EDUKASI: JURNAL PENDIDIKAN DAN PENGAJARAN
ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

# ITEM ANALYSIS OF CULTURAL ART SUBJECT: THE QUALITY AND QUANTITY OF MULTIPLE-CHOICE QUESTIONS AND DESCRIPTIONS

**IKA RATNANINGRUM, RHISMA DWI APRILLYA, AND ALYA PUTRI MARADIKA NINGRUM**
Universitas Negeri Semarang, Indonesia
Corresponding author: 3dara@mail.unnes.ac.id

**Abstract**
The purpose of this study is to analyze and describe the quality and quantity of multiple-choice questions and descriptions in the even semester of Cultural Arts subjects for fourth-grade students at SDN Dabin I, South Tegal District, Tegal City. This study uses a research design (mix methods). The research sample comprised 144 students answers sheets. Qualitative data were examined through content validity procedures and alignment with cognitive domains, while quantitative data were analyzed using the Anates V4 software. The results of the study show that the even semester final semester assessment questions for the fourth grade of Cultural Arts subjects for the 2024/2025 school year (multiple choice and essay) are good, but they are still found in multiple-choice questions that need to be improved. It is recommended to the question preparation team to improve the questions and conduct question item analysis activities before the questions are tested to students, in order to produce more quality question items. Contribution or findings in research is to provide information to the makers of end-of-semester assessment questions so that they can improve questions that are not of good quality before being used as an evaluation tool for students.

**Keywords:** cultural arts, essay, multiple choice, question items

## Introduction

Question item analysis activities can be carried out qualitatively and quantitatively. The analysis of question items qualitatively examines questions from the aspects of material, construction, language/culture, as well as answer keys or scoring guidelines. According to the Ministry of National Education (2008), there are two techniques used, namely the moderator technique (discussion with a mediator) and the panel technique (the study of questions based on the rules of writing by several reviewers). In addition, the analysis also includes the suitability of the questions to the cognitive realm in Bloom's taxonomy. The analysis of question items is quantitatively based on empirical data from the results of the question experiment to assess its quality. According to the Ministry of National Education (2008), there are two approaches, namely classical (using classical test theory by paying attention to validity, reliability, level of difficulty, discrimination, and effectiveness of deception) and modern (using Item Response Theory / IRT).

Based on the results of interviews with grade IV teachers of SD Tegal City, information was obtained that final semester assessment questions were made by teachers in each school that had been appointed. It began with a KKKS meeting by discussing the technicalities of preparing even semester final semester assessment questions. The preparation of final semester assessment questions starts from making a grid of questions to compiling final semester assessment Cultural Arts questions that will be tested for all elementary school students in Dabin I, South Tegal District, Tegal City, carried out by teachers at the designated school. Based on the results of the interview, the researcher chose the analysis of the item of the even semester final assessment as

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

the object of research, because the time when the researcher conducted the interview coincided with the even final semester assessment to be held, so that the data obtained by the researcher was new. The researcher chose class IV to be researched, because class IV is a transitional class from the initial class to the high class. Teachers who prepare even final semester assessment questions should be able to analyze the question items first before testing them to students, so that students can work on quality questions.

Some of the problem gaps that have been identified are as follows: The question compiler takes a long time to analyze the question items in more detail, so the question preparation team does not conduct the analysis. Teachers' knowledge still needs to be improved in conducting problem analysis. An analysis of question items has not been carried out to determine the quality of the questions reviewed from the aspects of material, construction, and language in the final semester assessment even semester question items for grade IV. An analysis of question items to determine the quality of the questions has not been carried out from the aspects of validity, reliability, level of difficulty of the questions, discriminating power, and effectiveness of deception in the final semester assessment even semester question items for grade IV. The distribution of cognitive domain levels has not been measured in the even semester final semester assessment questions for grade IV State Elementary School in Dabin I, South Tegal District, Tegal City. The problems in this study are (1) How is the analysis of multiple-choice questions and descriptive questions of Cultural Arts subjects in the even semester of grade IV reviewed from the aspects of validity, reliability, level of difficulty of the questions, discriminating power, and effectiveness of the deception empirically? (2) How is the analysis of multiple-choice question items and descriptions of Cultural Arts subjects in final semester assessment even semester grade IV reviewed from the aspects of material, construction, and language theoretically? (3) How is the distribution of the level of the cognitive domain that is measured in multiple-choice question items and descriptions of Cultural Arts subjects in final semester assessment even semester of grade IV?

The contribution or findings in the research is to provide information to the makers of end-of-semester assessment questions so that they can improve questions that are not of good quality before being used as an evaluation tool for students. Teachers who prepare even final semester assessment questions should be able to analyze the question items first before testing them to students, so that students can work on quality questions. The general purpose of this study is to analyze and describe the quality of Cultural Arts question items in the even semester final semester assessment of grade IV State Elementary School in Dabin I, South Tegal District, Tegal City for the 2024/2025 school year after being analyzed, both qualitatively and quantitatively.

## Literature Review

The purpose of the literature review is to provide an understanding of the theoretical basis and conceptual framework that strengthens the research argument. The things that will be discussed in the literature review in this study are the definition of education, the definition of curriculum, and learning evaluation.

### *Understanding of Education*

Education is very important for every human being. Education is a necessity that cannot be separated from human life. In fact, education can be a benchmark for whether or not a country progresses. This is because education is one of the determining factors in improving a country's human resources (Sanga & Wangdra, 2023). Education as a conscious effort has goals to be achieved (Abd Rahman et al., 2022). The goal provides an idea of how the learning process will

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

occur. Learning objectives refer to the functions and objectives of national education as stated in Law Number 20 of 2003 Chapter II Article 3 concerning the Basics, Functions, and Objectives of Education which says "*National education functions to develop abilities and shape the character and civilization of a dignified nation in order to educate the life of the nation, aiming to develop the potential of students to become human beings who believe and fear God Almighty, have noble character, are healthy, knowledgeable, capable, creative, independent, and become democratic and responsible citizens.*"

The functions and objectives of national education can be realized through the implementation of education with three educational paths, namely formal, non-formal, and informal education that complement and enrich each other (Syaadah et al., 2022). Schools are formal educational institutions that are part of a tiered and continuous education. The implementation of a learning course cannot run without a curriculum.

### *Understanding of curriculum*

The curriculum is a guide or guideline for the implementation of education in order to form students in accordance with the goals of national education (Firmansyah, 2020). Given the importance of a curriculum in the implementation of education, the preparation of the curriculum should not be arbitrary. The preparation of the curriculum must be based on national education goals and deep thinking (Hakim & Darojat, 2023). The curriculum used in education in Indonesia today is the independent curriculum. The independent curriculum aims to create learning that is more based on 21st century competencies, characters, and skills (Ihsan et al., 2025). The independent curriculum introduces formative assessments as an assessment method that assesses the learning process of students and not just looking at the final results (Alindra et al., 2025).

### *Learning assessment*

There are three terms that are often considered the same by most people, namely measurement; assessment; and evaluation. The three terms have different meanings, but they have interrelated relationships in determining the learning success of students (Sholihan et al., 2024). The evaluation process has a broader meaning, which includes measurement and assessment (Asrul et al., 2022). In the independent curriculum, there is an assessment that focuses more on the process, not like an evaluation that only focuses on results (Agustianti et al., 2022). The purpose of the evaluation is to find out the extent of the development of students' knowledge and measure the effectiveness of the learning methods used (Kurniawan et al., 2022). There are two learning evaluation techniques, namely test and non-test techniques (Maulani et al., 2024).

Arikuto (2016) stated that, "A test is a tool used to measure something, in a predetermined way and rules". Another opinion was conveyed by Sudjana (2017) who stated that, "Tests are used to assess and measure students' learning outcomes, especially in cognitive learning outcomes related to mastery of teaching materials in accordance with the teaching objectives given". Based on these opinions, it can be concluded that the test is a measuring tool to determine the achievement of students during the learning process. According to Sudijono (2015), reviewed in terms of function as a tool to measure student development, tests are divided into six groups, namely selection tests, initial tests, final tests, diagnostic tests, formative tests, and summative tests. Tests play a role in accumulating information that describes the characteristics of students' abilities, attitudes, interests, and motivation (Irawan et al., 2023).

Based on the classification of the test, one of the groups of summative tests is the Final Semester Assessment. End-of-Semester Assessments are routinely carried out by teachers and other agencies (Hidayati & Nisa, 2023). The main purpose of the Final Semester Assessment is to

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

determine the grades that symbolize the success of students after taking the learning activity process within one semester. The better the quality of the questions, the more accurate the information received by teachers from the results of the Final Semester Assessment (Muluki et al., 2020). Therefore, it is necessary to have appropriate test instruments (Magdalena et al., 2021). The instrument category is divided into two, namely objective tests and non-objective tests. Multiple-choice, true-wrong, matchmaking, and short-form questions are types of objective tests. Meanwhile, questions that have a long description answer format are non-objective tests (Zamzania & Aristia, 2018). A good and appropriate question instrument must be analyzed first. Analysis of question items can identify good questions, questions that must be revised, and questions that cannot be reused so that the instrument or question can be of high quality (Kumar et al., 2021). A good question is a question that is not too difficult or not too easy, but that has good differentiation (Jumini et al., 2024).

## Methodology

### *Research design and approach of the study*

The research design that has been used in this study is a quantitative descriptive research design, because the data analyzed is not to accept or reject the hypothesis, but the results of the analysis are in the form of descriptive of the observed symptoms, in the form of numbers or coefficients between variables. Quantitative descriptive research aims to describe events and occurrences objectively (Astono, 2021). The thing that distinguishes descriptive research from other research is that research focuses more on solving existing problems (Paramita et al., 2021). Descriptive research seeks to describe the object or subject being researched as it is in accordance with what is in the field (Syahrizal & Jailani, 2023). Because of its quantitative nature, this study uses numerical data to analyze and describe facts, events, or conditions as they are without variable manipulation (Waruwu et al., 2025). Because it uses two methods, namely descriptive (qualitative) and quantitative, the design of this research is called the combination method or the mixed method. Sugiyono (2017) said that the combination research method is a research method that mixes between two methods, namely quantitative methods and qualitative methods.

The questions analyzed in this study are multiple-choice questions and descriptions in final semester assessment II of the Cultural Arts class IV State Elementary School in Dabin I, South Tegal District, Tegal City for the 2024/2025 Academic Year. The procedure used in this study is to analyze each question item qualitatively by looking for the validity of the content in each question item. Furthermore, the questions were analyzed quantitatively by determining the validity, reliability, level of difficulty, discriminating power, and effectiveness of the deception.

### *Research site and participants*

The place used to carry out the research was the State Elementary School in Dabin I, South Tegal District, Tegal City, which uses an independent curriculum where there are eight schools. The eight elementary schools are SD Negeri Debong Tengah 1, SD Negeri Debong Tengah 2, SD Negeri Debong Tengah 3, SD Negeri Randugunting 1, SD Negeri Randugunting 2, SD Negeri Randuguntung 3, SD Negeri Randugunting 4, and SD Negeri Raduguting 5. The population in this study is the answer sheets of grade IV students of State Elementary School in Dabin I, South Tegal District, Tegal City with a total of 221 answer sheets from eight elementary schools. Sampling in this study uses a probability sampling technique with a Simple Random Sampling type. The probability sampling technique is used to provide equal opportunities for members of the

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

population to be selected as members of the sample (Sugiyono, 2017). The type of simple random sampling is a way of taking samples from a population that is done randomly and does not pay attention to the strata in that population (Sugiyono, 2017). This is because the members of the population are homogeneous answer sheets for grade IV students. The sampling technique used the Slovin formula in Thoifah (2015) and obtained as many as 144 students' answer sheets.

***Data collection and analysis***

According to Sugiyono (2020), data collection techniques are the most appropriate step in research, because the main purpose of research is to obtain data. The data collection in this study was carried out through the beginning of the problem to be researched can be found by the researcher using interviews as a data collection technique (Sugiyono, 2023). Interviews in this study were conducted with grade IV teachers, in order to obtain information about the condition of students and facts related to the implementation of the final assessment of the even semester of Cultural Arts subjects. The interview used is an unstructured interview, which is carried out by giving questions to the teacher in an irregular manner and the questions continue by adjusting the teacher's previous answers. Documentation is a data collection technique that is no less important than other techniques. Through the documentation of data related to research variables in the form of notes, transcripts, books, newspapers, magazines, inscriptions, and so on can be found (Sugiyono, 2023). The documentation in this study is in the form of student answer sheets, test answer keys, and end-of-semester assessment questions for class IV Cultural Arts subjects.

According to Sugiyono (2017), a research instrument is a tool used to measure and collect data from observed natural and social phenomena. The data collection carried out by the researcher uses tools to make it easier to obtain data. The data collection tools used by the researcher are interview guidelines and a checklist of documentation data. Interview guidelines are one of the data collection instruments in the form of questions about important points that are prepared and the answers can answer research problem questions. The checklist is a data collection instrument in the form of a rubric that contains statements related to research variables to match whether the data has been fulfilled or not.

The data analysis techniques used in this study include qualitative and quantitative analysis. Qualitative analysis in this study includes material aspects, construction, language, distribution of the cognitive domain level of Bloom's taxonomy, and the results of test implementation interviews. The analysis quantitatively includes testing the validity, reliability, difficulty, differentiating power, and effectiveness of the deception. According to Sugiyono (2017), activities in qualitative data analysis are carried out interactively and take place continuously until complete, so that the data is saturated". The analysis of material, construction, and language aspects in this study aims to determine the validity of the content in multiple-choice questions and descriptions. The results of the analysis carried out by the researcher are then consulted with the review expert for review. The results of the examination of question items using the panel technique are then specified in the domain using a four-point scale. Martuza, Hambleton, and Bausell (1977) in Gregory (2013) explain that domain specifications using a four-point scale consist of: irrelevant, somewhat relevant, quite relevant, and highly relevant.

The questions analyzed in this study are multiple-choice questions and descriptions in final semester assessment II of the Cultural Arts class IV State Elementary School in Dabin I, South Tegal District, Tegal City for the 2024/2025 Academic Year. The procedure used in this study is to analyze each question item qualitatively by looking for the validity of the content in each question item. Furthermore, the questions were analyzed quantitatively by determining the validity, reliability, level of difficulty, discriminating power, and effectiveness of the deception.

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

The quality of the question items can also be seen from the level of the cognitive domain of Bloom's taxonomy used in working on the questions. The classification of the cognitive domain proposed by Bloom was then revised by Anderson and Krathwohl (2001) in Yani (2019) reconstructing the cognitive taxonomy of six levels with several changes, namely inhibition, understanding, applying, analyzing, evaluating, and creating. The levels of remembering (C1), understanding (C2), and applying (C3) are the basis for the development of LOTS (Lower Order Thinking Skills) question items, while analyzing (C4), evaluating (C5), and creating (C6) are the basis for the development of HOTS (Higher Order Thinking Skills) question items). Meanwhile, quantitative analysis is carried out to analyze empirical data from the question item in question. Quantitative analysis in this study used the Anates V4 computer program. Anates V4 is a computer application program that aims to analyze question items (Rachmadani et al., 2024). Anates V4 is specifically designed to analyze multiple-choice questions and essays (Maulidah et al., 2023).

## Results

The results of the research that will be discussed in this study are to answer the objectives of the research, namely how the analysis of multiple-choice question items and essays of the end of even semester assessment of Cultural Arts subjects in Grade IV is reviewed from the aspects of validity, reliability, difficulty level, discriminating power, and effectiveness of deception empirically. Then about the analysis of multiple-choice question items and descriptions of Cultural Arts subjects in PAS even semester grade IV reviewed from the aspects of material, construction, and language theoretically. The last is about the distribution of the level of the cognitive domain that is measured in multiple-choice question items and descriptions of Cultural Arts subjects in PAS even semester of grade IV.

### *Analysis of multiple-choice questions and descriptive questions of cultural arts subjects in the even semester of grade IV reviewed from the aspects of validity, reliability, level of difficulty of the questions, discriminating power, and effectiveness of the deception empirically*

#### *Validity analysis*

The analysis of the validity of the final semester assessment II question items for the Cultural Arts subject for the 2024/2025 school year grade IV SD Negeri Dabin I, South Tegal District, Tegal City was carried out using the Anates V4 software program. Based on the significance limit of the correlation coefficient in Anates V4 with a degree of freedom of 30, it was then consulted at the significance level of p=0.01 and p=0.05. The result at p=0.01 obtained a significance value of 0.449 and at p=0.05 obtained a significance value of 0.349. After the results of the analysis were categorized, it was found that there were 5 questions categorized as "significant", 10 questions categorized as "very significant", and 15 questions categorized as "insignificant". The following is the percentage of the results of the validity analysis of multiple-choice questions for final semester assessment II for the Cultural Arts subject for the 2024/2025 school year grade IV SD Negeri Dabin I, South Tegal District, Tegal City presented in table 1.

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

**Table 1.** *Percentage of analysis of the validity of multiple-choice questions*

| Category | Sum | Question Number | Percentage |
|---|---|---|---|
| Very Significant | 10 | 6, 10, 13, 17, 18, 22, 23, 30, 32, 33 | 33% |
| Significant | 5 | 8, 15, 27, 28, 31 | 17% |
| Insignificant | 15 | 1, 2, 3, 4, 5, 9, 11, 12, 14, 19, 20, 21, 25, 26, 29 | 50% |

The results of the analysis of the validity of the essay questions are then categorized into the significance limits of the correlation coefficient that have been determined. Based on the significance limit of the correlation coefficient in Anates V4 with a degree of freedom of 5, then consulted at the significance level of p=0.01 and p=0.05. The result at p=0.01 obtained a significance value of 0.8744 and at p=0.05 obtained a significance value of 0.754. The following percentage of the results of the analysis of the validity of the essay question items can be seen in table 2.

**Table 2.** *Percentage of analysis of the validity of essay questions*

| Category | Sum | Question Number | Percentage |
|---|---|---|---|
| Very Significant | 1 | 16 | 20% |
| Significant | 2 | 24, 34 | 40% |
| Insignificant | 2 | 7, 35 | 40% |

Based on the results of the validity of multiple-choice questions and essays that have been analyzed, 17% of multiple-choice questions with the Significant category and 33% with the Very Significant category. While the Significant category in the essay questions is 40% and 20% with the Very Significant category. A total of 50% of multiple-choice questions and 60% of essay questions are in valid condition, so they are worth using. Meanwhile, 50% of multiple-choice questions and 40% of essay questions are in invalid condition, so they are not suitable for use.

### *Reliability analysis*

The results of the analysis showed that for the multiple-choice question item, the reliability coefficient was 0.73. The following are the results of the calculation of the reliability analysis of multiple-choice question items using the Anates V4 software program presented in table 3.

**Table 3.** *Multiple-choice reliability analysis results*

| Average | Baku intersection | XY Correlation | Test Reliability |
|---|---|---|---|
| 16,38 | 5, 13 | 0, 57 | 0, 73 |

The results of the reliability analysis for the description question item showed that the reliability coefficient obtained was 0.73. The following are the results of the calculation of the reliability analysis of the description question items using the Anates V4 software program presented in table 4.

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

**Table 4.** *Result of reliability analysis essay*

| Average | Baku intersection | XY Correlation | Test Reliability |
|---|---|---|---|
| 6, 58 | 2, 76 | 0, 57 | 0, 73 |

### *Difficulty level analysis*

The results of the difficulty level analysis are then categorized based on the difficulty level criteria according to Arikunto (2016). The results of the multiple-choice analysis showed that there were 7 questions categorized as "easy", 19 questions categorized as "moderate," and 4 questions categorized as "difficult". The following is the percentage of difficulty level analysis in the multiple-choice question items presented in table 5.

**Table 5.** *Percentage of difficulty level analysis of multiple-choice questions*

| Category | Sum | Question Number | Percentage |
|---|---|---|---|
| Easy | 7 | 2, 4, 21, 22, 25, 28, 33 | 23% |
| Moderate | 19 | 1, 3, 5, 6, 8, 10, 11, 12, 13, 15, 17, 18, 23, 26, 27, 29, 30, 31, 32 | 64% |
| Difficult | 4 | 9, 14, 19, 20 | 13% |

The item of the description question that has been analyzed can also be determined by the level of difficulty, the results of the analysis show that there are 1 question in the "easy" category and 4 questions in the "medium" category. The following is the percentage of difficulty level analysis in the description question items presented in table 6.

**Table 6.** *Percentage of analysis difficulty essay questions*

| Category | Sum | Question Number | Percentage |
|---|---|---|---|
| Easy | 1 | 35 | 20% |
| Moderate | 4 | 7, 16, 24, 34 | 80% |
| Difficult | - | - | 0% |

### *Differentiating power analysis*

The results of the differentiating power analysis for multiple-choice questions showed that there were 2 questions categorized as "very good", 15 questions categorized as "good", 11 questions categorized as "adequate", no questions categorized as "not good", and there were 2 questions categorized as "bad". The percentage of the results of the differentiating power analysis of multiple-choice question items that have been analyzed using Anates V4 can be read in table 7.

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

**Table 7.** *Percentage of differentiating power analysis of multiple-choice questions*

| Category | Sum | Question Number | Percentage |
|---|---|---|---|
| Bad | 2 | 2, 20 | 7% |
| Not good | - | - | 0% |
| Adequate | 11 | 1, 4, 5, 9, 11, 12, 14, 19, 21, 25, 26 | 36% |
| Good | 15 | 3, 6, 8, 10, 13, 15, 18, 22, 23, 27, 28, 29, 31, 32, 33 | 50% |
| Very good | 2 | 17, 30 | 7% |

Regarding the description questions that have been analyzed for differentiating power, there are 3 questions categorized as "very good", 1 question categorized as "good", and 1 question categorized as "adequate". No questions categorized as "not good" and "bad". The percentage of the results of the differentiating power analysis of the description question items that have been analyzed using Anates V4 can be read in table 8.

**Table 8.** *Percentage of analysis of discriminating power essay questions*

| Category | Sum | Question Number | Percentage |
|---|---|---|---|
| Bad | - | - | 0% |
| Not good | - | - | 0% |
| Adequate | 1 | 7 | 20% |
| Good | 1 | 35 | 20% |
| Very good | 3 | 16, 24, 34 | 60% |

***Analysis of the effectiveness of deception***

The results of the analysis of the effectiveness of the deceitful are then categorized using the criteria of the effectiveness of the deceitful According to the Ministry of National Education (2008) which states, "The deceitful is said to function if: (1) the choice of the answer is chosen by at least 5% of the students; (2) more answer choices are chosen by students who do not understand the teaching material". So, it can be concluded that the questions analyzed will be effective if chosen by 7 or more students, because the 7 students constitute 5% of the total sample in this study. The results of the analysis of the question items that have been categorized show that there are 25 questions that are categorized as "effective" and 5 questions that are categorized as "ineffective". The following is the percentage of the results of the analysis of the effectiveness of the question item deceivers presented in table 9.

**Table 9.** *Percentage of fraud effectiveness analysis*

| Category | Sum | Question Number | Percentage |
|---|---|---|---|
| Effective | 25 | 1, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 23, 27, 28, 29, 30, 31, 32, 33 | 83% |
| Ineffective | 5 | 2, 4, 21, 25, 26 | 17% |

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

***Analysis of multiple-choice question items and descriptions of cultural arts subjects in the even semester grade IV reviewed from the aspects of material, construction, and language theoretically***

***Material, construction, and language analysis***

Analysis activities have been carried out first by the researcher, the results of the study are then consulted with two experts to be checked again for veracity. The following is the calculation of the content validity index in multiple-choice questions, reviewed from the aspects of material, construction, and language:

Index of validity of the content of the material aspect $= \frac{30}{0+0+0+30} = = \frac{30}{30} 1.00$

Validity index of construction aspect content $= \frac{30}{0+0+0+30} = = \frac{30}{30} 1.00$

Index of validity of the content of the language aspect $= \frac{30}{0+0+0+30} = \frac{30}{30} = 1,00$

The following is the calculation of the content validity index in the description question, reviewed from the aspects of material, construction, and language:

Index of validity of the content of the material aspect $= \frac{5}{0+0+0+5} = = \frac{5}{5} 1.00$

Validity index of construction aspect content $= \frac{5}{0+0+0+5} = = \frac{5}{5} 1.00$

Validity index of language aspect content $= \frac{5}{0+0+0+5} = = \frac{5}{5} 1.00$

***Distribution of the level of the cognitive domain that is measured in multiple-choice items and descriptions of cultural arts subject in the seven semester of grade IV***

The analysis of the distribution of the cognitive domain level of Bloom's taxonomy was carried out by matching the question items with the categories of the cognitive level Anderson and Krathwohl in Yani (2019). The results of the cognitive domain distribution analysis for multiple-choice questions showed that there were 5 "remember" or C1 level questions, 5 "understanding or C2" level questions, 12 "apply or C3" level questions, 5 "analyze or C4" level questions, 3 "assess or C5" level questions, and there were no "create or C6" level questions. The following are the results of the percentage analysis of the distribution of the cognitive domain level in the multiple-choice questions presented in table 10.

**Table 10.** *Percentage of cognitive domain level distribution analysis multiple choice questions*

| Levels | Criterion Taxonomy Bloom | Question Number | Percentage |
|---|---|---|---|
| Remember (C1) | Finding, recalling, reading, mentioning, reciting, memorizing, compiling lists, underlining, matchmaking, choosing, defining, stating that. | 1, 5, 21, 22, and 33 | 16, 67% |
| Understand (C2) | Explain, interpret, interpret, tell, display, exemplify, summarize, conclude, compare, classify, show, describe, distinguish, predict, estimate, explain, replace | 2, 3, 4, 6, and 27 | 16, 67% |

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

| | | | |
|---|---|---|---|
| Apply (C3) | Perform, use, conceptualize, implement, determine, demonstrate, process, calculate, connect, perform, prove, produce, demonstrate, complete, adjust, discover | 8, 10, 13, 17, 18, 19, 20, 23, 25, 28, 30, and 32 | 40% |
| Analyze (C4) | Differentiating, associating, organizing, attributing, studying, diagnosing, detailing, delineating, detecting, breaking, separating, selecting, selecting, comparing, opposing, delineating, discovering. | 9, 11, 12, 26, and 31 | 16, 67% |
| Judge (C5) | Checking, criticizing, verifying, defending, validating, supporting, projecting, comparing, concluding, criticizing, assessing, evaluating, advising, arguing, interpreting, recommending | 14, 15, and 29 | 10% |
| Create (C6) | Building, planning, producing, combining, designing, reconstructing, creating, abstracting, categorization, combining, composing, designing, creating, creating, rearranging, assembling | - | 0% |

The results of the cognitive domain level distribution analysis for description questions showed that there were 2 tiered questions "understanding or C2", 2 tiered questions "applying or C3", 1 tiered question "analyzing or C4", and there were no tiered questions "remembering or C1", "assessing or C5", and "creating or C6". The following are the results of the percentage analysis of the distribution of the cognitive domain level in the description questions presented in table 11.

**Table 11.** *Percentage of distribution analysis of the cognitive domain level essay questions*

| Levels | Bloom Taxonomy Criteria | Question Number | Percentage |
|---|---|---|---|
| Remember (C1) | Finding, recalling, reading, mentioning, reciting, memorizing, compiling lists, underlining, matching, choosing, defining, stating that. | - | 0% |
| Understand (C2) | Explain, interpret, interpret, tell, display, exemplify, summarize, conclude, compare, classify, show, describe, distinguish, predict, estimate, explain, replace | 7 and 34 | 40% |
| Apply (C3) | Perform, use, conceptualize, implement, determine, demonstrate, process, calculate, connect, perform, prove, produce, demonstrate, complete, adjust, discover | 24 and 35 | 40% |
| Analyze (C4) | Differentiating, associating, organizing, attributing, studying, diagnosing, detailing, delineating, detecting, breaking, separating, selecting, selecting, comparing, opposing, delineating, discovering. | 16 | 20% |
| Judge (C5) | Checking, criticizing, verifying, defending, validating, supporting, projecting, comparing, concluding, criticizing, assessing, evaluating, advising, arguing, interpreting, recommending | - | 0% |
| Create (C6) | Building, planning, producing, combining, designing, reconstructing, creating, abstracting, categorization, combining, composing, designing, creating, creating, rearranging, assembling | - | 0% |

## Discussion

The place used to conduct the research was the public elementary schools in Dabin I South Tegal Subdistrict, Tegal City, which totalled eight schools. The eight schools were Debong Tengah 1 State Elementary School, Debong Tengah 2 State Elementary School, Debong Tengah 3 State

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

Elementary School, Randugunting 1 State Elementary School, Randugunting 2 State Elementary School, Randuguntung 3 State Elementary School, Randugunting 4 State Elementary School, and Randuguting 5 State Elementary School. This research was carried out in a period of approximately two months starting from May 2025 to July 2025. The scope of this research is about the analysis of multiple choice items and descriptions from both quantitative and qualitative aspects. The selection of grade IV as the research subject is based on the consideration that students at this level have sufficient cognitive abilities to take part in research activities, and the material under study is in accordance with the grade IV curriculum. In addition, class IV does not have the burden of final exams, so that the implementation of research can run optimally.

An indication instrument is reliable if it has validity (Pratiwi & Rufi'i, 2023). The validity of the multiple-choice question item shows that there are 15 questions (50%) categorized as valid and 15 questions (50%) categorized as invalid. Meanwhile, in the analysis of the question item description, the results were obtained that there were 3 questions (60%) categorized as valid and 2 questions (40%) were categorized as invalid. Based on these results, it can be concluded that multiple-choice and description questions have a very significant level of validity that is quite high. However, there are still some problems with high insignificant categories that need to be improved. There are three factors that can affect the validity of test results, namely the instruments used, the administration and assessment process, and the answers given by students (Arifin, 2017).

To assess the consistency of test actions whose results show regularity, the reliability of the test is used. If an instrument has high reliability, then it can provide the same or consistent results even if it is done by different people and different times (Siregar et al., 2024). In the reliability aspect, the multiple-choice question item shows that 0.73 has a high correlation rate. Likewise, the reliability analysis of the description question also got the same result, which was 0.73 and had a high correlation level. This means that multiple-choice and description questions are suitable for testing to students, it is in accordance with Wells & Wollack (2003) in Azwar (2014) stating that the test used should have a reliability coefficient of 0.70 or more (≥0.70).

The difficulty index is a measure that describes how many test takers managed to answer the question correctly (Yusdiana et al., 2019). A good exam script is one that has a balanced level of difficulty (Magdalena, 2022). Difficult question items will make students give up, while easy questions are not able to stimulate students' ability to solve problems (Nurhalimah et al., 2022). In terms of difficulty level, multiple choice questions showed that there were 7 questions (23%) categorized as "easy", 19 questions (64%) categorized as "moderate", and 4 questions (13%) categorized as "difficult". Sudjana (2017) stated that "a comparison between questions that are moderately difficult can be made 3-4-3, which means, 30% of the questions are in the easy category, 40% are in the medium category, and 30% are in the difficult category". Meanwhile, the aspect of the difficulty level of the description question item shows that there is 1 question (20%) in the "easy" category and 4 questions (80%) in the "medium" category. Sudjana (2017) stated that "a comparison between questions that are moderately difficult can be made 3-4-3, which means, 30% of the questions are in the easy category, 40% are in the medium category, and 30% are in the difficult category". Based on these results, the final semester assessment II questions for the subject of Cultural Arts have not met the requirements for comparing questions, so improvements must be made so that the distribution of questions in the easy-medium-difficult category is even.

Differentiating power is the extent to which a question item is arranged to show the difference between upper group students (smart) and lower group students (less smart) (Qomariyah, 2022). The results of the analysis of multiple-choice question items showed that there were 11 questions categorized as "adequate", 15 questions categorized as "good", 2 questions categorized as "very good", and 2 questions categorized as "not good". Meanwhile, the results of the analysis of the description question items showed that there were 3 questions categorized as

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

"very good", 1 question categorized as "good", and 1 question categorized as "adequate". Based on the results of the analysis, question items that are still categorized as not good can be replaced or corrected to meet the classification requirements. Questions that have enough differentiating power can be changed first, then the question can be added to the question bank (Humaria et al., 2023).

Tricksters who do not have the appeal for students to choose are not good tricksters because they are too flashy or heterogeneous. If students do not feel interested in choosing a deceiver, then the deceiver does not carry out his function properly (Widiyanto, 2018). In the aspect of deceiving effectiveness, the results of the analysis of multiple-choice question items showed that there were 25 questions categorized as "effective" with a percentage of 83% and 5 questions categorized as "ineffective" with a percentage of 17%. Meanwhile, in the description question, there is no analysis of the effectiveness of the deception. Multiple-choice questions can be a guide in choosing questions objectively (Ikawati et al., 2022).

The quality of multiple-choice question items was reviewed from the aspects of material, construction, and language, the validity index of the content was 1.00. Thus, every aspect of the multiple-choice question item has a content validity that is categorized as "very high", this is in accordance with the content validity criteria according to Wikrama (2015) which states that "if the content validity criteria have a range of 0.0 – 1.00, then the validity of the content is categorized as very high. Likewise, the description questions received a validity index of 1.00 with a high category in the aspects of material, construction, and language.

The results of the analysis of the distribution of the cognitive domain level in multiple-choice questions and descriptions are evenly distributed, because there are already questions ranging from the cognitive domain C1 to C6. The results of the cognitive domain level distribution analysis in multiple-choice questions showed that there were as many as 10 questions with a cognitive level of "LOTS (Lower Order Thinking Skills)", 12 questions with a cognitive level of "MOTS (Middle Order Thinking Skills)", and 8 questions with a cognitive level of "HOTS (Higher Order Thinking Skills). Analysis of the description questions showed that there were 2 questions that had a cognitive level of "LOTS (Lower Order Thinking Skills)", 2 questions had a cognitive level of "MOTS (Middle Order Thinking Skills)", and 1 question had a cognitive level of "HOTS (Higher Order Thinking Skills). As stated by (Center for Education Assessment 2020), the thinking process in the cognitive realm is classified into three levels of the cognitive realm, namely high-level thinking skills (HOTS), medium-level thinking skills (MOTS), and low-level thinking skills (LOTS) covering aspects of knowing and understanding.

## Implication

There are 2 implications of the research results, namely (1) To improve the quality of learning, it is urged to prepare indirect question items that are not directly arranged carelessly, but must pay attention to the rules of question preparation and prepare the components of the preparation of questions, because this can have an impact on the quality of the question items that have been prepared so that the learning results of students are not necessarily reliable. (2) To improve pedagogical competence, teachers need to pay more attention in the preparation of evaluation instruments, for example, in the grid, teachers can prepare grids that are in accordance with the format of the grid that has been determined consisting of the level of education, program/department, subject, curriculum, and number of questions. In addition, there are also basic competencies, materials, question indicators, cognitive levels, question numbers, and question forms so that the grid becomes clearer and easier for students to understand. In addition, in terms of the questions, it also needs to be considered where the questions that are not good also

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

need to be improved again in order to produce more quality questions, therefore, teachers need to analyze the question items both qualitatively and quantitatively before the questions are tested to students. And the distribution of the distribution of the cognitive domain level also needs to be considered by teachers in the preparation so that it can be evenly distributed from C1 to C6.

The results of the research on the analysis of final semester assessment II questions for the subject of Cultural Arts certainly have advantages and disadvantages that can be used as a basis for appropriate follow-up research. This can be done by the next researcher who is expected to be able to conduct research on the analysis of question items in the independent curriculum that is currently used so that they can find out the changes in the question items in the next curricula.

## Conclusion

Based on quantitative analysis on multiple choice questions which includes analysis of validity, reliability, difficulty level, differentiating power, and the effectiveness of the checkers, it shows that multiple choice items can be said to be good and feasible to be tested on students, but there are still some items that need to be corrected, namely at numbers 2, 4, 20, 21, 25, and 26. Item numbers 2 and 20 need to be corrected because the items are invalid and have poor differentiating power. Meanwhile, question items number 4, 21, 25, and 26 need to be corrected because the question items are invalid and the effectiveness of the checkers is not effective. In terms of material, construction, and language aspects in theoretical analysis, it has a content validity of 1.00 which is categorised as "very high", so this PAS II question is suitable for testing to students.

The distribution of cognitive domain levels measured on multiple choice questions of even semester PAS in Mathematics for the 2021/2022 school year for class IV SD Negeri Dabin I, Tegal Selatan Subdistrict, Tegal City, namely, there are 5 questions at the level of "remembering or C1" with a percentage of 16.67%, 5 questions at the level of "understanding or C2" with a percentage of 16.67%, 12 questions at the level of "applying or C3" with a percentage of 40%, 5 questions at the level of "analysing or C4" with a percentage of 16.67%, 3 questions at the level of "assessing or C5" with a percentage of 10%. In addition, it was found that 10 questions had a cognitive level of "LOTS (Lower Order Thinking Skills)" with a percentage of 33.33%, 12 questions had a cognitive level of "MOTS (Middle Order Thinking Skills)" with a percentage of 40%, and 8 questions had a cognitive level of "HOTS (Higher Order Thinking Skills) with a percentage of 26.67%. The distribution of cognitive domain levels for description questions shows that, there are 2 questions at the level of "understanding or C2" with a percentage of 40%, 2 questions at the level of "applying or C3" with a percentage of 40%, 1 question at the level of "analysing or C4" with a percentage of 20%. In addition, it was found that 2 questions had a cognitive level of "LOTS (Lower Order Thinking Skills)" with a percentage of 40%, 2 questions had a cognitive level of "MOTS (Middle Order Thinking Skills)" with a percentage of 40%, and 1 question had a cognitive level of "HOTS (Higher Order Thinking Skills) with a percentage of 20%.

The results of research on the analysis of end-of-semester items in Cultural Arts subjects can be used as a basis for appropriate research. In addition, future researchers are expected to conduct research on the analysis of items in the independent curriculum with the deep learning method currently used so that they can find out changes in items in the independent curriculum and the independent curriculum with the deep learning method.

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Reference**

Abd Rahman, B. P., Munandar, S. A., Fitriani, A., Karlina, Y., & Yumriani, Y. (2022). Pengertian pendidikan, ilmu pendidikan dan unsur-unsur pendidikan. *Al-Urwatul Wutsqa: Kajian Pendidikan Islam*, *2*(1), 1–8. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://d1wqtxts1xzle7.cloudfront.net/91021639/7757_24249_1_PB_2_-libre.pdf?1663123266=&response-content-disposition=inline%3B+filename%3DPENGERTIAN_PENDIDIKAN_ILMU_PENDIDIKAN_DA.pdf&Expires=1760346750&S

Agustianti, R., Abyadati, S., Nussifera, L., Irvani, A. I., Handayani, D. Y., Hamdani, D., & Amarulloh, R. R. (2022). *Asesmen Dan Evaluasi Pembelajaran*. Tohar Media.

Alindra, A. ., Attalia, S. ., Meilisa, M., Sari, G. R. ., Marisa, D., Safariyah, S. ., Raihana, N. ., & Sari, T. . (2025). Perspektif Guru dan Siswa : Penerapan Kurikulum Merdeka di Sekolah Dasar. *JISPENDIORA Jurnal Ilmu Sosial Pendidikan Dan Humaniora*, *4*(2), 148–157. https://doi.org/10.56910/jispendiora.v4i2.2220

Arifin, Z. (2017). *Evaluasi Pembelajaran*. PT Remaja Rosdakarya.

Arikunto, S. (2016). *Dasar-dasar Evaluasi Pendidikan Edisi 2*. Bumi Aksara.

Asrul, A., Saragih, A. H., & Mukhtar, M. (2022). *Evaluasi pembelajaran*.

Astono, A. D. (2021). *METODOLOGI PENELITIAN Metode Penelitian Manajemen Pendekatan Kuantitatif dan Kualitatif Buku Ajar Perkuliahan*. Cahya Ghani Recovery.

Azwar, S. (2014). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Firmansyah. (2020). Pengembangan Kurikulum Pendidikan Agama Islam Berbasis Multikultural. *Anthropos: Jurnal Antropologi Sosial Dan Budaya (Journal of Social and Cultural Anthropology)*, *5*(2), 164–169. https://doi.org/10.24114/antro.v5i2.14384

Gregory, R. J. (2013). *Tes Psikologi Sejarah, Prinsip, dan Aplikasi Edisi Keenam Jilid I*. Erlangga.

Hakim, A. R., & Darojat, J. (2023). Pendidikan multikultural dalam membentuk karakter dan Identitas Nasional. *Jurnal ilmiah Profesi pendidikan*, *8*(3), 1337–1346. https://doi.org/https://doi.org/10.29303/jipp.v8i3.1470

Hidayati, K., & Nisa, Z. E. K. (2023). Analisis Butir Soal Penilaian Akhir Semester Matematika. *Aksioma: Jurnal Program Studi Pendidikan Matematika*, *12*(3), 3516–3529. https://doi.org/http://dx.doi.org/10.24127/ajpm.v12i3.7575

Humaria, I. U., Muti'ah, A., & Syukron, A. (2023). Analisis Butir Soal Ujian Akhir Semester Ganjil Bahasa Indonesia Kelas X Tahun Ajaran 2022/2023 di SMAN 2 Jember. *Lingua Skolastika*, *2*(1), 35. https://doi.org/10.19184/linsko.v2i1.38997

Ihsan, M., Muharyati, S., & Zaitun, Z. (2025). Analisis Kebijakan Kurikulum Merdeka Pengengembangan dan Implementasi. *Jurnal Ilmu Pendidikan Muhammadiyah Kramat Jati*, *6*(1), 62–69. https://doi.org/https://doi.org/10.55943/jipmukjt.v6i1.359

Ikawati, H. D., Jayadi, & Hermansyah, H. (2022). Analisis Kualitas Tes dan Butir Soal Sejarah di SMAN 1 Praya Timur. *Jurnal Pendidikan dan Konseling*, *4*(4), 6262–6267. https://doi.org/https://doi.org/10.31004/jpdk.v4i4.6489

Irawan, Y., Sa'adah, R., Thoyibah, A., Sari, M., Muslimah, M., & A'isyah, S. (2023). Analisis Pengembangan Butir Soal Perbandingan Bertingkat untuk Meningkatkan Kemampuan Pemecahan Masalah Peserta Didik. *Indiktika: Jurnal Inovasi Pendidikan Matematika*, *5*(2), 222–232. https://doi.org/https://doi.org/10.31851/indiktika.v5i2.11592

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

Jumini, S., Madnasri, S., Cahyono, E., Nadrah, Y. M., Hamzah, H., & Fatiatun, F. (2024). Analysis of Items for Measuring Technological Literacy Using the Rasch Model. *Jurnal Pendidikan IPA Indonesia*, *13*(2). https://doi.org/https://doi.org/10.15294/04k5q459

Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, *77*, S85–S88. https://doi.org/https://doi.org/10.1016/j.mjafi.2020.11.007

Kurniawan, A., Febriant, A. N., Hardianti, T., Ichsan, I., Desy, D., Risan, R., & Fuad, H. (2022). *Evaluasi pembelajaran*.

Magdalena, I., Nurul Annisa, M., Ragin, G., & Ishaq, A. R. (2021). Analisis Penggunaan Teknik Pre-Test dan Post-Test pada Mata Pelajaran Matematika dalam Keberhasilan Evaluasi Pembelajaran di SDN Bojong 04. *Nusantara*, *3*(2), 150–165. https://ejournal.stitpn.ac.id/index.php/nusantara/article/view/1250

Maulani, G., Septiani, S., Susilowaty, N., Rusmayani, N. G. A. L., Evenddy, S. S., Nababan, H. S., & Nurlely, L. (2024). *Evaluasi Pembelajaran*. Sada Kurnia Pustaka.

Maulidah, H., Sukarno, S., & Syefrinando, B. (2023). Analisis Kualitas Instrumen Tes Fisika Kelas X Menggunakan Software Anates. *Physics and Science Education Journal (PSEJ)*. https://doi.org/10.30631/psej.v2i3.1660

Muluki, A., Bundu, P., & Sukmawati, I. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV MI Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, *4*(1), 86–96. https://doi.org/https://doi.org/10.23887/jisd.v4i1.23335

Nasional, D. P. (2008). *Panduan Analisis Butir Soal*. https://teguhsasmitosdp1.files.wordpress.com/2010/05/butir_soal1.pdf

Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan Antara Validitas Item Dengan Daya Pembeda dan Tingkat Kesukaran Soal Pilihan Ganda PAS. *Jurnal NSER*, *4*(3), 249–257. https://doi.org/https://doi.org/10.21107/nser.v4i3.8682

Paramita, R. W. D., Rizal, N., & Sulistyan, R. B. (2021). *Metode penelitian kuantitatif*.

Pembelajaran, P. A. dan. (2020). *AKM dan Implikasinya pada Pembelajaran*. Badan Penelitian dan Pengembangan dan Perbukuan Kementerian Pendidikan dan Kebudayaan.

Pratiwi, L. E. P., & Rufi'i, H. (2023). Analisis Soal Sumatif IPA Mengenai Gaya pada Peserta Didik Kelas IV di SD YP Nasional. *Didaktika: Jurnal Kependidikan*, *12*(4), 599–610. https://doi.org/https://doi.org/10.58230/27454312.287

Qomariyah, T. (2022). Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal TOAFL Universitas Hasyim Asy'ari Tebuireng Jombang. *Jurnal Pendidikan Bahasa Arab*, *6*(1), 1–18. https://doi.org/https://doi.org/10.32699/liar.v6i1.2549

Rachmadani, F. C., Lestari, C. D., Hakim, L., & Pratiwi, V. (2024). Analisis Butir Soal Pilihan Ganda pada Elemen Akuntansi Keuangan Guna Mengoptimasi Evaluasi Menggunakan Anates V4.0. *Jurnal Nirta: Studi Inovasi*, *4*(1), 71–88.

Sanga, L. D., & Wangdra, Y. (2023). Pendidikan adalah faktor penentu daya saing bangsa. *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK)*, *5*, 84–90. https://doi.org/https://doi.org/10.33884/psnistek.v5i.8067

Sholihan, M. P., Rusmayani, N. G. A. L., ST, S., Udil, P. A., Shalehati, N. A., Hafizi, M. Z., & Aran, A. M. (2024). *Evaluasi Pembelajaran*. Cendekia Publisher.

Siregar, N. H., Remiswal, R., & Khadijah, K. (2024). Analisis Butir Soal Ujian Tengah Semester Pada Mata Pelajaran Pendidikan Agama Islam. *Urwatul Wutsqo: Jurnal Studi Kependidikan Dan Keislaman*, *13*(2), 179–189. https://doi.org/https://doi.org/10.54437/urwatulwutsqo.v13i2.1637

Sudijono, A. (2015). *Pengantar Evaluasi Pendidikan*. Rajagrafindo.

Sudjana, N. (2017). *Penilaian Hasil Proses Belajar Mengajar*. PT Remaja Rosda Karya.

ISSN |2355-3669| E-ISSN |2503-2518|
Volume 12 | Number 2 | December 2025|

Sugiyono. (2017). *Metode Penelitian Kombinasi (Mixed Methods)*. Alfabeta.
Sugiyono. (2020). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. CV. Alfabeta.
Sugiyono. (2023). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. ALFABETA, cv.
Syaadah, R., Ary, M. H. A. A., Silitonga, N., & Rangkuty, S. F. (2022). Pendidikan formal, Pendidikan non formal Dan Pendidikan informal. *Pema*, *2*(2), 125–131. https://doi.org/https://doi.org/10.56832/pema.v2i2.298
Syahrizal, H., & Jailani, M. S. (2023). Jenis-jenis penelitian dalam penelitian kuantitatif dan kualitatif. *QOSIM: Jurnal Pendidikan Sosial & Humaniora*, *1*(1), 13–23. https://doi.org/https://doi.org/10.61104/jq.v1i1.49
Thoifah, I. (2015). *Statistika pendidikan dan Metode Penelitian Kuantitatif*. Madani.
*Undang-Undang Nomor 20 Tahun 2003 BAB II Pasal 3 tentang Dasar, Fungsi, dan Tujuan Pendidikan*. (2025). https://kelembagaan.ristekdikti.go.id/wpcontent/uploads/2016/08/UU_no_20_th_2003.pdf
Waruwu, M., Puat, S. N., Utami, P. R., Yanti, E., & Rusydiana, M. (2025). Metode penelitian kuantitatif: Konsep, jenis, tahapan dan kelebihan. *Jurnal Ilmiah Profesi Pendidikan*, *10*(1), 917–932. https://doi.org/https://doi.org/10.29303/jipp.v10i1.3057
Widiyanto, J. (2018). *Evaluasi Pembelajaran (Sesuai dengan Kurikulum 2013)*. Unipma Press.
Wikrama, I. N. (2015). *Validitas dan Reliabilitas Tes Hasil Belajar*. http://karya-wikrama.blogspot.com/2015/04/validitas-dan-reliabilitas-teshasil.html
Yani, A. (2019). *Cara Mudah Menulis Soal HOTS*. PT. Refika Aditama.
Yusdiana, Y., Zamsir, Z., & Kodirun, K. (2019). Kualitas Tes Sumatif Mata Pelajaran Matematika Kelas VIII SMP Negeri 5 Kendari Tahun Ajaran 2016/2017. *Jurnal Penelitian Pendidikan Matematika*, *6*(3), 141. https://doi.org/10.36709/jppm.v6i3.9146
Zamzania, W. H., & Aristia, R. (2018). *Jenis-Jenis Instrumen dalam Evaluasi Pembelajaran*. Universitas Muhammadiyah Sidoarjo. http://eprints.umsida.ac.id/id/eprint/4050