

## Analysis of Criteria for School Exam Questions in Science Subjects at SMP Negeri 2 Ngoro

Sulistiyowati

Universitas PGRI Adibuana Surabaya

Email: [sulis.styowati778@gmail.com](mailto:sulis.styowati778@gmail.com)

### Abstract

Quality test questions are questions that can provide precise information about student learning outcomes. The purpose of this study was to analyze the quality of the Science Subject School Exam questions at SMPN 2 Ngoro for the 2022/2023 academic year which consisted of 20 questions. The research sample consisted of 62 student responses in class IXA and IXF with purposive sampling method. This study uses a quantitative and qualitative approach with data analysis techniques using the SPSS version 25 application and Microsoft excel. The results showed that the items used were classified as difficult 15%, items classified as moderate 60% and items classified as easy 25%. The analysis of the discriminatory power of the questions showed that 70% of the items already had good discriminating power and the other 30% did not have good discriminating power and needed to be revised. The analysis of the distractor options from the items found that 96% of the total distracting options on these items were good and the rest were not good and needed to be revised.

**Keywords:** Item analysis, difficulty level, discrimination, distractor effectiveness

**Artikel History:** November 5, 2024: Received in revised form March 23, 2025: Accepted March 31,2025: Available online March 31,2025.

### Introduction

Evaluation and assessment activities are inseparable activities in the learning process, because the results of these two activities provide an overview of the quality of the learning process. The quality of learning in the classroom is highly dependent on teacher performance because the teacher is the most important director in the learning process. As stated by (Laka & Tuasikal, 2019) that effective classroom learning requires professional commitment from teachers. Assessment and evaluation activities are often carried out by teachers to see and determine the level of achievement of student learning outcomes in their learning experiences. (Rahayu & Djazari, 2016) said that evaluation is carried out with the aim of obtaining evidentiary data, so that it becomes an indication of the level of ability and success of students in achieving learning objectives. Evaluation also aims to

measure and assess the effectiveness of teaching and the methods that have been applied or implemented by educators. Supporting the above opinion, (Kurniawan, 2015) also explains that evaluation of learning outcomes aims to assess competency achievement and improve the learning process as well as guidelines for compiling reports on student learning progress.

One of the techniques used to evaluate student learning outcomes is by testing. Tests are the most frequently used measuring tools to determine student success in the learning process. (Sudijono, 2001) explains that a test is a method or procedure that needs to be taken in measuring and assessing in the field of education, which is in the form of giving assignments or a series of assignments in the form of questions that must be answered or commands that must be carried out by the testee, so that a value can be produced that

symbolizes the behavior or achievement of the testee. Then (Septiana, 2016) also explained that tests can be arranged in objective or subjective form. An objective test is a question for which all the information needed to answer the question is available. A subjective test is a form of question consisting of questions that require answers in the form of relatively long descriptions. In compiling test questions, it is also necessary to pay attention to their quality. (Sudijono, 2001) in (Rahayu & Djazari, 2016), said that a quality test can provide real information about student learning outcomes, so that this information can be used to measure the level of development or progress that has been achieved by students after they have gone through the teaching and learning process for a certain period of time. In addition, information obtained from quality tests can be used to find out how far the predetermined teaching program can be achieved. According to Arikunto in (Afrian et al., 2018), a quality test must meet the following requirements: validity, reliability, objectivity, practicality, and economy. A test is said to be valid if it can provide appropriate information and can be used to achieve certain goals. A test is said to be reliable if it always gives the same results at different times or opportunities. A test is said to be objective if in its implementation, there are no subjective factors that influence it, especially in the scoring system. A test is said to be economical if it does not require a lot of money, effort, and time. Adiputra (2012) explained several benefits of analyzing the quality of test items, namely: (1) determining whether a test item function is as expected, (2) providing input to students about their abilities and as a basis for discussion materials in class, (3) providing input to teachers about student difficulties, (4) providing input on certain aspects for curriculum development, (5) revising the material that is assessed or measured, (6) improving test writing skills. Given the importance of test quality, teachers need to evaluate each test item given to students. The results of this study will be an indication of whether the quality of these questions is good enough to be used to test student learning outcomes or not.

In the 2022/2023 academic year, SMP Negeri 2 Ngoro has implemented School Exams. The implementation of School Exams is a very important requirement as a standard for determining graduation for students. This is based on the Circular of the Minister of Education and Culture Number 1 of 2021 concerning "Elimination of National Exams and Equivalency Exams and Implementation of School Exams during the Emergency Period of the Spread of Covid Virus Disease (Covid-19)". Given the importance of implementing school exams, in compiling the outline of these exam questions, teachers need to pay attention to the quality of the questions used. Based on the background above, this study aims to analyze the quality of the School Exam questions for grade IX Science subjects at SMP Negeri 2 Ngoro. The analysis of the question criteria includes an analysis of the level of difficulty, discrimination power, and effectiveness of distractors.

## Methods

This study uses a mixed method research type, namely by using a quantitative and qualitative approach. This research was conducted at SMP Negeri 2 Ngoro with a sample selection technique using a purposive sampling technique. The research variables in this study are the criteria for the Science School Examination questions for the 2022/2023 Academic Year consisting of 20 questions. The criteria for these questions were obtained from 62 student respondents, consisting of classes IXA and IXF. The data collection technique in this study used interview techniques and documentation (test results). The research data that had been collected was then analyzed using the SPSS version 25 application and also Microsoft Excel. The analysis of these question criteria includes the level of difficulty, discrimination power and effectiveness of distractor options for each question used. The SPSS application is used to test the level of difficulty and discrimination power of each question used, while Microsoft Excel is used to test the effectiveness of distractors for each question used.

## Results

### Level of difficulty of questions

Table 1 below shows the results of the test of the level of difficulty of the questions from 62 students consisting of classes IXA and IXF for the 2022/2023 Academic Year Science Subject School Examination consisting of 20 questions.

**Tabel 1**  
 Hasil Uji Tingkat Kesukaran Butir Soal

Nomor soal	N		Mean	Keterangan
	Valid	Missing		
1	62	0	0,82	mudah
2	62	0	0,71	sedang
3	62	0	0,34	sedang
4	62	0	0,56	sedang
5	62	0	0,81	mudah
6	62	0	0,92	mudah
7	62	0	0,13	sukar
8	62	0	0,40	sedang
9	62	0	0,53	sedang
10	62	0	0,47	sedang
11	62	0	0,71	sedang
12	62	0	0,63	sedang
13	62	0	0,77	mudah
14	62	0	0,15	sukar
15	62	0	0,71	sedang
16	62	0	0,73	mudah
17	62	0	0,61	sedang
18	62	0	0,37	sedang
19	62	0	0,27	sukar
20	62	0	0,58	sedang

### Discriminating power

The results of the test of the discriminating power of the test items are presented in table 2 below.

**Tabel 2**  
 Hasil Uji Daya Beda Soal

Nomor soal	Skor total		N	Keterangan
	Pearson Correlation	Sig. (2-tailed)		
1	0,248	0,052	62	revisi
2	.548**	0,000	62	baik
3	.548**	0,000	62	baik
4	0,217	0,090	62	revisi
5	0,218	0,088	62	revisi
6	.324*	0,010	62	sedang
7	-.317*	0,012	62	tolak
8	.648**	0,000	62	baik
9	.376**	0,003	62	sedang
10	.336**	0,008	62	sedang
11	.365**	0,004	62	sedang
12	.375**	0,003	62	sedang
13	.619**	0,000	62	baik
14	0,221	0,085	62	revisi
15	.408**	0,001	62	baik
16	0,249	0,051	62	revisi
17	.604**	0,000	62	baik
18	.472**	0,000	62	baik
19	0,155	0,231	62	tolak
20	.591**	0,000	62	baik

### Effectiveness of distractors

The results of the effectiveness (functioning) test of distractor options on each question item are shown in Table 3.

**Tabel 3**  
 Hasil Keterfungsian Option Pengacoh Setiap Butir Soal

Nomor soal	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
Option-A	0,145	0,113	0,242	0,129	0,806	0,919	0,032	0,403	0,161	0,468
Keterangan	baik	baik	baik	baik	baik	baik	baik	baik	baik	baik
Option-B	0,823	0,129	0,339	0,274	0,113	0,000	0,597	0,145	0,532	0,242
Keterangan	baik	baik	baik	baik	baik	ditolak	baik	baik	baik	baik
Option-C	0,016	0,048	0,129	0,565	0,048	0,048	0,129	0,129	0,161	0,145
Keterangan	revisi	baik	baik	baik	baik	baik	baik	baik	baik	baik
Option-D	0,016	0,710	0,290	0,032	0,032	0,032	0,242	0,323	0,145	0,145
Keterangan	revisi	baik	baik	baik	baik	baik	baik	baik	baik	baik

  

No.11	No.12	No.13	No.14	No.15	No.16	No.17	No.18	No.19	No.20
0,048	0,177	0,065	0,419	0,097	0,726	0,613	0,129	0,452	0,581
baik	baik	baik	baik	baik	baik	baik	baik	baik	baik
0,048	0,629	0,774	0,339	0,710	0,113	0,129	0,194	0,274	0,242
baik	baik	baik	baik	baik	baik	baik	baik	baik	baik
0,710	0,097	0,113	0,145	0,097	0,063	0,177	0,323	0,210	0,145
baik	baik	baik	baik	baik	baik	baik	baik	baik	baik
0,194	0,097	0,048	0,097	0,097	0,097	0,081	0,355	0,065	0,032
baik	baik	baik	baik	baik	baik	baik	baik	baik	baik

## Discussion

### Analysis of the level of difficulty of the questions

The level of difficulty of the questions is the opportunity to answer a question correctly at the level of ability or it can be said that to find out whether the question is classified as an easy question or a difficult question (Fitriawati, 2017). A good question is a question that is not too easy and not too difficult (Susanto et al., 2015). The test question script should use questions with a balanced level of difficulty, namely difficult = 25%, medium = 50%, and easy = 25%. The level of difficulty of the questions is seen from the students' ability to answer them, not from the teacher's perspective as the question maker (Kurniawan, 2015).

Formula for difficulty level (Uno, H. B., & Koni, 2018):

$$I = \frac{B}{N}$$

Description:

I = Item difficulty index

B = Number of students who answered the item correctly

N = Number of students who took the test

Table 1 presents the relationship between the level of difficulty and the quality of the test items.

**Tabel 4**  
 Hubungan Antara Tingkat Kesukaran Dengan Kualitas Butir Soal

Indeks kesukaran	Kategori Butir Soal
0,00- 0,30	Sukar, butir soal kurang baik, direvisi
0,31-0,70	Sedang, butir soal cukup baik, digunakan
0,71-1,00	Mudah, butir soal baik, direvisi

This difficulty level index is generally expressed in the form of a proportion ranging from 0.00 to 1.00. The greater the difficulty level index obtained from the calculation results, the easier the question is (Kadir, 2015). Table 1 presents the results of the test of the level of difficulty of the School Exam questions at SMPN 2 Ngoro. This test of the level of difficulty analysis of the questions aims to determine the extent of the level of difficulty of the questions made by the teacher to test student learning outcomes. Based on the results of the test of the level of difficulty analysis of the questions above, the questions that are classified as difficult are numbers 7, 14, 19. The index of difficulty of the questions starts from a range of 0.00 - 0.30. Questions that have a moderate level of difficulty, meaning questions that are not too difficult and also not too easy, are found in numbers 2, 3, 4, 8, 9,10, 11, 12, 15, 17, 18, and 20. The Mean value in table 1 shows a range of values from 0.31 - 0.70. Meanwhile, questions that are classified as easy are found in numbers 1, 5, 6, 13, 16. The Mean value in table 1 shows a range of values from 0.71 - 1.00. Questions are classified as difficult, moderate and easy according to the relationship between the level of difficulty and the quality of the questions in table 4. The ease or difficulty of a test question depends on the percentage of students who answer correctly (value 1) or incorrectly (value 0) on a question item. If more students

answer correctly, the difficulty level index value will be high, meaning that the question is easy, and vice versa, if more students answer incorrectly, the difficulty level index value will decrease, meaning that the question is difficult. (Widoyoko, 2018) & (Kumiawan, 2015) explain that the test question script should use questions with a balanced level of difficulty, namely difficult = 25%, medium = 50%, and easy = 25%. In the data analysis results above, it is known that the questions that are classified as difficult are 15%, the questions that are classified as medium are 60% and the questions that are classified as easy are 25%. Based on this data, it can be concluded that the criteria for the level of difficulty of the questions almost have a balance between the percentage of the number of questions that are classified as easy, medium, and difficult. If the question is to be used again, it is necessary to make improvements to the questions whose level of difficulty does not match the theory. Discriminating Power Analysis

The discriminating power of a question item is the ability of a question item to differentiate groups in the aspects measured according to the differences in the groups (Bagiyono, 2017). One of the objectives of the item discriminating power analysis is to determine whether or not a question item is able to differentiate between high-ability training participants and low-ability training participants.

The formula for discriminating power according to Glass and Stanley in (Kadir, 2015). To find out the discriminating power of a multiple-choice test is to use the following formula:

$$DP = \frac{BA - BB}{\frac{1}{2} N} \text{ atau } DP = \frac{2(BA - BB)}{2}$$

Description:

DP = Discriminatory power of the question,

BA = Number of correct answers in the upper group,  
 BB = Number of correct answers in the lower group,  
 N = Number of students who took the test  
 To determine the discriminatory power of the Multiple Choice test, use the following formula:

$$DP = \frac{\text{mean kelompok atas} - \text{mean kelompok bawah}}{\text{skor maksimum soal}}$$

The relationship between discrimination power and item quality is classified into four categories according to Dali in (Fitrianawati, 2017). Table 5 presents the relationship.

**Tabel 5**  
 Hubungan antara Daya Beda dengan Kualitas Butir Soal

Kualitas Butir Soal	Daya Beda
Baik, dapat digunakan tanpa revisi	0,40 – 1,00
Cukup Baik (sedang), dapat digunakan dengan revisi	0,30 – 0,39
Kurang Baik (revisi), perlu pembahasan dan direvisi	0,20 – 0,29
Tidak baik (tolak), dibuang atau diganti	-1,00 – 0,00

One of the objectives of item discrimination power analysis is to determine whether or not a question item is able to differentiate between high-ability training participants and low-ability training participants. Based on the data from the item discrimination power test results in table 2, it is known that numbers 2, 3, 8, 13, 15, 17, 18, and 20 are questions that have good discrimination power. In table 2, discrimination power can be seen from the Pearson Correlation value ranging from 0.40 to 1.00. Numbers 6, 9, 10, 11, 12 are questions that have moderate discrimination power. In table 2, discrimination power can be seen from the Pearson Correlation value ranging from 0.30 to 0.39. Meanwhile, 1, 4, 5, 14, 16 are questions that have moderate discrimination power and need to be revised. Discriminatory power can be seen from the Pearson Correlation value ranging from 0.20 to 0.29. For numbers 7 and 19, these are items that have poor discrimination power and must be discarded. The

discrimination power can be seen from the Pearson Correlation value ranging from -1.00 to 0.00. The discrimination power is said to be good, moderate or revised and rejected according to the relationship between discrimination power and the quality of the items in Table 5.

Items that have good discrimination power are items that are able to distinguish between high-ability test participants and low-ability test participants. The data from the analysis results above show that around 70% of the items already have good discrimination power, while the other 30% do not have good discrimination power and need to be revised.

#### Distractor Effectiveness Analysis

The effectiveness of the test item distractor is how the ability of the question distractor functions to deceive students who are less able to choose the answer alternative. Writing multiple-choice questions must have distractor effectiveness. Distractor effectiveness analysis or answer pattern analysis is carried out by calculating the test participants who choose each answer alternative for each item. The criteria for a good distractor is if the distractor is chosen by at least 5% of the test participants (Uno, H. B., & Koni, 2018). To find the value of the distractor quality, the equation is used:

$$IPc = \frac{nPC}{N - nB/Alt - 1} \times 100\%$$

Description:

IPc = Distractor Index,  
 nPc = Number of students who chose the distractor, N = Total number of subjects who took the test,  
 nB = Number of subjects who answered correctly on the item. (Wardoyo & Suprptono, 2014).

According to the Ministry of Education and Culture (1997) in (Amaliya, 2013) to assess the distractors of each question item can be categorized as in Table 6.

**Tabel 6**  
 Kategori Pengecoh (Distraktor)

Kategori Distraktor	Nilai <i>Proportion Endorsing</i>
>0,025	Baik
< 0,025	Revisi
0.000	Tidak Baik/tolak

The distractor effectiveness test aims to determine how well the answer choices or options can deceive test takers who do not know the answer key available on each question item used. The more test takers choose the distractor, the better the distractor can function. From the data (Table 3) the results of the distractor effectiveness analysis on each question item are known that in numbers 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 are questions that have good distractor options. While in numbers 1 and 6 are questions that have less good distractor options. Distractor option number 1 in options C and D needs to have its statement revised. The distractor category value shows <0.025. In number 6 in option B the statement is very bad and must be removed. The distractor category value is 0.000. The distractor options are said to be good, revised and not good or rejected because they are adjusted to the distractor category in table 6. The distractor option that has a good distractor effectiveness value is the distractor option that if chosen by approximately 5% of the total respondents who took the test and this means that this distractor option is very good for testing student learning outcomes. Meanwhile, the question item option that has a moderate or not good distractor effectiveness value is the distractor option that has a small distractor index value (chosen <2% of the total test participants). This means that this distractor option needs to have its statement revised and if necessary discarded and should not be used to test student learning outcomes. From the analysis data above, question item number 1 in options C and D has a

small distractor index of 0.016 (chosen by only around 1.6% of the total participants who took the test) so this means that the distractor option is not good and its statement needs to be revised. The distractor option at number 6 in option B is a distractor option that is not good to use to test student learning outcomes, this is because no students are interested in choosing this option. This data shows that the distractor option is very striking in its statement so that students are not easily fooled by the statement, so the statement in option B is not good to use. While the distractor options on the other questions are good to use. From this data, it can be concluded that 96% of the total distractor options on the questions are good and the rest are not good and need to be revised.

### Conclusion

From the results of the analysis of the criteria for the School Exam questions used by SMP N 2 Ngoro in the 2022/2023 Academic Year, it is known that the results of the analysis show that the level of difficulty of the questions shows that the questions that are classified as difficult are around 15% of the total exam questions, these questions are in numbers 7, 14, 19. Questions that are classified as moderate are 60% of the total exam questions, these questions are in numbers 2, 3, 4, 8, 9, 10, 11, 12, 15, 17, 18, and 20. These questions are not too difficult and also not too easy. Questions that are classified as easy are 25% of the total exam questions, these questions are in numbers 1, 5, 6, 13, 16. Based on this data, it can be concluded that the criteria for the level of difficulty of the exam questions do not yet have a balance between the percentage of the number of questions that are classified as easy, moderate, and difficult. If the question is to be used again, then it is necessary to make improvements to the question items whose level of difficulty does not match the theory.

Data from the results of the discrimination power analysis shows that around 70% of the questions have good discrimination power while the other 30% do not have good discrimination power. The questions that do

not have good discrimination power are found in numbers 1, 4, 5, 14, 16 and 19. Therefore, these questions need to be revised. For questions with very low discrimination power, they must be discarded. The data from the results of the distractor effectiveness analysis on each question shows that around 96% of the total distractor options on the question items are good and the rest are not good and need to be revised. These bad options are found in number 1 options C and D and also in number 6 in option B.

## References

- Adiputra, I. B. R. (2012). Analisis Butir Soal Tes Ulangan Akhir Semester IPS Terpadu Buatan Mgmp IPS Kabupaten Gianyar Kelas VII Semester 1 Tahun Pelajaran 2011-2012. *Penelitian Dan Evaluasi Pendidikan Indonesia*, 2(1), 1–17.
- Afriani, R., Islami, Z. R., & Mustika, F. (2018). Pembinaan Pembuatan Tes Buatan Guru (Soal) Mata Pelajaran Geografi Sma/Ma Kota Langsa. *Jurnal Vokasi - Politeknik Negeri Lhokseumawe*, 1(2).  
<https://doi.org/10.30811/vokasi.v1i2.687>
- Amaliya, N. (2013). Hubungan Antara Motivasi Belajar Dengan Prestasi Belajar Siswa Kelas X Mata Pelajaran Melakukan Persiapan Pengolahan Di Smk Negeri 6 Yogyakarta [Universitas Negeri Yogyakarta].  
<http://eprints.uny.ac.id/id/eprint/13173>.
- Bagiyono. (2017). Analisis Tingkat Kesukaran dan Daya Pembeda Sial Ujian Pelatihan Radiografi Tingkat 1 The Analysis of Difficulty Level and Discrimination Power of Test Items of Radiography Level 1 Examination *Bagiyono 16(No. 1)*, 1–12. [http://repo-](http://repo-nkm.batan.go.id/140/1/05_analisis_tingkat_ke_sukaran.pdf)
- [journalng.uwks.ac.id/index.php/jnsl](https://journalng.uwks.ac.id/index.php/jnsl)  
<https://doi.org/10.32884/ideas.v5i4.220>
- Fitrianawati, M. (2017). Peran analisis butir soal guna meningkatkan kualitas butir soal, kompetensi guru dan hasil belajar peserta didik. *Seminar Nasional Pendidikan PGSD UMS & HDPGSDI Wilayah Jawa*, 282–295.
- Kadir, A. (2015). Menyusun Dan Menganalisis Tes Hasil Belajar Abdul Kadir. *Al-Ta'dib*, 8(2), 70–81.
- Kurniawan, T. (2015). Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Ips Sekolah Dasar. *Journal of Elementary Education*  
[Http://Journal.Unnes.Ac.Id/Sju/Index.Php/Jee](http://Journal.Unnes.Ac.Id/Sju/Index.Php/Jee)  
ANALISIS, 4(1), 1–6.  
<http://journal.unnes.ac.id/sju/index.php/jee>
- Laka, B. M., & Tuasikal, P. (2019). Efektivitas Penggunaan Media Gambar dalam Meningkatkan Keaktifan Siswa pada Pembelajaran IPS. *Ideas*, 5(4), 363–370.  
<https://doi.org/10.32884/ideas.v5i4.220>
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, XIV(1), 85–94.
- Septiana, N. (2016). Analisis Butir Soal Ulangan Akhir Semester (Uas) Biologi Tahun Pelajaran 2015/2016 Kelas X dan XI Pada MAN Sampit. *EduSains*, 4(2), 115–121.
- Sudijono, A. (2001). Pengantar Evaluasi Pendidikan. *Rajawali Press*.

- Susanto, H., Rinaldi, A., & Novalia. (2015). Analisis Validitas Reabilitas Tingkat Kesukaran dan Daya Beda pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika. *Al-Jabar: Jurnal Pendidikan Matematika*, 6(2), 203–217.
- Uno, H. B., & Koni, S. (2018). Assessment Pembelajaran. *PT Bumi*.
- Wardoyo, W., & Suprpto, E. (2014). Rancang Bangun Program Analisis Butir Soal Pilihan Ganda Sebagai Pendukung Proses Evaluasi Pembelajaran. *Jurnal Teknik Elektro*, 6(2).  
<https://doi.org/10.15294/jte.v6i2.3589>
- Widoyoko, E. P. (2018). Penilaian Hasil Pembelajaran Di Sekolah. *Pustaka Pelajar*.