JGSD: Jurnal Guru Sekolah Dasar
ISSN: 3063-0169 (Online)
Vol 1 No. 1 2024, pp. 1-9

Quality Analysis of Grade 6 Literacy and Numeracy End-of-Semester Exam Questions Based on Classical Test Theory

Kusumaningrum Prasetyani [1]✉, Luri Laras Ati[2], Wawan Shokib Rondli[3], Mohammad Kanzunnudin[4]
Magister Pendidikan, Universitas Muria Kudus, Indonesia
email : 202303075@std.umk.ac.id

Abstract

This study aims to analyze the quality of literacy and numeracy end-of-semester exam questions for grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory. The research method used is quantitative by analyzing the validity, reliability, difficulty level, and differentiating power of the questions. The research sample consisted of 10 6th grade students of SD Negeri 3 Pojok. The data was collected through documentation of student answer sheets in the final semester exam. Data analysis was conducted using Anates program. The results showed that out of 30 items, there were 22 valid questions and 8 invalid questions. The reliability coefficient of the question is 0.82 which is included in the high category. The level of difficulty of the questions consisted of 10 easy questions, 15 medium questions, and 5 difficult questions. The differentiating power of the questions consisted of 5 bad questions, 10 fair questions, 10 good questions, and 5 excellent questions. It can be concluded that the quality of literacy and numeracy end-of-semester exam questions for grade 6 of SD Negeri 3 Pojok, Tawangharjo sub-district is classified as good based on Classical Test Theory.

Keywords: *Item Analysis;Classical Test Theory;Literacy;Numeracy*

`Article History:
Received 2024-06-01
Revised 2024-06-04
Accepted 2024-06-07

DOI:
10.70277/jgsd.v1i1.0001

Abstrak

This study aims to analyze the quality of literacy and numeracy end-of-semester exam questions for grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory. The research method used is quantitative by analyzing the validity, reliability, difficulty level, and differentiating power of the questions. The research sample consisted of 10 6th grade students of SD Negeri 3 Pojok. The data was collected through documentation of student answer sheets in the final semester exam. Data analysis was conducted using Anates program. The results showed that out of 30 items, there were 22 valid questions and 8 invalid questions. The reliability coefficient of the question is 0.82 which is included in the high category. The level of difficulty of the questions consisted of 10 easy questions, 15 medium questions, and 5 difficult questions. The differentiating power of the questions consisted of 5 bad questions, 10 fair questions, 10 good questions, and 5 excellent questions. It can be concluded that the quality of literacy and numeracy end-of-semester exam questions for grade 6 of SD Negeri 3 Pojok, Tawangharjo sub-district is classified as good based on Classical Test Theory.

Kata Kunci: *Item Analysis;Classical Test Theory;Literacy;Numeracy.*

INTRODUCTION

Literacy and numeracy are two fundamental skills that students must master from an early age, especially at the primary school level. Literacy refers to the ability to understand, use and reflect on written texts in order to achieve goals, develop knowledge and potential, and participate in society (Murtafiah et al., 2023). Meanwhile, numeracy refers to the knowledge and skills to manage situations or solve problems in everyday life involving the concepts of number, quantity, and data (Fernanda & Hidayah, 2020). These two

abilities are interrelated and become an important foundation for students to be able to follow learning at higher education levels, as well as play an important role in social life.

Assessment of primary school students' literacy and numeracy skills can be done through end-of-semester examinations. The questions used in the exam must be of good quality in order to measure students' abilities accurately and validly. One method that can be used to analyze the quality of questions is Classical Test Theory (CTT) (Handayani & Iba, 2020).

CTT is a measurement theory that analyzes the characteristics of questions based on empirical data from student answers. CTT analysis includes validity, reliability, difficulty level, and question differentiation (Arbain & Sirad, 2023). Validity measures the extent to which a question can measure what should be measured, or how well a question can measure the ability to be measured. Reliability shows the consistency or persistence of the question in measuring student abilities, or how consistently a question can produce the same score if tested on the same group of students at different times. The level of difficulty identifies the proportion of students who answer correctly on a question, which can be used to determine whether the question is classified as easy, medium, or difficult. Differentiating power measures the ability of a question to distinguish between high and low ability students, or how well a question can distinguish between students who master the material and those who do not.

Previous research conducted by Barus et al. (2023) analyzed the quality of grade IV end-of-semester exam questions at SD AL-ITTIHADIYAH and MI NURHIDAYAH using CTT. The results showed that of the 15 items, there were 10 valid questions and 5 invalid questions, with a high reliability coefficient. Another study by Manarisip et al. (2023) analyzed the quality of statistics exam questions using CTT and Rasch Model. The results showed that out of 50 questions, only 21 questions met the criteria for difficulty and differentiation based on CTT.

Based on the importance of literacy and numeracy for elementary school students as well as the need to have quality test questions, this study aims to analyze the quality of literacy and numeracy end-of-semester test questions for grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory. This analysis is expected to provide useful information for teachers and schools in preparing better exam questions in the future, so that they can measure students' literacy and numeracy skills accurately and validly.

Literacy and numeracy are two interrelated skills that are an important foundation for students in participating in learning at higher education levels and play an important role in social life. Therefore, it is very important to assess the literacy and numeracy skills of primary school students. One form of assessment that can be done is through end-of-semester exams.

The end-of-semester exam is a form of summative assessment conducted to measure the achievement of student competencies after following the learning process in one semester. The questions used in the final semester exam must have good quality in order to measure students' abilities accurately and validly. The quality of good questions can be seen from several aspects, such as validity, reliability, difficulty level, and question differentiation.

Question validity refers to the extent to which questions can measure what should be measured, or how well a question can measure the ability to be measured. Question validity can be analyzed through content validity and construct validity. Content validity refers to the extent to which questions can represent all the material tested, while construct validity refers to the extent to which questions can measure the abilities or constructs to be measured (Sesanti et al., 2023).

Question reliability refers to the consistency or persistence of questions in measuring students' abilities. A question is said to have high reliability if the question can produce the same or consistent score when tested on the same group of students at different times. Question reliability can be analyzed using various methods, such as the two-split method, parallel method, or Cronbach's Alpha method.

The difficulty level of a question identifies the proportion of students who answer correctly on a question. Questions that are too easy or too difficult can cause the question to not be able to distinguish students' abilities properly. Therefore, the ideal level of question difficulty is a question with a moderate level of difficulty, which can distinguish between high and low ability students.

The differentiating power of the question measures the ability of the question to distinguish between high and low ability students (Murtafiah et al., 2023). Questions that have good differentiating power can clearly distinguish between students who master the material and those who do not master the material. Conversely, questions that have low differentiating power cannot distinguish well between high and low ability students.

In the context of this study, the analysis of the quality of literacy and numeracy end-of-semester exam questions for grade 6 of SD Negeri 3 Pojok, Tawangharjo sub-district, was carried out using Classical Test Theory (CTT). CTT is a measurement theory that analyzes the characteristics of questions based on empirical data from student answers (Irwan & Kerans, 2022).. This research is important to ensure that the questions used in the literacy and numeracy end-of-semester exams of grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district are of good quality and can measure student abilities accurately and validly. The results of this analysis can provide useful information for teachers and schools in developing better exam questions in the future, so as to improve the quality of assessment of students' literacy and numeracy skills.

RESEARCH METHODS

This study uses a quantitative approach with a descriptive method to analyze the quality of literacy and numeracy end-of-semester exam questions for grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory (CTT). The quality of the questions was analyzed by testing the validity, reliability, difficulty level, and discriminating power of the questions.

The research sample consisted of 10 grade 6 students of SD Negeri 3 Pojok, Tawangharjo sub-district, who took the end of semester literacy and numeracy exams. The sample selection was done using saturated sampling technique, where the entire population was used as the research sample. Data was collected through documentation of students' answer sheets in the final exam of the semester.

Data analysis was conducted using the Anates version 4.09 program, which is special software for analyzing items based on Classical Test Theory. Item validity was analyzed using point biserial correlation, which is a technique for calculating the correlation coefficient between item scores and total scores. The higher the correlation coefficient, the more valid the item.

The reliability of the questions was analyzed using Cronbach's Alpha formula, which is a method for estimating the reliability of a test based on the internal consistency of the items. The reliability coefficient value ranges from 0 to 1, where the higher the coefficient value, the more reliable the question is.

The level of difficulty of questions is classified into three categories, namely easy, medium, and difficult. Questions that are too easy or too difficult can cause the question to not be able to distinguish student abilities properly. Therefore, the ideal level of question difficulty is a question with a moderate level of difficulty.

The differentiating power of questions is classified into four categories, namely poor, fair, good, and excellent. Question discriminating power measures the ability of a question to distinguish between high and low ability students. The higher the discriminating power of the question, the better the question can distinguish between students who master the material and those who do not master the material.

Data analysis was conducted objectively and systematically using the Anates version 4.09 program. The results of the analysis were then interpreted and presented in descriptive form to provide a clear picture of the quality of the end-of-semester test questions for literacy and numeracy in grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory.

RESULT AND DISCUSSION

Research Result

The results of the analysis of the quality of literacy and numeracy end-of-semester exam questions for grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory (CTT) are presented as follows:

1. Question Validity

Question validity measures the extent to which questions can measure what should be measured. In this study, the validity of the questions was analyzed using point biserial correlation. The results of the question validity analysis are presented in Table 1.

Tabel 1. Question Validity Analysis Results

Validity Criteria	Number of Questions	Percentage
Valid	22	73,3%
Invalid	8	26,7%

Validity Criteria	Number of Questions	Percentage
Total	30	100^

From Table 1, it can be seen that of the 30 items analyzed, 22 items (73.3%) were valid and 8 items (26.7%) were invalid. Valid questions have a point biserial correlation coefficient between 0.30 and 0.70. These valid questions can be used to accurately measure students' literacy and numeracy skills.

2. Question Reliability

The reliability of the questions shows the consistency or persistence of the questions in measuring students' abilities. In this study, the reliability of the questions was analyzed using the Cronbach's Alpha formula. The results of the question reliability analysis are presented in Table 2.

Tabel 2. : Results of Problem Reliability Analysis

Reliability Coefficient	Category
0,82	High

From Table 2, it can be seen that the reliability coefficient of the literacy and numeracy end-of-semester exam questions for grade 6 of SD Negeri 3 Pojok, Tawangharjo sub-district is 0.82. This value is included in the high category, which means that the questions have good consistency in measuring students' abilities. In other words, if this question is retested on the same group of students at different times, the results will be relatively the same or consistent.

3. Problem Difficulty Level

The level of question difficulty identifies the proportion of students who answer correctly on a question. In this study, the level of question difficulty was classified into three categories, namely easy, medium, and difficult. The results of the question difficulty level analysis are presented in Table 3.

Tabel 3. Results of Problem Difficulty Analysis

Category of Level of Difficulty	Number of Questions	Percentage
Easy (P > 0.70)	10	33,3%
Medium (0.30 ≤ P ≤ 0.70)	15	50,0%
Difficult (P < 0.30)	5	16,7%
Total	30	100%

From Table 3, it can be seen that of the 30 items, there were 10 questions (33.3%) in the easy category (the proportion of students who answered correctly > 0.70), 15 questions (50.0%) in the medium category (the proportion of students who answered correctly between 0.30 and 0.70), and 5 questions (16.7%) in the difficult category (the proportion of students who answered correctly < 0.30). The ideal level of question difficulty is a question with a moderate level of difficulty, which can distinguish between high and low ability students. In this study, most of the questions (50.0%) fell into the medium category, indicating that the questions had a good level of difficulty.

4. Question Distinguishing Power

Question discriminating power measures the ability of questions to distinguish between high and low ability students. In this study, the differentiating power of questions is classified into four categories, namely poor, fair, good, and excellent. The results of the question discriminating power analysis are presented in Table 4.

Tabel 4. Results of Question Distinguishing Power Analysis

Distinguishing Power Category	Number of Questions	Percentage
Poor (D < 0.20)	5	16,7%
Fair (0.20 ≤ D < 0.40)	10	33,3%

Distinguishing Power Category	Number of Questions	Percentage
Good (0.40 ≤ D < 0.70)	10	33,3%
Excellent (D ≥ 0.70)	5	16,7%
Total	30	100%

From Table 4, it can be seen that out of 30 items, there are 5 questions (16.7%) with poor category (differentiating power <0.20), 10 questions (33.3%) with sufficient category (differentiating power between 0.20 to 0.40), 10 questions (33.3%) with good category (differentiating power between 0.40 to 0.70), and 5 questions (16.7%) with excellent category (differentiating power >0.70).

Questions that have good differentiating power can clearly distinguish between students who master the material and those who do not master the material. In this study, most of the questions (50.0%) had good and excellent discriminating power, indicating that the questions could distinguish between high and low ability students well.

Overall, the results of the analysis of the quality of the literacy and numeracy end-of-semester exam questions of grade 6 SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory show that the questions have good quality. Most of the questions have good validity, reliability, difficulty level and differentiating power.

Discussion

Based on the results of the study, it can be seen that the quality of the end-of-semester exam questions for literacy and numeracy in grade 6 of SD Negeri 3 Pojok, Tawangharjo sub-district, is classified as good based on Classical Test Theory (CTT). Most of the questions have good validity, with a high reliability coefficient. The level of difficulty of the questions is also quite varied, with a balanced proportion of easy, medium, and difficult questions. The discriminating power of the questions was also quite good, with most of the questions having good and excellent categories.

The results of this study are in line with research conducted by Chauhan et al. (2023) who developed a numeracy literacy Minimum Competency Assessment (AKM) question for grade 5 elementary school. The study produced valid questions and in accordance with the numeracy literacy AKM framework. Another research by Muchlizani A et al. (2023) also analyzed the quality of Akidah Akhlak end-of-semester exam items for grade V MI Radhiatul Adawiyah Makassar using CTT. The results showed that most of the questions had good validity, reliability, difficulty level, and differentiating power.

Question validity is one of the important aspects in measuring question quality. Validity refers to the extent to which the question can measure what should be measured (Handayani & Iba, 2020). In this study, most of the questions have good validity, which means that the questions can accurately measure students' literacy and numeracy skills. The validity of questions can be analyzed through several approaches, such as content validity and construct validity.

Content validity refers to the extent to which questions can represent all the material being tested (Mamot et al., 2022). In the context of this study, the content validity of the questions can be seen from the suitability of the questions with the basic competencies and indicators of competency achievement that have been determined in the curriculum. In addition, content validity can also be seen from the representation of questions in measuring various aspects of literacy and numeracy, such as reading, writing, counting, and reasoning.

Construct validity refers to the extent to which the question can measure the ability or construct to be measured. In this study, the construct validity of the questions can be seen from the suitability of the questions with the theories and concepts underlying literacy and numeracy. For example, literacy questions can measure the ability to read comprehension, write, and analyze information, while numeracy questions can measure the ability to count, mathematical reasoning, and problem solving.

In addition to validity, reliability is also an important aspect in measuring the quality of questions. Reliability refers to the consistency or persistence of questions in measuring student abilities (Chauhan et al., 2023). In this study, the reliability coefficient of the questions is high, which means that the questions have good consistency in measuring students' literacy and numeracy skills. Item reliability can be analyzed using various methods, such as the bisection method, parallel method, or Cronbach's Alpha method.

Question difficulty is also an important aspect in measuring question quality. The level of question difficulty identifies the proportion of students who answer correctly on a question (Fatimah & Alfath, 2019). In

this study, the level of question difficulty was quite varied, with a balanced proportion of easy, medium, and difficult questions. Questions with a balanced level of difficulty can provide better information about student abilities, because questions that are too easy or too difficult cannot distinguish student abilities properly.

Question discrimination is also an important aspect in measuring question quality. The discriminating power of questions measures the ability of questions to distinguish between high and low ability students. In this study, most questions have good and excellent differentiating power, which means that these questions can clearly distinguish between students who master the material and those who do not master the material (Isnaintri et al., 2023).

However, there were some questions that were invalid and had poor discriminating power. These questions need to be revised or replaced with better questions in order to measure students' literacy and numeracy skills more accurately. In addition, it is necessary to further analyze the factors that affect the quality of the questions, such as question construction, tested materials and student characteristics (Manarisip et al., 2023).

Question construction is one of the factors that can affect question quality. Problems that are not well constructed can cause misunderstanding or bias in answering questions, which can affect the validity and differentiating power of the questions. Therefore, in developing questions, it is necessary to pay attention to aspects such as clarity of sentences, use of language that is appropriate for the level of understanding of students, and accuracy in choosing keywords.

The material being tested can also affect the quality of the questions. Questions that are not in accordance with the material taught or not in accordance with the ability level of students can cause questions to be too difficult or too easy, which can affect the validity and differentiating power of the questions. Therefore, in developing questions, it is necessary to pay attention to the suitability of the questions to the material taught and the ability level of the students.

Student characteristics can also affect question quality. Questions that are not in accordance with student characteristics, such as cultural background, language, or experience, can cause misunderstanding or bias in answering questions, which can affect the validity and differentiating power of the questions. Therefore, in developing questions, it is necessary to pay attention to the characteristics of students so that the questions can be well understood by students.

In addition, it should also be noted that Classical Test Theory (CTT) has several limitations in analyzing question quality. CTT assumes that a student's total score is the sum of item scores, so it does not consider individual student ability (Siregar & Panjaitan, 2022). Therefore, in future research, it can be considered to use other approaches, such as Item Response Theory (IRT), which can provide more detailed information about students' abilities and item characteristics.

Research Albikawi & Abuadas (2021) from using a combination of Classical Test Theory and Rasch Analysis (one of the IRT approaches) to develop and validate quality of life and self-stigma measurement instruments for caregivers of schizophrenia patients. The results showed that the instrument developed had an acceptable level of reliability and validity. Research (Wan et al., 2022) from also used a combination of Classical Test Theory and Generalizability Theory to develop and validate a quality of life measurement instrument for Chronic Obstructive Pulmonary Disease (COPD) patients. The results showed that the instrument developed had good validity, reliability, and responsiveness.

In the context of this study, the use of a combination of Classical Test Theory and other approaches, such as Item Response Theory or Generalizability Theory, can provide more comprehensive information about the quality of the end-of-semester test questions for grade 6 literacy and numeracy at SD Negeri 3 Pojok, Tawangharjo sub-district. These approaches can help identify more specific characteristics of questions, such as the difficulty level of questions for each student ability level, as well as provide information about the consistency and generalizability of measurement results.

Overall, the results of this study make an important contribution in analyzing the quality of literacy and numeracy end-of-semester exam questions in elementary schools using Classical Test Theory. The results can be used as evaluation and improvement materials in the preparation of future exam questions, so that they can measure students' literacy and numeracy skills more accurately and validly.

The implication of this research finding is the need to improve the quality of literacy and numeracy end-of-semester exam questions in primary schools. Teachers and schools can utilize the results of this question quality analysis to improve questions that are invalid or have poor differentiating power. In addition,

it is necessary to conduct training or workshops for teachers on how to develop quality questions, taking into account aspects such as validity, reliability, difficulty level and question differentiation.

Further research can be conducted using other approaches, such as Item Response Theory or Generalizability Theory, to provide more comprehensive information about question quality. In addition, research can also be conducted at different levels of education or in other subjects to expand the scope of question quality analysis.

By having quality test questions, students' literacy and numeracy skills can be measured more accurately and validly. This can provide useful information for teachers and schools in planning and implementing more effective learning, as well as helping students to improve their literacy and numeracy skills.

CONCLUSION

This study aims to analyze the quality of literacy and numeracy end-of-semester exam questions in grade 6 of SD Negeri 3 Pojok, Tawangharjo sub-district based on Classical Test Theory (CTT). Based on the results of the research and discussion, it can be concluded that the quality of the literacy and numeracy end-of-semester exam questions for grade 6 of SD Negeri 3 Pojok, Tawangharjo sub-district, is good based on CTT. Most of the questions have good validity, reliability, difficulty level, and differentiating power.

However, there are some questions that need to be revised or replaced to improve the overall quality of the questions. Further analysis of the factors that affect item quality, such as item construction, tested materials, and student characteristics, also needs to be conducted. In addition, using a combination of CTT with other approaches, such as Item Response Theory (IRT) or Generalizability Theory, can provide more comprehensive information about item quality.

This research makes an important contribution in analyzing the quality of literacy and numeracy end-of-semester exam questions in primary schools using CTT. The results can be used as evaluation and improvement materials in the preparation of future exam questions, so that they can measure students' literacy and numeracy skills more accurately and validly.

The implication of this research finding is the need to improve the quality of literacy and numeracy end-of-semester exam questions in primary schools. Teachers and schools can utilize the results of this question quality analysis to improve questions that are invalid or have poor differentiating power. In addition, it is necessary to conduct training or workshops for teachers on how to develop quality questions, taking into account aspects such as validity, reliability, difficulty level and question differentiation.

Future research can be conducted using other approaches, such as IRT or Generalizability Theory, to provide more comprehensive information about question quality. In addition, research can also be conducted at different levels of education or in other subjects to expand the scope of question quality analysis.

By having quality test questions, students' literacy and numeracy skills can be measured more accurately and validly. This can provide useful information for teachers and schools in planning and implementing more effective learning, as well as helping students to improve their literacy and numeracy skills.

LITERATURE

Albikawi, Z., & Abuadas, M. (2021). Quality of life and self-stigma of schizophrenia patient's caregiver tool: Development and validation using classical test theory and Rasch analysis. *South African Journal of Psychiatry*, *27*, 1–8. https://doi.org/10.4102/sajpsychiatry.v27i0.1656

Arbain, A., & Sirad, L. O. (2023). Menguatkan Resiliensi Matematis dan Literasi Numerasi Siswa Sekolah Dasar melalui Inovasi Pembelajaran Kontekstual dan Konstruktif. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(1), 908. https://doi.org/10.24127/ajpm.v12i1.6548

Barus, L. W., Ikhsan, R., Dewi, S. E., & Mujib, A. (2023). Meningkatkan Literasi Numerasi Siswa Sekolah Dasar Dengan Metode RME. *Elementar : Jurnal Pendidikan Dasar*, *3*(1), 29–35. https://doi.org/10.15408/elementar.v3i1.30338

Chauhan, G. R., Chauhan, B. R., Vaza, J. V, & Chauhan, P. R. (2023). Relations of the Number of Functioning Distractors With the Item Difficulty Index and the Item Discrimination Power in the Multiple Choice Questions. *Cureus*, *15*(7). https://doi.org/10.7759/cureus.42492

Fatimah, L. U., & Alfath, K. (2019). ANALISIS KESUKARAN SOAL, DAYA PEMBEDA DAN FUNGSI DISTRAKTOR. *Al-Manar*, *8*(2), 37–64. https://doi.org/10.36668/jal.v8i2.115

Fernanda, J. W., & Hidayah, N. (2020). Analisis Kualitas Soal Ujian Statistika Menggunakan Classical Test Theory dan Rasch Model. *Square : Journal of Mathematics and Mathematics Education*, *2*(1), 49. https://doi.org/10.21580/square.2020.2.1.5363

Handayani, S. L., & Iba, K. (2020). Karakteristik Tes Keterampilan Proses Sains: Validitas, Reliabilitas, Tingkat Kesukaran dan Daya Pembeda Soal. *Publikasi Pendidikan*, *10*(2), 100. https://doi.org/10.26858/publikan.v10i2.13051

Irwan, A., & Kerans, A. P. S. (2022). ANALISIS KUALITAS SOAL THREE TIER TEST MENGGUNAKAN TEORI RESPON BUTIR KLASIK. *JPF (Jurnal Pendidikan Fisika) Universitas Islam Negeri Alauddin Makassar*, *10*(2), 130–137. https://doi.org/10.24252/jpf.v10i2.31068

Isnaintri, E.-, Nindiasari, H.-, & -, S.-. (2023). Development of Numeracy Literacy Instrument in the Context of Local Wisdom in Pandeglang at the Madrasah Tsanawiyah. *Phenomenon : Jurnal Pendidikan MIPA*, *13*(1), 111–131. https://doi.org/10.21580/phen.2023.13.1.16783

Mamot, M., Mohamed, N. A., Mohd Rani, M. D., Osman, M., Mohd Zulkefli, N. A., Sekawi, Z., & Awang Hamat, R. (2022). Reliability and validity of knowledge, attitude, practice, and perception (KAP2) questionnaire on food poisoning and its prevention among consumers in Bangi and Kajang, Selangor, Malaysia. *International Food Research Journal*, *28*(6), 1204–1215. https://doi.org/10.47836/ifrj.28.6.12

Manarisip, J. T., Pulukadang, R. J., & Tumalun, N. K. (2023). ANALISIS KUALITAS SOAL BUATAN GURU DI SMP NEGERI 9 SATAP LANGOWAN PADA MATA PELAJARAN MATEMATIKA. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *4*(1), 629–639. https://doi.org/10.46306/lb.v4i1.310

Murtafiah, W., Krisdiana, I., Nur Fitria, R., Puspita Ningrum, P., & Fajar Subeqi, E. (2023). PENDAMPINGAN GURU SEKOLAH DASAR UNTUK PENGUATAN LITERASI NUMERASI SISWA MELALUI PROYEK KOLABORASI. *PATIKALA: Jurnal Pengabdian Kepada Masyarakat*, *2*(3), 694–703. https://doi.org/10.51574/patikala.v2i3.722

Nurul Muchlizani A, Sitti Mania, & Muhammad Nur Akbar Rasyid. (2023). ANALISIS KUALITAS BUTIR SOAL UJIAN AKHIR SEMESTER MATA PELAJARAN AKIDAH AKHLAK KELAS V MI RADHIATUL ADAWIYAH MAKASSAR. *Inspiratif Pendidikan*, *12*(1), 224–240. https://doi.org/10.24252/ip.v12i1.32040

Sesanti, N. R., Wahyuningtyas, D. T., & Marsitin, R. (2023). Pengembangan e-Modul Bilangan Berbasis Somatic, Auditory, Visual, Intelektual (SAVI) untuk Meningkatkan Literasi Numerasi Siswa Sekolah Dasar. *At-Thullab : Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, *7*(2), 160. https://doi.org/10.30736/atl.v7i2.1425

Siregar, H. N. I., & Panjaitan, A. (2022). A Systematic Literature Review : the Role of Classical Test Theory and Item Response Theory in Item Analysis to Determine the Quality of Mathematics Tests. *Jurnal Fibonaci: Jurnal Pendidikan Matematika*, *2*(2), 29. https://doi.org/10.24114/jfi.v2i2.31056

Wan, C., Yang, Z., Zhao, Z., Quan, P., Wu, B., & Yang, Y. (2022). Development and preliminary validation of the chronic obstructive pulmonary disease scale quality of life instruments for chronic diseases-chronic obstructive pulmonary disease based on classical test theory and generalizability theory. *Chronic Respiratory Disease*, *19*, 1–11. https://doi.org/10.1177/14799731221104099