 64%, and the questions are in the fair category with a presentation of 24%. More detailed results are presented in table 7.

Table 7. Results of the differentiating power test of questions

No item	Result R Calculate		Validity
1	0.576		good
2			
3	0.593		good
4	0.659		good
5	0.471		good
6	0.454		good
7	0.481		good
8	0.403		good
9	0.387		enough
10	0.411		good

Differentiation power interpretation criteria
0.70-1.00 = Very good
0.40-0.69 = good
0.20-0.39 = enough
0.00-0.19 = ugly

No item	Result R Calculate	Validity
11		
12	0.612	good
13	0.292	enough
14	0.761	Very good
15	0.489	good
16	0.398	enough
17	0.378	enough
18	0.510	good
19	0.676	good
20	0.515	good
21	0.398	enough
22	0.378	enough
23	0.510	good
24	0.676	good
25	0.515	good

In this study, the results of the differentiating power test of the questions stated that all the questions tested were in the adequate and good categories. So, it can be stated that all the questions are of good enough quality so they do not require improvement and can be used as an evaluation tool in learning.

CONCLUSION

The development of an evaluation measuring tool in the form of multiple-choice questions to measure conceptual understanding of Mathematics learning in Class VIII at SMP Muhammadiyah 13 Surabaya has been successful. The results of the analysis show that all the questions tested are valid and reliable, so they can be used as a quality evaluation tool. The results of the difficulty index test show that 20% of the questions are in the very easy category, 72% of the questions are in the medium category and 20% of the questions are in the easy category. Meanwhile, the test of differentiating power of questions. The test results show that the questions are at a very good level with a presentation of 4%, the questions are in the good category with a presentation of 64%, and the

questions are in the fair category with a presentation of 24%. The questions presented can be analyzed with valid and reliable results and can be distributed well. So, it can be concluded that these question items are suitable for use in the mid-semester exam. The benefits that can be obtained from the results of this research are that it is best to always prepare reserve questions to replace questions in the bad category, revisions can be made to questions in the good enough category so that they can be used again, and questions in the good category can be added to the question bank.

REFERENCES

- Hanan, M. P., Jannah, R. R., & Alim, J. A. (2023). Analisis Butir Soal Matematika Materi Kpk Dan Fpb Berbasis Cerita Di Sdn 111 Pekanbaru. *Journal Of Educational Learning And Innovation (Elia)*, 3(1). <https://doi.org/10.46229/Elia.V3i1.538>
- Hidayatunnisa'i, Kusrini, & Kusnawi. (2023). Perbandingan Kinerja Metode Naïve Bayes Dan Support Machine (Svm) Dalam Analisis Kualitas Butir Soal Pilihan Ganda. *Jurnal Fasilkom*, 13(02). <https://doi.org/10.37859/Jf.V13i02.5087>
- Inawati, Arya, A., & Lasmiatun. (2022). Analisis Tingkat Kesulitan Butir Soal Mata Pelajaran Bahasa Indonesia Buatan Guru. *Bastrando*, Vol 1(No 2).
- Ismiyati, I., Raharjo, T. H., Tusyanah, T., & Sholikah, M. (2023). Pelatihan Analisis Butir Soal Berdasarkan Teori Tes Klasik Berbantuan Itekan Untuk Meningkatkan Kualitas Instrumen Penilaian. *Japi (Jurnal Akses Pengabdian Indonesia)*, 8(2). <https://doi.org/10.33366/Japi.V8i2.5064>
- Jumini, S., Madnasri, S., Cahyono, E., & Parmin, P. (2023). Analisis Kualitas Butir Soal Pengukuran Literasi Sains Melalui Teori Tes Klasik Dan Rasch Model. *Prosiding Seminar Nasional Pascasarjana Unnes*.
- Kusumayanti, A., & Jannah, N. (2022). Analisis Butir Soal Ujian Masuk Mandiri Uin Alauddin Makassar Dengan Teori Tes Modern. *Mapan : Jurnal Matematika Dan Pembelajaran*, 10(1).
- Marambaawang, D., Bano, V. O., & Hada Enda, R. R. (2023). Analisis Kualitas Butir Soal Penilaian Akhir Semester Gasal Tahun 2021/2022 Menggunakan Itekan Di Smp Negeri 1 Kambera. *Dharmas Education Journal (De_Journal)*, 4(1). <https://doi.org/10.56667/Dejournal.V4i1.961>

- Mira, K., Hayon, V. H. B., & Tinenti, Y. R. (2022). Analisis Butir Soal Pilihan Ganda Pada Materi Pokok Larutan Elektrolit Dan Nonelektrolit Kelas X Ipa Sma Negeri 1 Lamba Leda Manggarai Timur. *Wahana*, 73(2). <https://doi.org/10.36456/Wahana.V73i2.4201>
- Nur Cahyo, A., Luriawati, D., & Wagiran, W. (2022). Analisis Butir Soal Penilaian Keterampilan Kebahasaan Pada Pembelajaran Teks Eksplanasi Kelas Xi. *Jbsi: Jurnal Bahasa Dan Sastra Indonesia*, 2(01). <https://doi.org/10.47709/Jbsi.V2i01.1493>
- Pradani, R. A., & Efendi, A. (2023). Analisis Butir Soal Ujian Sekolah Menggunakan Program IteMan (Analysis Of School Exam Questions Using The IteMan Program). *Indonesian Language Education And Literature*, 8(2). <https://doi.org/10.24235/Ileal.V8i2.11002>
- Prawiki, S. M., & Helendra. (2022). Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Tahun Pelajaran 2020/2021 Mata Pelajaran Biologi Kelas X Sma Negeri 1 Teluk Sebong. *Biodidaktika: Jurnal Biologi Dan Pembelajarannya*, 17(2).
- Putri Andini, D., & Mukhlis, M. (2023). Analisis Butir Soal Pada Instrumen Asesmen Kompetensi Minimum Literasi Membaca Di Smp It Insan Utama Pekanbaru. *Diglosia: Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, 6(2). <https://doi.org/10.30872/Diglosia.V6i2.658>
- Rahayu, W., Sulaeman, E., Arnawisuda Ningsi, B., Arofah, I., & Akbari, W. (2023). Analisis Kemampuan Pemahaman Guru Sekolah Dasar Dalam Mengembangkan Butir Soal Berbasis Literasi Numerik. *Perduli: Jurnal Pengabdian Kepada Masyarakat*, 4(01). <https://doi.org/10.21009/Perduli.V4i01.30490>
- Riau, U. I., & Mukhlis, M. (2023). Analisis Butir Soal Literasi Membaca Pada Instrumen Asesmen Kompetensi Minimum Siswa Smk. In *Sastra Indonesia Dan Daerah* (Vol. 13, Issue 2).
- Rismaulhijjah, W., & Kuswanti, N. (2022). Analisis Butir Soal Ulangan Harian Hasil Pengembangan Guru Materi Sistem Gerak Manusia Kelas Xi Ipa. *Berkala Ilmiah Pendidikan Biologi (Bioedu)*, 11(3). <https://doi.org/10.26740/Bioedu.V11n3.P643-661>
- Sari, R. P. (2022). Analisis Butir Soal Ujian Akhir Semester Genap Mata Pelajaran Matematika Kelas Xi Smk Negeri 1 Percut Sei Tuan T.A. 2019/2020. *Jurnal Penelitian Pendidikan Mipa*, 7(1). <https://doi.org/10.32696/Jp2mipa.V7i1.1360>

- Sari, V. N. I., Utomo, A. P. Y., & Sumarwati. (2022). Kualitas Soal Bahasa Indonesia Di Smp Muhammadiyah 1 Pontianak: Analisis Butir Soal Vina. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 11(2).
- Sugiyono. (2017). *Metode Penelitian Kuantitatif Kualitatif Dan R&D* (25th Ed.). Alfabeta.
- Sugiyono, S. (2019). *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D* (Sutopo (Ed.); 1st Ed.). Alfabeta.
- Verawati, Y., Siskawati, F. S., & Susilaningtyas, T. (2023). Analisis Butir Soal Ujian Akhir Semester (Uas) Mata Pelajaran Matematika Pada Tahun Ajaran 2020/2021 Kelas Vii Smp Islam At Tanwir Kecamatan Ledokombo Kabupaten Jember. *Jurnal Jendela Pendidikan*, 3(01). <https://doi.org/10.57008/jjp.v3i01.422>