

**ANALYSIS OF DESCRIPTIVE ITEM TEST FOR THE MID-TERM EVEN SEMESTER
SUMMATIVE OF THE MERDEKA CURRICULUM FOR EIGHT-GRADERS SCIENCE
SUBJECT AT SMP NEGERI 20 LEBONG ACADEMIC YEAR 2024/2025**

**Asmi Astuti¹, Dharmawati Ambarita¹, Jeli Dahliah¹, M. Thoha¹, Kashardi¹,
Merri Sri Hartati¹, Reny Dwi Riastuti²**

¹ Universitas Muhammadiyah Bengkulu, Indonesia

² Universitas PGRI Silampari, Indonesia

email: asmiaastuti83@gmail.com, dharmam.ambarita@gmail.com,
jelidahliah@gmail.com, mthoha10@gmail.com, kashardi@umb.ac.id,
renydwirastuti@unpari.ac.id, merrisrihartati@umb.ac.id,

Abstract: This study aims to analyse the quality of the descriptive items used in the mid-semester summative assessment (even semester) of the *Ilmu Pengetahuan Alam* (Science) subject for eighth-grade students at SMP Negeri 20 Lebong in the 2024/2025 academic year. This analysis is important to ensure that the test items used in the learning evaluation not only align with the learning objectives but also effectively assess students' critical thinking skills, conceptual understanding, and scientific abilities. The research employed a descriptive quantitative design with an evaluative approach. Data were obtained through documentation of descriptive items and students' answer sheets, which were then analysed using test item quality indicators, including content validity, reliability, difficulty level, and discriminating power. Data were collected through a descriptive test consisting of 10 items. The data source consisted of 37 answer sheets from eighth-grade students of SMP Negeri 20 Lebong. The findings indicate that six descriptive questions (60%) are valid, while four questions (40%) are invalid. The descriptive questions are categorised at an adequate level of dependency with a consistency score of 0.443. The results of very good, good, quite good, and bad are indicated by the discrimination power of descriptive questions. There are three different categories for the level of difficulty of descriptive questions: moderate, easy, and too easy. The study suggests that "very good" questions could be reused, while "bad or invalid" questions could be revised from scratch based on their indicators. On the other hand, most of the students' learning outcomes meet the KKTP or the Learning Target Achievement Criteria.

Keywords: *Item Analysis, Summative Assessment, Descriptive Test, Merdeka Curriculum, Science, Junior High School.*

INTRODUCTION

Education is one of the important factors in forming superior, creative, and adaptive human resources for the times. To improve the quality of education, the Indonesian government has gradually implemented the Merdeka Curriculum as a replacement for the 2013 Curriculum. The Merdeka Curriculum emphasises student-

centred learning, strengthening basic competencies, and character development through the Pancasila Student Profile. In this context, assessments not only serve as a tool to measure learning outcomes achievement but also as an integral part of the learning process that supports the development of students' competencies as a whole.

The community must prioritise education. Teachers' presence in classes, attending training activities, and conducting research are valuable ways to educate future generations, besides contributing to the nation. Improving educational standards to adapt to changing times and meet the demands of current students has been carried out for many years. Improving education needs to be done by paying attention to many parts of its composition, including qualified educators, curriculum resources, learning facilities, and new approaches to teaching and learning (Sari et al., 2021). Improving the quality of the evaluation or assessment system is another way to improve the learning process and the quality of education (Zainal, 2020). Learning outcomes were evaluated using test and non-test methods. Questionnaires, interviews, and observation methods are examples of non-test assessments. Students are required to answer questions in a variety of formats, including multiple-choice, short-answer, and descriptive, as part of the assessment process. Applying re-exams every day is one of the evaluation strategies that teachers often use to monitor the achievement of learning goals. Teachers provide daily assessments to measure whether students have met learning objectives after completing one basic competency (KD) (Depdiknas, 2013).

Providing students more opportunities to investigate ideas and hone their skills, the Merdeka Curriculum is implemented as a varied intracurricular learning experience with better content (Redana et al., 2023). To guarantee the achievement, educators must take the lead in creating and delivering a curriculum that meets the needs of their students. Science education in junior high school/MTs/ as well as other levels of independent study, emphasises students' ability to understand and apply course content through the use of differentiated instruction that considers each student's unique background, perspective, and set of experiences (Mahdiannur et al., 2022).

To help students develop a basic understanding of science and its practical applications, science education often takes the form of contextual learning, involving hands-on activities and real-world examples. Therefore, to motivate students actively seeking established scientific principles, facts, and concepts, it is important to offer creative learning activities, original and entertaining (Fitra, 2022). According to Supartama et al. (2023), science education involves teachers' creativity in teaching their students by adapting their approaches to the different characteristics of students and by applying various learning models. Students are encouraged to take an active role in their science education and to cultivate their

unique scientific interests and abilities through the Merdeka Curriculum (Putri et al., 2024) (Maulidia, 2025)

Optimising learning outcomes based on student capacity is the main principle of the Merdeka Curriculum. Thus, there must be a learning design that prioritises student needs. The beginning and end of the learning process can include this evaluation. The first step in developing an individual curriculum is to provide learning evaluations to each student. Most importantly, the Merdeka Curriculum uses a variety of evaluation strategies. Assessment is an ongoing and methodical way to collect data on how and what students have learned. By doing so, the decisions can be made according to certain standards (Nasution, 2022).

It is believed that the focus of evaluation in the Merdeka Curriculum will change when compared to the previous curriculum. The previous course gave more value to summative evaluation. Student study reports were based on summative exam results. In the new assessment paradigm, educators have more flexibility to prioritize formative evaluation over summative evaluation. To improve future learning, it would be helpful if formative assessment findings were used as a basis (Kurka, 2022a) (Nur Budiono & Hatip, 2023).

The main purpose of assessment (evaluation) is to find out how much, how well, and to what extent educational goals have been met (Ralph Tyler, 2013). Assessment, according to Griffin and Nix, is a process of finding out how useful a program is, and that includes educational programs. These two caveats indicate that evaluation involves some kind of decision making. Plans, actions, and rewards are all part of the decision. In accordance with Fernandes' definition of testing as a systematic technique for characterizing individual behavior in numerical or categorical terms, we find that evaluation is associated with testing. When used as a verb, "test" indicates the act of testing, and when used as a noun, it indicates examination (Anita Yus, 2011:39-40). A more inclusive explanation is proposed by two other authorities, especially Cronbach and Stufflebeam. (Riinawati, 2021)

To effectively evaluate student progress in the classroom, teachers must master several key areas of educational assessment. These areas include: (1) knowing how to select appropriate assessment tools to use in making decisions about learning; (2) creating appropriate assessment tools to use in making decisions about learning; (3) administering, scoring, and interpreting test results; (4) making appropriate decisions based on assessment data; (5) creating valid assessment tools and using assessment information; and (6) communicating assessment findings effectively.

Minister of Education and Culture Regulation Number 66 of 2013 regulates the basic concept of assessing student learning outcomes at elementary and secondary education levels, namely: 1. objective, in the sense that it does not depend on or does not take into account the assessor's views or values; 2. integrated, in the sense

that the assessment is carried out by teachers methodically, coordinated, and continuously with the learning implementation plan; 3. cost-effective, in the sense that the assessment utilises resources optimally from the planning, implementation, and reporting stages; 4. open, so that all parties involved can see the assessment process, assessment criteria, and the reasons behind the conclusions; 5. accountable, in the sense that the technical components, processes, and results of the assessment can be explained to internal and external school stakeholders; 6. pedagogical, in the sense of providing lessons and inspiration to educators and students.

A specific protocol for evaluating educational progress must be in place. The following steps are required to follow the steps outlined by Uno and Satria: (a) converting core skills into learning outcome indicators; (b) setting endpoints for each indicator; (c) mapping standards for competencies, core skills, indicators, endpoints, and report card components; (d) mapping standards for competencies, core skills, indicators, endpoints, assessment components, and assessment methods; and (e) selecting evaluation tools according to the nature of the indicators (Mumtahanah, 2018).

At the end of each program unit, whether it is a quarter, a semester, or a full academic year, students take a summative assessment to measure their progress and identify areas for improvement. This assessment is used to determine students' report card grades and whether they will be promoted to the next grade.

The purpose of summative assessment is to evaluate students' progress toward learning outcomes through the use of numerical scales and ratings. At the end of the school year, students' awards are based on their performance on summative assessments (Kusairi, 2013). (Otaya et al., 2021)

According to Wahyuni and Ibrahim (2012:02), both evaluating competence and providing a positive influence on the learning process are prerequisites for assessment. Instructors may confuse the terms "test", "measurement", and "evaluation" when implementing learning assessment. Evaluation is different from a test. The purpose of giving a test is to evaluate students' final competencies after they have completed all the learning activities outlined in the curriculum. The wider activities are encompassed by the implementation process of an ongoing assessment. Furthermore, the concept of measurement also exists. The purpose of measurement is to collect numerical data on how well or poorly a person performs in a particular field of study (Munaroh, 2024).

Since both produce data that illustrates how well the learning process is going, evaluation and assessment are intrinsic to every learning process. As the most influential guide for students' educational journeys, teacher effectiveness has a direct impact on the quality of teaching in the classroom. Teachers must dedicate themselves to their careers if they want their students to learn well in class (Laka &

Tuasikal, 2019). It is common practice for educators to conduct assessment and evaluation tasks to measure the extent to which their students have mastered the subject matter. According to (Rahayu & Djazari, 2016), the purpose of evaluation is to collect evidence that can be used to determine how well students have learned the material. The second purpose of evaluation is to determine how well teachers have achieved their pedagogical goals. In addition to providing recommendations for reporting student learning progress, (Kurniawan, 2015) said that the purpose of evaluating learning outcomes is to determine whether students have acquired the necessary competencies or not and to improve the overall learning process (Bano, Marambaawang, & Njoeroemana, 2022).

In determining the quality of education, teachers play an important role. Improving the quality of national education is the responsibility of teachers as learning agents (Amrullah et al., 2021). Teachers must also take the time to assess their students' progress in learning. According to Phafiandita et al. (2022), the purpose of the assessment is to determine how well the teaching techniques are running over a certain time. According to Nafs et al. (2023), evaluation is a learning stage when teachers collect information as feedback to improve the quality of learning and student teaching methods.

The teacher's ability to deliver quality knowledge is demonstrated through the results of learning evaluations. Therefore, educators must have good abilities in all aspects of implementing learning evaluations, including but not limited to: planning, preparing instruments, analysis, and interpretation. One of the abilities that educators must have is the ability to assess something (Jumini et al., 2023; Kasmayanti et al., 2023).

To measure the extent to which students have internalised the previously taught material, educators often give tests. In order for instructors to make accurate assessments of student learning, tests provide a systematic, comprehensive, and objective means of evaluation (Marsiyah, 2016). Instructors can measure the level of student understanding of the subject matter through the use of test questions, which are instruments (Anita, 2018).

Although related, measurement, assessment, and evaluation are not the same thing when it comes to schooling. (Arthur, 2022) Measuring student performance or achievement produces numerical or statistical data (Prastiwi et al., 2023). Objective data can be obtained from measurements, which can then be used for assessment and evaluation purposes. Educational assessment can be done in several ways. Written tests that evaluate students' skills through descriptives, multiple choice, or other written forms of examination are common components of traditional assessment (Gronlund & Linn, 1990). The cognitive component of learning is the focus of this method. On the other hand, authentic assessment emphasises tasks that demonstrate how students' knowledge and skills are used in real-world

scenarios (Wiggins, 1993). Projects, presentations, portfolios, and other assignments that test analytical and problem-solving skills are part of this evaluation (Arta, 2024).

Final exams or daily assessments can use several types of test tools to measure students' cognitive skills (Marsiyah, 2016). After each chapter, students take a test to see how well they have learned the material and whether they have mastered one of the core competencies (Depdiknas, 2013). The questions and answers presented in writing are written exams (Virginia, Angraini, Pratesya, & Walid, 2021).

The purpose of summative assessment is to measure the extent to which students have mastered the subject matter. Each topic is represented in this exam, which assesses the overall learning achievement of students. The content assessed covers all disciplines and learning objectives in a year or semester program.

Valid and reliable assessment tools are needed to measure student competency. A well-designed exam can serve as a reliable standard for evaluating student performance. Exam quality analysis is conducted to determine exam criteria. The state of the exam questions can be known through exam quality analysis. The four qualities are as follows: validity, dependability, objectivity, and practicality (Surapranata, 2004) (Shofiyah & Sartika, 2018).

Teacher-made tests are tests created by instructors to measure how well their lessons are being received by students. In most schools, teacher-developed assessments are widely used. Class or school restrictions are usually imposed when administering these teacher-made tests (Harjanti, 2006). It is important for teachers to carefully consider each item of the test when creating test questions. These questions should measure the following: the ability to remember, understand, or think critically about the topic being discussed; the clarity of the sentences and language used; and finally, whether the questions are aligned with the competencies that have been taught (Putriani, Turahmah, Sunarti, Ismarliana, & Walid, 2020).

To provide a more targeted implementation, evaluation tasks require assessment tools or methodologies. Tests and non-tests are both used as evaluation techniques in education. A test is a method of evaluating students' understanding and performance in a particular subject by asking them to complete a task or series of tasks designed to measure specific skills or knowledge. Objective and subjective test formats are both valid choices. When all the information needed to complete the test is readily available, we say that the test is objective. In subjective tests, respondents are asked to provide lengthy explanations in response to a series of questions or directions. Accountability in the two test formats used here means that, judged by the quality of the questions, the test can serve as an adequate evaluation instrument (Kelas, Xi, Man, & Palangkaraya, 2016).

The answers on the descriptive test can be free or limited in scope according to the test taker's wishes. Descriptive questions on the free descriptive exam are designed to assess students' critical thinking skills by asking them to write, organize, and structure their own answers in their own words. Students' ability to explain procedures, present relevant explanations, formulate hypotheses, draw acceptable conclusions, apply principles or theories, and so on can be adequately measured by limited descriptive tests. (Rahman & Nasryah, 2019).

Teachers often encounter the idea of evaluation, measurement, and testing at the same time when they carry out the learning assessment process. As a result, in reality, it is quite rare for people not to distinguish between the three, because conducting an evaluation requires doing all three. Teachers must develop assessment tools, including tests, exam questions, observations of the learning process, and other non-tests, to conduct assessments. Measurement requires quantifying some aspect of the learning process or student work as an indicator of their level of mastery of the material; then, these metrics are compared to predetermined standards, such as minimum mastery limits or group abilities, to produce scores that represent the quality of the learning process and results. Instructors then make decisions about the procedure and its quality in relation to learning outcomes (Poerwanti, 2015).

Therefore, to ensure the value of the material learned, learning evaluation must include some type of measurement or assessment. (Study conducted by Magdalena and colleagues in 2023) From a qualitative perspective, learning and assessment of learning is about finding out how much something is worth. From a quantitative perspective, measurement in learning activities is about comparing how much something is worth with what has been determined. "Febriana" (2021) said Students can learn a lot about their academic performance through assessment. Students will be motivated to continue to push themselves academically when they see positive results in their grades. If the results are not satisfactory, students will try to improve their learning process. However, it is important for educators to produce encouraging signals to keep students engaged. This is according to Sukma (2022) (Nur Aidila Fitria, Muhammad Yoga Julyanur, & Eka Widyanti, 2024).

In this situation, the purpose of the test is to determine how effective the previous learning program was. The role of the test in the learning process is very important, so it is very important for instructors to design high-quality tests. The characteristics of a good test are its ability to differentiate student skills, its ability to identify students who are learning and those who are not, and its consistency in results between administrations. In addition, when creating a test, it is necessary to consider a number of criteria. A well-designed test that measures students' talents and skills according to learning objectives is one of the requirements for a good

test, according to Sopiah, et al. (2019). This criterion is known as validity, and is met if the measuring instrument can provide reliable results for the target variable (Juanta, Sijabat, Festiyed, & ..., 2023).

Although item analysis is important, some tests only examine the content validity of items, rather than evaluating them for testing and item analysis, which takes into account criterion validity, reliability, difficulty level, discriminating power, and distractor effectiveness. According to Agustin's (2019) research, teachers never use item analysis. Teachers still rely heavily on textbooks and questions from previous years' tests, the quality of which is unknown, when constructing questions. The assumption made by instructors is that they have met the requirements for effective test questions if they limit themselves to creating questions from their own question banks and syllabi (Differentiation & Difficulty, 2024).

Field testing of a question or test involves item analysis, according to (Maimun, 2011). The fact that it exists is evidence that the question is of high quality. Feedback is needed for test makers to check for errors in questions after determining that the question has weak discrimination power, is too difficult or too easy, and has negative interference (Mahfudhah, 2017). There is a qualitative approach to item analysis that considers its statistical features (Susanty, 2016) (Azianto, 2022).

To find out how much students have learned at the end of a certain period of time (e.g., mid-term or end of year), teachers use summative assessments. One form of summative assessment that is commonly used is descriptive tests, which are considered effective in measuring higher-order thinking skills, conceptual understanding, and skills in applying knowledge. Therefore, the quality of descriptive test items greatly determines the effectiveness of the assessment in describing students' abilities objectively.

However, in practice, various problems are still found in the preparation and implementation of descriptive tests, such as questions that are not in accordance with learning outcomes, do not have good differentiating power, are too difficult or too easy, and do not reflect the assessment principles applied in The Merdeka Curriculum. This condition can cause the assessment results to be invalid and unreliable, so that they cannot be used as a basis for making accurate learning decisions.

SMP Negeri 20 Lebong, as one of the educational units that has implemented The Merdeka Curriculum, needs to evaluate the assessment instruments used, including descriptive test questions in the mid-summative assessment of even semesters. This evaluation is important for two reasons: first, to ensure that all questions are up to standard, and second, to support lessons that aim to build student competence and character.

Based on this background, this study intends to analyse the descriptive test items of the mid-term summative assessment of the even semester of the VIII grade science subject at SMP Negeri 20 Lebong in the 2024/2025 Academic Year. The focus of the analysis includes content validity, reliability, level of difficulty, and discrimination power. It is hoped that the results of this study can provide a positive contribution to the development and improvement of learning assessments in schools.

RESEARCH METHOD

This study applies a descriptive approach. Quantitative techniques are used in this study because the questions are analysed systematically, planned, and organised, resulting in findings. According to Sudijono (2018), numerical data is generated through statistical calculations in this study. On April 26, 2025, researchers from SMP Negeri 20 Lebong conducted a study. Eighth-grade students became the sample of this study. Ten descriptive questions included in the mid-semester summative exam for the even semester science course were the subjects of this study. The materials used in this study include student responses, answer keys, and mid-semester summative evaluation questions from the even semester science course. Quantitative analysis is the right way to solve mid-semester summative exam questions for even-semester science classes. Classical methods are often used to test quantitative analysis, considering its discrimination power, level of complexity, dependency, and validity. The collected data were then examined with the help of SPSS 25.0 for Windows.

Validity

Content validity is a method used to evaluate the efficacy of learning outcome assessment. If r count is greater than or equal to r table, then the research is considered valid (Rahayu & Djazari, 2016).

Reliability

The dependency coefficient is a numerical indicator of the level of reliability; its value ranges from 0 to 1, with higher values indicating more consistent measurement findings (Ida & Musyarofah, 2021).

Tabel 2. Reliability Criteria

Correlation Index	Test item quality
$r_{11} < 0.20$	Very low
$0.20 \leq r_{11} < 0.40$	Low
$0.40 \leq r_{11} < 0.70$	Moderate
$0.70 \leq r_{11} < 0.90$	High
$0.90 \leq r_{11} < 1.00$	Very high

(Azis, 2016)

Level of Difficulty

One way to describe the level of difficulty of a question can be seen from the index, which describes the probability that a student with a certain level of talent will answer the question correctly.

Table 3. Difficulty Index

Level of Difficulty	Interpretation
$IK = 0,00$	Too difficult
$0,00 < IK \leq 0,30$	Difficult
$0,30 < IK \leq 0,70$	Moderate
$0,70 < IK \leq 1,00$	Easy
$IK = 1,00$	Too easy

(Subana & Sudrajat, 2011)

Discrimination Power

After the learning process is complete, the discrimination power is used to determine which students have mastered the material competencies and which students have not (Kurniawan, 2015).

Table 4. Classification of Discrimination Power

	Interpretasi
$0,40 \leq D \leq 1,0$	Very good
$0,3 D < 0,4$	Good
$0,2 D < 0,3$	Quite good
$D < 0,2$	Bad

(Fiska et al., 2021)

FINDINGS AND DISCUSSION

The research data on the analysis of the quality of mid-semester summative assessment questions for the Merdeka Curriculum for grade VIII science subjects at SMP Negeri 20 Lebong are as follows.

1. Validity

The results of data analysis on the validity aspect of descriptive question items can be seen in Table 5 below.

Table 5. Validity of Descriptive Question Items

No	Validity	Item test number	Total	Percentage
1	Valid	2, 3, 5, 8, 9, 10	6	60%
2	Invalid	1, 4, 6, 7	4	40%

The results of the analysis in Table. 5 show that 6 descriptive questions (60%) are declared valid with a calculated r value of $r \geq 0.3$, meaning that they are in accordance with the learning indicators and measure the competencies designed in the Merdeka Curriculum Learning Outcomes (CP). In the questions that were declared invalid, there were 4 descriptive questions (40%) with a calculated r value of ≤ 0.3 , indicating that there were still a number of questions that did not match the indicators or learning objectives, or did not measure essential competencies.

By comparing the number of valid questions with the number of invalid questions, we can conclude that the questions produced have achieved validity with a fairly high level of quality. Questions are considered valid because they cover topics that reflect the intended assessment objectives and produce findings that are equivalent to their functions and objectives. Valid questions can be stored in a question bank for future use (Fiska et al., 2021), while invalid questions need to be edited using the question compilation approach and achievement indicators. Question validity has an impact on student scores (Nurhasanah & Ahmad, 2017). Before working on the test questions, students' work needs to be assessed to ensure that their results meet the Learning Objective Achievement Criteria (KKTP). The use of questions that are considered invalid will prevent students from measuring the intended construct, resulting in findings that do not reflect students' actual skills (Fiska et al., 2021). This is in accordance with the idea put forward by Grounlund and Arifin (2017) which states that the validity of test results can be influenced by three elements: the instrument used, the assessment standard, and factors derived from students' answers.

2. Reliability

Table 6 illustrates the findings of the data analysis on the dependency aspect of descriptive questions as shown below:

Table 6. Reliability of Descriptive Question Items

No	Reliability	Coefficient	Type test
1	Reliable	0,443	Descriptive

The reliability value of 0.443 indicates that the descriptive questions are reliable with sufficient consistency, as shown in Table 6. Based on research conducted by Rahmasari and Ismiyati (2016), if the item dependability rating is less than 0.70, then its reliability will be considered poor and must be updated. There are many variables that can affect the reliability value, such as how the test is administered, how many students take it, and how challenging the questions are. Some components of test implementation include student unpreparedness and

inadequate direction to complete it. This will affect the dependability value if certain sections are not supportive. Lower dependability ratings and more diverse answers are the result of larger classes. The dependability value decreases as the level of difficulty increases because students are more likely to guess or simply answer questions that are too difficult (Anita et al., 2018).

3. Level of difficulty

Table 7 below shows the results of the analysis of the difficulty level of descriptive questions.

No	Item Category	Item Number	Total	Percentage
1	0,30 < IK ≤ 0,70 Moderate	3	1	10%
2	0,70 < IK ≤ 1,00 Easy	1, 2, 5, 7, 8, 9, 10	7	70%
3	IK = 1,00 Too easy	4, 6	2	20%

Table 7 shows how challenging descriptive questions are. Question 3 covers 10% of the total and is categorised as in the medium group. To encourage students to use question-answering skills, choose questions in the medium category. If the difficulty level of the question is between 0.30 and 0.70, then the question is good (Arikunto, 2013). With seven questions (or 70%) covering numbers 1, 2, 5, 7, 8, 9, and 10, it can be said that the questions are well-crafted and of high quality. The issues discussed in this discussion relate to topics such as vibrations, waves, light, elements, compounds, and mixtures; students are encouraged to provide their answers. There are only two questions (20%) on numbers 4 and 6, which makes the category very easy.

To get good questions, the questions should fall into the moderate group, meaning they are not too easy or too difficult. The question bank is a good place to store questions that fall into the moderate category. This is in line with research (Ningrum, Rahmawati, Minarti, & Mekar, 2023) that the right questions must have a balance between being too simple and too complicated. Two items (10%), numbers 4 and 6, from the descriptive items, produced data with a level of difficulty that was too simple. Questions that are too simple and too challenging have the potential to dampen students' enthusiasm for learning (Virginia et al., 2021). To ensure that the exam accurately measures students' progress or development after the learning process, it is necessary to adjust the complexity of each question to their developmental stage (Fiska et al., 2021).

4. Discrimination Power

Table 8 below displays the results of the data analysis carried out on the discriminating power of descriptive question items.

Table 8. 4. Discrimination Power of Descriptive Question Items

No	Discrimination Power	Item Number	Total	Percentage
1	$0,40 \leq D \leq 1,0$ Very good	3	1	10%
2	$0,3 D < 0,4$ Good	1, 2, 7	3	30%
3	$0,2 D < 0,3$ Quite good	5, 8, 9	3	30%
4	$D < 0,2$ Bad	4, 6, 10	3	30%

The discriminatory power of descriptive questions is broken down into several different categories in Table 8. One question (10%) falls into the excellent category, three questions (30%) fall into the good category, two questions (30%) fall into the fair category, three questions (30%) fall into the fair category, and three questions (30%) fall into the poor category. One way to differentiate between high and low achievers on an achievement test is to consider the discriminating power of the questions asked (Rahmasari & Ismiyati, 2016). Thus, the questions on the test can identify whether students have learned the information or not. This is in line with research (Ningrum, Rahmawati, Minarti, & Mekar, 2023) that there is a correlation between the level of difficulty and the discriminating power of a question. Questions with a low level of difficulty will have few correct answers, while questions with a high level of difficulty will have many correct answers, thus eliminating any discriminating power. Questions with low discriminating power need to be edited and added to the question bank if their discriminating power is very good, good, or quite good.

CONCLUSION

Based on the findings of the study and discussion regarding the quality of the analysis of the descriptive questions of the mid-term summative assessment of the even semester of The Merdeka Curriculum for the subject of Science for class VIII at SMP Negeri 20 Lebong in the 2024/2025 Academic Year, it can be concluded that the questions are of valid quality, because the proportion of things that are considered valid exceeds the percentage of items that are declared invalid. There is sufficient consistency in the descriptive questions to meet the reliability standards. It can be said that the discriminatory power of the questions varies, but in general it is very good, because the dominant category is good and the non-dominant category is

weak. Because the reference value is somewhere in the "Moderate" group (between 0.30 and 0.70), the questions are of fairly high quality, which contributes to their relatively simple level of difficulty. The findings of the study indicate that valid questions can be reused, while objects that are less good or invalid can be improved or made new based on their signs. However, overall, most of the students' learning outcomes have met the KKTP value for the Learning Target Achievement Criteria.

ACKNOWLEDGEMENT

The authors would like to express their deepest gratitude to:

1. The Principal of SMP Negeri 20 Lebong who gave permission and full support in carrying out this research.
2. All students of class VIII of SMP Negeri 20 Lebong who became the respondents in this research.
3. Lecturers of the Master of Pedagogy, Faculty of Teacher Training and Education, University of Muhammadiyah Bengkulu, who provided a lot of guidance and knowledge in compiling this article.
4. Colleagues who provided input, criticism, and constructive suggestions for the improvement of this article.

REFERENCES

- Arta, G. Y. (2024). Asesmen dalam Pendidikan: Konsep, Pendekatan, Prinsip, Jenis, dan Fungsi. *Jurnal Pendidikan, Bahasa Dan Budaya*, 3(3), 170–190.
- Azianto, A. (2022). Analisis Butir Soal Tes Uraian pada Mata Pelajaran SKI Kelas IX MTsS TI Candung Tahun Ajaran 2020/2021. *Didaktika: Jurnal Kependidikan*, 11(1), 1–8. <https://doi.org/10.58230/27454312.113>
- Bano, V. O., Marambaawang, D. N., & Njoeroemana, Y. (2022). Analisis Kriteria Butir Soal Ujian Sekolah Mata Pelajaran IPA di SMP Negeri 1 Waingapu. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, 8(1), 145. <https://doi.org/10.32884/ideas.v8i1.660>
- Juanta, P., Sijabat, A., Festiyed, F., & ... (2023). VALIDITAS BUTIR SOAL IPA FISIKA SMP VII Dr WAHIDIN SUDIROHUSODO. *Jurnal Tunas ...*, 6(1), 262–267. Retrieved from <https://ejournal.ummuba.ac.id/index.php/pgsd/article/view/1253%0Ahttps://ejournal.ummuba.ac.id/index.php/pgsd/article/download/1253/843>
- Kelas, P., Xi, X. D. A. N., Man, P., & Palangkaraya, I. (2016). 514-1179-1-Sm, 4(20), 115–121.
- Maulidia, D. (2025). JURNAL PENDIDIKAN DAN PEMBELAJARAN SAINS INDONESIA Analisis Pelaksanaan Pembelajaran IPA Berdasarkan Kurikulum Merdeka Kelas VIII SMP Negeri 1 Seririt, 8(April), 38–49.
- Mumtahanah, N. (2018). Prosedur Pengembangan Dan Karakteristik Assesment Pembelajaran. *Al-Hikmah*, 8(19), 1–9.
- Munaroh, L. N. (2024). Asesmen dalam Pendidikan : Memahami Konsep,Fungsi dan Penerapannya. *Jurnal Pendidikan Sosial Humaniora*, 3(3), 281–297.
- Ningrum, W. A., Rahmawati, R. C., Minarti, I. B., & Mekar, R. M. (2023). Analisis Butir

- Soal Ulangan Harian Pembelajaran IPA Pada Kelas VIII Di SMPN 21 Semarang. *Educatio*, 18(1), 91–101. <https://doi.org/10.29408/edc.v18i1.18291>
- Nur Aidila Fitria, Muhammad Yoga Julyanur, & Eka Widyanti. (2024). Langkah-langkah Evaluasi Pembelajaran. *PUSTAKA: Jurnal Bahasa Dan Pendidikan*, 4(3), 285–294. <https://doi.org/10.56910/pustaka.v4i3.1572>
- Nur Budiono, A., & Hatip, M. (2023). Asesmen Pembelajaran Pada Kurikulum Merdeka. *Jurnal Axioma: Jurnal Matematika Dan Pembelajaran*, 8(1), 109–123. <https://doi.org/10.56013/axi.v8i1.2044>
- Otaya, Ilian G., Imam, T., Jayanti, D., Rahmatullah, Wahab, A., Lutfie, M., ... Indra P, I. M. (2021). *Evaluasi Pembelajaran*.
- Pembeda, D., & Kesukaran, T. (2024). PROSIDING SEMINAR NASIONAL PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT (SNPP) TAHUN 2024 “ Tantangan dan Peluang Dalam Mengoptimalkan Peran Artificial Intelligence (AI) Untuk Menghasilkan Karya Ilmiah Berkualitas Menuju Masa Depan Riset Indonesia Maju 204, 175–181.
- Poerwanti, E. (2015). Konsep Dasar Asesmen. *PT. Remaja Rosdakarya.*, (1), 1–44.
- Putriani, D., Turahmah, F., Sunarti, N. R., Ismarliana, P., & Walid, A. (2020). Analisis Butir Soal UAS Biologi 2018/2019 Kelas X dan XI SMAN 11 Kota Bengkulu. *Journal of Biology Learning*, 2(1), 1–7. <https://doi.org/10.32585/v2i1.559>
- Rahman, A. A., & Nasryah, C. E. (2019). *Evaluasi Pembelajaran. Uwais Inspirasi Indonesia*.
- Riinawati. (2021). *Sasaran dan objek penilaian. Pengantar Evaluasi Pendidikan*.
- Shofiyah, N., & Sartika, S. B. (2018). *Buku Ajar Asesmen Pembelajaran. Paper Knowledge . Toward a Media History of Documents (Vol. 5)*.
- Virginia, S., Angraini, W., Pratesya, W., & Walid, A. (2021). Analisis Butir Soal Ulangan Harian IPA Terpadu Kelas VII SMP 05 Kota Bengkulu Tahun Pelajaran 2020/2021. *Jurnal PTK Dan Pendidikan*, 6(2), 61–65. <https://doi.org/10.18592/ptk.v6i2.4173>