

Summative Assessment: Quality Test of Cognitive Instruments for the Subject "Bahasa Indonesia"

Devi Yansah¹, Harry Soedarto Harjono², Mukhlas Abrar³, Rustam⁴

^{1,2,3,4}Bahasa dan Sastra Indonesia Department, Post Graduate, Universitas Jambi-Indonesia
Email: deviyansahunja1512@gmail.com

ABSTRACT

Positive assessment instruments are one of the most important parts in measuring the quality of the learning process and objectives. The research aims to test the quality of cognitive instruments for Indonesian language subjects in the summative assessment for the odd semester 2023/2024 at SMP Negeri 11 Jambi City. This research uses a quantitative approach, with data collection techniques through field tests. Instrument quality testing is carried out with the help of ANATES software which includes validity, reliability and difficulty level tests. The research results show that in general the cognitive summative test instruments in the "Indonesian" subject do not meet the elements of good quality. From the validity aspect, questions in the "very valid" and "valid" categories amounted to 29 items or around 54%, and questions in the "less valid" and "invalid" categories totaled 21 items or around 52%. From the reliability aspect, the categories "very high", "high" and "medium" amounted to 36 items or around 72%, and questions that were not of good quality, namely in the "low" and "very low" categories, amounted to 14 items or around 28%. From the aspect of difficulty level, in general the instrument has an even distribution, however, this distribution is less than ideal so it still needs to be refined. This research has implications for the need to strengthen the quality of cognitive assessment instruments as an alternative to improving the quality of "Indonesian" learning.

Keywords: *instrument quality, cognitive assessment, Bahasa Indonesia.*

1. INTRODUCTION

Education is a process to prepare quality future generations through coaching, learning and mentoring activities so that they are able to keep up with the times. Diana & Afendi (2023) explain that education is a form of human capital that keeps pace with developments in increasingly advanced and complex times. Sujana (2019) states that education is a continuous and never-ending process, so that it can produce sustainable quality, which is aimed at realizing the human figure of the future. To fulfill life's needs, including keeping up with the times, everyone needs education (Uddin, 2010). Therefore, it can be understood that the educational process continues to progress along with the times.

In Indonesia, one of the subjects that is mandatory for all students at the junior high school level is "Bahasa Indonesia". In principle, Indonesian language lessons at school are essentially teaching children to be able to communicate using Indonesian (Suparlan, 2020). In more detail, Hanna (2014) explains that the objectives of learning "Bahasa Indonesia" and Literature in general should include (1) students respecting and being proud of Indonesian as a unified (national) language and state language, (2) students understanding Indonesian in terms of form and meaning, and function, and use it appropriately and creatively for various purposes, needs and circumstances, (3) students have the ability to use Indonesian to improve intellectual abilities, emotional maturity and social maturity, (4) students have discipline in thinking and language (speaking and writing), (5) students are

able to enjoy and utilize literary works to develop their personality, broaden their outlook on life, and improve their knowledge and language skills, and (6) students appreciate and take pride in Indonesian literature as a cultural and intellectual treasure of Indonesian people. Thus, it can be understood that the subject "Bahasa Indonesia" is a means of improving the quality of students' thinking and so that they are able to apply and appreciate Indonesian as a form of national culture.

To find out and ensure that the coaching, learning and mentoring process for students occurs in a quality manner, the process needs to be assessed or assessed. Assessment has an important role in the education and learning process. Zaimul (2018) and Yambi (2020) stated that one of the important parts in implementing learning activities is the assessment technique for learning outcomes. According to Adinda et al. (2021) every learning process requires assessment to ensure that learning objectives are achieved, assessment is also needed as a form of teacher accountability. Apart from that, Barokah (2019) also stated that assessment is one of the most important aspects of teacher success in implementing learning. This activity is one of the four main tasks of teachers, namely planning, implementing, and assessing the success of teaching and providing guidance. A teacher is required to master the ability to provide assessments to determine the abilities that his students have mastered. Therefore, it can be understood that assessment is an important and strategic part of the quality of the learning process and the achievement of educational goals, both generally and specifically.

Conceptually, Ifat (2015), Box (2018) and Fry (2019) explain that assessment is the process of interpreting various information systematically, periodically, continuously and comprehensively about the processes and results of growth and development that have been achieved by students through activities learning and interpreting that information to make decisions. Meanwhile, Hafidhoh et al. (2021) and Looney (2009) state that assessment is a series of activities to obtain, analyze and interpret data about the student learning outcomes process which is carried out systematically and continuously so that it becomes meaningful information in making decisions. The conceptual definition of assessment above is also reinforced by Fetrianto (2017) that program evaluation is a series of activities carried out deliberately and carefully to determine the level of implementation or success of a program by knowing the effectiveness of each component, both for ongoing programs and programs which has passed. Considering that assessment is part of evaluation activities, conceptually, both can be used to determine the quality of the learning process.

According to Zaimul (2018) assessment has several types, namely formal and informal assessment, class assessment, diagnostic assessment, formative assessment, and summative assessment. According to Adinda et al. (2021) summative assessments are usually carried out at the end of the learning process with the aim of finding out how much students have achieved from the studies they have completed. The explanation above is also in line with the opinion of Fetrianto (2017) that the purpose of summative evaluation is to measure program achievements. Therefore, one of the functions of summative evaluation in evaluating learning programs is intended as a means of knowing the individual's position or position in the group. The explanation of the general purpose of summative assessment is also strengthened by Gaspersz et al. (2023) that assessment aims to determine the level of success of students. Thus, summative assessment is an important factor and must be considered to determine the success of students in participating in the learning process.

In the Indonesian curriculum, students' success in following the learning process is measured in three domains, namely affective, psychomotor and cognitive. In more detail, Prastiwi et al. (2023) explains that learning outcomes are achievements obtained by students academically by doing assignments, exams, and being active in asking questions and answers, which can generally be grouped into three, namely cognitive, affective, and psychomotor. Apart from that, Zakiah & Khairi (2019) also stated that changes that occur in humans include changes in physical aspects (motor), emotions, cognition, psychosexual, psychosocial, moral, and others. The development of each aspect runs simultaneously in harmony following the rhythm of individual development. Of the three aspects of the learning outcomes domain, the cognitive domain is often used as the main reference in determining students' success in participating in learning. In various schools, even in many countries, students' success in participating in learning is also measured from the aspect of cognitive ability.

Why the cognitive abilities are considered the most important? Cognitive abilities are considered the most important, because they involve the ability to think and mental processes simultaneously and these abilities are

considered to represent and influence all human actions. Zakiah & Khairi (2019) and Grageda et al. (2022) states that cognitive is a mental process related to abilities in the form of general recognition which is mental in nature and is characterized by the representation of an object in a person's mental image whether in the form of symbols, responses, ideas, and values or considerations. Marinda (2020) and Pakpahan and Saragih (2022) also explain that cognitive development is the stages of change that occur in the span of human life to understand, process information, solve problems and know things.

Apart from that, Cahyaningsih et al. (2019) and Teng (2023) explain that the goals of developing thinking (cognitive) power in detail are (a) developing logical thinking abilities and knowledge of space and time, (b) children are able to develop knowledge they already know with the new knowledge they have acquired, (c) develop the ability to understand something by seeing various relationships between one object and another object based on differences and similarities, (d) develop imagination through various activities, (e) provide opportunities to actively process the environment and build the world, and (f) so that children can appreciate and love the contents of nature as God's creation. Therefore, it can be understood that cognitive factors have an important role in successful learning, because most activities in learning are always related to remembering and thinking.

To ensure that the assessment process in the cognitive domain is accurate, quality instruments are needed. The quality of an instrument for assessing learning outcomes - including in the cognitive domain - is largely determined by the level of validity, reliability, level of difficulty and so on. Determining the quality of the instrument can be done using quantitative analysis. According to Khaerudin (2015), quantitative analysis is carried out by testing instruments that have been analyzed qualitatively on a number of students who have the same characteristics as the students who will be tested with the instrument. Furthermore, Khaerudin (2015) stated that the quantitative analysis was intended to determine the validity, reliability, distinguishing power, level of difficulty and effectiveness of the distractor function. Apart from that, Maulana (2023) also explained that to determine the quality of an assessment instrument it is necessary to carry out a series of testing activities for validity, reliability, distinguishing power and level of difficulty. The current aim of this activity is to identify which questions fall into the good, poor and bad categories, so that they can be used as a basis for determining the weaknesses or shortcomings of the tests that have been prepared and can be followed up by making improvements.

The urgency of quality testing of test instruments is not yet fully understood by "Indonesian" teachers, especially at SMP Negeri 11 Jambi City. Analysis of test instruments so far has not yet reached quantitative analysis, so the quality of the instruments has not been fully tested empirically. Research on validity, reliability and level of difficulty has not been fully carried out in this school, so improvements and refinements to assessment instruments, especially the "Indonesian" subject in the cognitive domain, have not run optimally. The development of test instruments in each field of study has so far been the teacher's responsibility and there has been no adequate quality control system. Therefore, the research aims to carry out a quantitative analysis of the "Bahasa Indonesia" subject questions in the cognitive domain of the summative test in the first semester of the 2023/2024 academic year.

2. RESEARCH METHODS

This research uses a quantitative descriptive approach, namely trying to find out the value of independent variables, either one or more variables (independent) without making comparisons, or connecting them with other variables (Iyus, 2020). The data collection technique is through field or empirical tests. Therefore, the validity data of the cognitive summative test instrument in this research is empirical data obtained after the questions were tested on students. According to Selaras et al. (2019), this data was obtained from the results of analysis of the answer sheet questions that were filled in by students.

The subjects in this research were 27 students at SMP Negeri 11 Jambi City. The object of the research is the test instrument (questions) for the subject "Bahasa Indonesia" which is in the form of multiple choice questions, totaling 50 questions. The test instrument (questions) for the "Indonesian Language" subject being tested are questions that have been developed by field teachers and used in the summative test for the odd semester of the 2023/2024

academic year. Instrument quality testing is carried out with the help of ANATES software which includes validity, reliability and difficulty level tests

Decision making on each test instrument quality indicator is based on Rizki's (2017) opinion, namely as follows:

a. Validity Category

The first test of the quality of the "Bahasa Indonesia" subject instrument at SMP Negeri 11 is validity. Decision making regarding the level of validity is as follows in Table 1

Table 1. Basic table for validity decision making

No	Validity Category	Criteria
1	Very Valid	$3 \leq VR \leq 4$
2	Valid	$2 \leq VR < 3$
3	Less Valid	$1 \leq VR < 2$
4	Invalid	$0 \leq VR < 1$

Referring to Table 1, instruments or cognitive summative questions are said to be of quality if they have the categories "very valid" and "valid", that is, they have the criteria $3 \leq VR \leq 4$ and $2 \leq VR < 3$. As for questions that have the categories "less valid" and "invalid" or having criteria $0 \leq VR < 2$ is considered not of good quality, so it must be corrected.

b. Reliability Category

The next test for the quality of the "Bahasa Indonesia" subject instrument at SMP Negeri 11 is reliability. Decision making regarding reliability is as shown in Table 2 below

Table 2. Basis for Reliability Decision Making

No	Reliability Category	Criteria
1	Very high	$0,80 < r_{11} \leq 1,00$
2	Tall	$0,60 < r_{11} \leq 0,80$
3	Currently	$0,40 < r_{11} \leq 0,60$
4	Low	$0,20 < r_{11} \leq 0,40$
5	Very low	$0,00 < r_{11} \leq 0,20$

Referring to Table 2 above, instruments or questions said to be of good quality and can be used in summative tests are questions with reliability categories of "very high", "high" and "medium". Thus questions with the criteria $0.40 < r_{11} \leq 1.00$ are said to be good or reliable and can be used. Questions with criteria of $0.0 < r_{11} \leq 0.60$ or questions with reliability in the "low" and "very low" categories are bad questions and cannot be used in summative tests.

c. Difficulty Level Category

The next test for the quality of the instrument for the subject "Bahasa Indonesia" at SMP Negeri 11 is the level of difficulty. Decision making regarding the level of difficulty is as follows in Table 3:

Table 3. Basis for decision making on Difficulty Level

No	Category	Difficulty Index
1	Difficult	Less than 0,30
2	Medium	0,30 - 0,70
3	Easy	More than 0,70

For the level of difficulty, an instrument is said to be good if it has an even level of difficulty, with a ratio of 25% in the "difficult" category, 50% in the "medium" category, and 25% in the "easy" category. Therefore, in research the balance of distribution is the basis for making decisions.

3. RESULTS AND DISCUSSION

Validity Test

Based on the results of the quality test of the summative test instrument in the cognitive domain of the subject "Indonesian Language" using the ANATES application, each quality indicator, namely validity, reliability and level of difficulty, is systematically described descriptively. For validity testing, the results of quantitative analysis can be seen in Table 4 below:

Table 4. Validity Test Results of Cognitive Summative Instruments

No	Validity Category	Criteria	Frequency	Percentage
1	Very Valid	$3 \leq VR \leq 4$	12	24
2	Valid	$2 \leq VR < 3$	17	34
3	Less Valid	$1 \leq VR < 2$	18	36
4	Invalid	$0 \leq VR < 1$	3	6
Total			50	100

Referring to Table 4 above, it is known that of the 50 multiple choice questions in the "Bahasa Indonesia" subject, 12 questions or around 24% are in the "very valid" category. The questions in the "valid" category were 17 items or around 34, while the questions in the "less valid" category were 18 items or around 36%, and the questions in the "invalid" category were 3 questions or around 6%. This means that the questions that meet the quality criteria, namely questions in the "very valid" and "valid" categories, amount to 29 items or around 54%. Questions that were not of good quality were in the "less valid" and "invalid" categories as many as 21 items or around 52%. This means that in general, there are still many (52%) of the total questions developed and used in cognitive summative tests that do not meet the elements of a quality instrument and need to be improved.

Referring to the opinion of Puspasari and Puspita (2022) that the validity test aims to see the accuracy of the measurement. Slamet and Wahyuningsih (2022) stated that an instrument from a questionnaire is said to be valid if the instrument can accurately measure the object it wants to measure. Thus, a valid instrument produces valid or correct data. Therefore, referring to the results of the validity test of the cognitive summative test instrument on the subject "Indonesian" which was developed as in Table 4, there are 21 items in the "less valid" and "invalid" categories or around 52%. This means that in general the measurement results or learning outcomes data in the cognitive domain of students that are measured are less valid or do not reflect the truth.

Reliability Test

Apart from validity, quality questions have good reliability requirements. According to experts, there are five criteria for reliability categories, namely: "very high", "high", "medium", "low" and "very low". The results of the reliability test on the "Indonesian Language" subject questions at SMP Negeri 11 Jambi City using ANATES software can be seen in Table 5 below:

Table 5. Summary of Reliability Test Results

No	Validity Category	Criteria	Frequency	Percentage
1	Very high	$0,80 < r_{11} \leq 1,00$	2	4
2	Tall	$0,60 < r_{11} \leq 0,80$	19	38
3	Currently	$0,40 < r_{11} \leq 0,60$	15	30
4	Low	$0,20 < r_{11} \leq 0,40$	12	24

5	Very low	$0,00 < r_{11} \leq 0,20$	2	4
Total			50	100

Referring to Table 5 above, it is known that of the 50 multiple choice questions in the subject "Bahasa Indonesia", as many as 2 questions or around 4% are in the "very high" category. The questions that have reliability in the "high" category are 19 items or around 38%. There are 15 questions in the "medium" category or around 30%, and 15 questions in the "medium" category or around 30%. Questions with "low" category reliability were 12 items or around 24%, and questions in the "very low" category were 2 items or around 4%. This means that the questions that meet the quality criteria, namely questions with reliability in the "very high", "high" and "medium" categories, total 36 items or around 72%. Questions that were not of good quality were in the "low" and "very low" categories, totaling 14 items or around 28%. This means that in general, there are still many (28%) of the total questions developed and used in cognitive summative tests that do not meet the elements of a quality instrument and need to be improved.

Referring to the opinion of Puspasari and Puspita (2022), the reliability test aims to determine the extent to which the instrument can be trusted. Slamet and Wahyuningsih (2022) explained that reliability tests are used to determine the consistency of measuring instruments, whether the measuring instruments used are reliable and remain consistent if the measurements are repeated. A measuring instrument is said to be reliable if it produces the same results even if measurements are taken many times. Data from the reliability test using ANATES software shows that the questions that meet the quality criteria, namely questions with reliability categories "very high", "high" and "medium" total 36 items or around 72%. Questions that were not of good quality were in the "low" and "very low" categories, totaling 14 items or around 28%. This means that 28% of the total questions developed and used in cognitive summative tests are still not reliable and consistent in determining students' cognitive abilities, especially in the subject "Bahasa Indonesia".

Test the difficulty level

The third aspect measured in this research is the level of difficulty. According to experts, there are three categories of difficulty level of an assessment instrument, namely the "difficult", "medium" and "easy" categories. Instruments are said to be of quality if they have a balanced comparison of each. The results of the cognitive test questions on the difficulty level of the subject "Bahasa Indonesia" at SMP Negeri 11 Jambi City using ANATES software are summarized in Table 6 below:

Table 6. Summary of Test Difficulty Level

No	Validity Category	Difficulty Index	Frequency	Percentage
1	Hard	Less than 0,30	8	16
2	Currently	0,30 - 0,70	31	62
3	Easy	More than 0,70	11	22
Total			50	100

Based on the data in Table 6, it can be seen that the bad luck for the summative cognitive test, questions in the "difficult" category or having an index of less than 0.3 are 8 items or around 16%. Questions in the "medium" category or with a difficulty index between 0.3 and 0.7 are 31 items or around 62%. Meanwhile, questions in the "easy" category or with a difficulty index of more than 0.70 were 11 items or around 22%. Based on the distribution of this data, it can be understood that in general the instrument has an even distribution. However, this distribution is less than ideal, so it still needs to be refined.

Referring to the results of the difficulty level test shows that in general the questions above do not meet the elements of a quality instrument. According to Lumbanraja and Daulay (2018) that if a question has a balanced (proportional) level of difficulty, then it can be said that the question is good, and vice versa. Apart from that, referring to the opinion of Hanifah et al. (2014) that the difficulty level of a question item is the proportion between the number of test takers who answered the question item correctly and the number of test takers. Therefore, the data in Table 6 shows that the proportion of instruments used in the cognitive summative test is not optimal.

Referring to the results of the research above (test validity, reliability and level of difficulty), in general it can be said that the cognitive summative test instruments in the "Bahasa Indonesia" subject generally do not meet the elements of good quality. With this situation, the measurement data using this instrument is also not convincing. Therefore, planned and systematic steps need to be taken to improve the instrument. According to Hanifah et al. (2014) six requirements for a good test, namely (1) the test must be reliable, (2) the test must be valid, (3) the test must be objective, (4) the test must be discriminatory, (5) the test must be comprehensive, (6) the test must be easy used. Thus, the first step that must be taken is to improve the instrument so that it is better and suitable for use.

4. CONCLUSION

Referring to the results of the research above (test validity, reliability and level of difficulty), in general it can be said that the cognitive summative test instruments in the "Bahasa Indonesia" subject generally do not meet the elements of good quality. From the validity aspect, questions in the "very valid" and "valid" categories amounted to 29 items or around 54%, and questions in the "less valid" and "invalid" categories totaled 21 items or around 52%. From the reliability aspect, the categories "very high", "high" and "medium" amounted to 36 items or around 72%, and questions that were not of good quality, namely in the "low" and "very low" categories, amounted to 14 items or around 28%. In terms of level of difficulty, in general the instrument has an even distribution. However, this distribution is less than ideal, so it still needs to be refined.

5. REFERENCES

- Adinda, A. H., Siahaan, H. E., Raihani, I. F., Aprida, N., Fitri, N., & Suryanda, A. (2021). Penilaian Sumatif dan Penilaian Formatif Pembelajaran Online. *Report Of Biology Education*, 2(1), 1–10.
- Barokah. (2019). Manajemen Penilaian Sumatif Pada Ranah Kognitif Pembelajaran Pai Kelas X Semester Ganjil Di Sma Negeri 2 Pontianak Tahun Pelajaran 2017/ 2018. *Al-Idarah: Jurnal Kependidikan Islam* V, 9(2), 159–179.
- Box, C. (2018). Formative Assessment in United States Classrooms: Changing the Landscape of Teaching and Learning. *Formative Assessment in United States Classrooms: Changing the Landscape of Teaching and Learning*, 1–181. <https://doi.org/10.1007/978-3-030-03092-6>
- Cahyaningsih, K. W., Fajar utami, D. D., & Rohmalina, R. (2019). Meningkatkan Kemampuan Kognitif Anak Melalui Pembelajaran Tematik Dengan Metode Eksperimen Bercocok Tanam Di Kelompok B. *CERIA (Cerdas Energik Responsif Inovatif Adaptif)*, 2(4), 115. <https://doi.org/10.22460/ceria.v2i4.p115-121>
- Diana, S. R., & Afendi, A. R. (2023). Urgensi Pendidikan Agama Islam Dalam Membentuk Karakter Pada Peserta Didik di SMPN 6 Tenggarong Seberang. *Jurnal Pendidikan Tambusai*, 7, 1897–1903.
- Fetrianto, F. (2017). Penerapan Formative Summative Evaluation Model Dalam Penelitian Tindakan. *Seminar Nasional Pendidikan Olahraga*, 1(1), 408–421.
- Fry, K. (2019). Developing and Assessing an Environmental Science-Based Education Module to Support the Nature of Science and Increase Science Literacy. *Proceedings of the ... National Conference on Undergraduate Research. National Conference on Undergraduate Research*.
- Gaspersz, M., AW, S., & Gaspersz, N. (2023). Model Evaluasi Formatif-Sumatif Terhadap Hasil Belajar Matematika Melalui Pembelajaran Berdiferensiasi Pada Peserta Didik Sma. *Jurnal Magister Pendidikan Matematika (JUMADIKA)*, 5(1), 1–7. <https://doi.org/10.30598/jumadikavol5iss1year2023page1-7>
- Grageda, C., Tinapay, A. O., Tirol, S. L., & Abadiano, M. N. (2022). Socio-Cultural Theory in the Cognitive Development Perspective. *NeuroQuantology*, 20(16), 1482. <https://doi.org/10.14704/NQ.2022.20.16.NQ880145>
- Hafidhoh, N., Muhammad, D., & Rifa', R. (2021). Karakteristik Penilaian Pembelajaran Pada Kurikulum 2013 Di Mi. *Awwaliyah: Jurnal PGMI*, 4(1), 10–16.

- Hanifah, N., Studi, P., & Konseling, B. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi*, 6(1), 41–55.
- Hanna. (2014). Pembelajaran Bahasa Indonesia Mau Dibawa Ke Mana? *Bahtera : Jurnal Pendidikan Bahasa Dan Sastra*, 3(2), 1–46.
- Ifat Fatimah. (2015). Penilaian dalam pembelajaran anak usia dini. *Tunas Siliwangi*, 1(1), 92–111.
- Iyus Jayusman. (2020). Studi Deskriptif Kuantitatif Tentang Aktivitas Belajar Mahasiswa Dengan Menggunakan Media Pembelajaran Edmodo Dalam Pembelajaran Sejarah. *Jurnal Artefak*, 7(1), 13–20.
- Khaerudin. (2015). Kualitas instrumen tes hasil belajar. *Jurnal Madaniyah*, 2, 212–235.
- Looney, J. W. (2009). Assessment And Innovation In Education. *OECD Education Working Papers*.
- Lumbanraja, L. H., & Daulay, S. (2018). Analisis Tingkat Kesukaran Dan Daya Pembeda Pada Butir Tes Soal Ujian Tengah Semester Bahasa Indonesia Kelas Xii Sma Negeri 7 Medan Tahun Pembelajaran 2016/2017. *Kode: Jurnal Bahasa*, 6(1), 15–24. <https://doi.org/10.24114/kjb.v6i1.10814>
- Marinda, L. (2020). Kognitif dan Problematika. *An-Nisa' : Jurnal Kajian Perempuan Dan Keislaman*, 13(1), 116–152.
- Maulana, H. (2023). Analisis Kualitas Instrumen Evaluasi Pembelajaran Menggunakan Media Digitalisasi Untuk Memotivasi Hasil Belajar Peserta Didik. *Bersatu: Jurnal Pendidikan Bhinneka Tunggal Ika*, 1(4).
- Pakpahan, F. H., & Saragih, M. (2022). Theory Of Cognitive Development By Jean Piaget. *JOURNAL OF APPLIED LINGUISTICS*, 2(2), 55–60. <https://doi.org/10.52622/joal.v2i2.79>
- Prastiwi, Y. E. N., Arba'iyah, Barru, A. A. Al, & Hidayatullah, A. S. (2023). Penilaian dan Pengukuran Hasil Belajar Pada Peserta Didik Berbasis Analisis Psikologi. *Bersatu: Jurnal Pendidikan Bhineka Tunggal Ika*, 1(4), 218–231.
- Puspasari, H & Puspita, W. (2022). Uji Validitas dan Reliabilitas Instrumen Penelitian Tingkat Pengetahuan dan Sikap Mahasiswa terhadap Pemilihan Suplemen Kesehatan dalam Menghadapi Covid-19 Validity Test and Reliability Instrument Research Level Knowledge and Attitude of Students Towards. *Jurnal Kesehatan*, 13, 65–71.
- Rizki Riyani. (2017). Uji Validitas Pengembangan Tes Untuk Mengukur Kemampuan Pemahaman Relasional Pada Materi Persamaan Kuadrat Siswa Kelas VIII SMP. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, 1(1), 60–65.
- Selaras, G. H., Ahda, Y., & Alberida, H. (2019). The Validity and Reliability of The Instrument Assessment of Higher Order Thinking Skill on The Biological Scope Materials. *Bioeducation Journal*, 3(2), 124–125.
- Slamet, R., & Wahyuningsih, S. (2022). Validitas Dan Reliabilitas Terhadap Instrumen Kepuasan Ker. *Aliansi : Jurnal Manajemen Dan Bisnis*, 17(2), 51–58. <https://doi.org/10.46975/aliansi.v17i2.428>
- Sujana, I. W. C. (2019). FUNGSI DAN TUJUAN PENDIDIKAN INDONESIA. *Adi Widya: Jurnal Pendidikan Dasar*. <https://doi.org/10.25078/aw.v4i1.927>
- Suparlan, S. (2020). Pembelajaran Bahasa Indonesia di Sekoah Dasar. *Fondatia*, 4(2), 245–258. <https://doi.org/10.36088/fondatia.v4i2.897>
- Teng, M. F. (2023). Metacognition. *Cognitive Individual Different in Second Language Accuision: Theories, Assessment, and Pedagogy, March*, 175–199. <https://doi.org/10.1515/9781614514749-009>
- Uddin, M. J. (2010). *Meaning , Purpose , function and Morality of Education : An Overview of National Education Policy-2010 of Bangladesh*.
- Yambi, T. de A. C. (2020). Assessment and Evaluation in Education EDUC 540 Spring 2020. *Researchgate*,

4(July), 12–34.

Zaimul. (2018). Teknik Penilaian Hasil Pembelajaran. *Rausyan Fikr : Jurnal Pemikiran Dan Pencerahan*, 14(02), 53–62. <https://doi.org/10.31000/rf.v14i02.901>

Zakiah, Z., & Khairi, F. (2019). Pengaruh Kemampuan Kognitif Terhadap Prestasi Belajar Matematika Siswa Kelas V Sdn Gugus 01 Kecamatan Selaparang. *El Midad*, 11(1), 85–100. <https://doi.org/10.20414/elmidad.v11i1.1906>

