Journal of Educational Research and Evaluation
Volume 5, Issue 3, 2021, pp. 422-428
P-ISSN: 2597-422x E-ISSN: 2549-2675
Open Access: https://ejournal.undiksha.ac.id/index.php/JERE

The Teacher-Made Multiple-Choice Tests Quality for Natural Science Subject

Erik Nanda Putra[1*], Syamsurizal[2], Lufri[3], Zulyusri[4]

[1,2,3,4] Pendidikan Biologi, Universitas Negeri Padang, Padang, Indonesia

ARTICLE INFO

Article history:
Received March 22, 2021
Revised March 25, 2021
Accepted July 24, 2021
Available online August 25, 2021

Kata Kunci :
Instrument Penilaian, Tes Pilihan Ganda, IPA

Keywords:
Assessment Instrument, Multiple Choice Test, IPA

ABSTRAK

Soal yang digunakan sebagai evaluasi cenderung lebih banyak menguji aspek memori sedangkan ranah kognitif, afektif, dan psikomotor dalam taksonomi Bloom belum optimal. beberapa guru IPA masih mengalami kesulitan dalam merumuskan soal, sehingga instrument yang dikembangkan oleh guru kurang valid. Tujuan penelitian yaitu menganalisis kualitas soal secara kuantitatif (reliabilitas, daya pembeda, tingkat kesukaran, dan keefektifan distraktor). Penelitian ini bertujuan untuk Penelitian ini merupakan penelitian deskriptif. Populasi dalam penelitian ini adalah seluruh lembar jawaban siswa pada Ujian Penilaian IPA. Teknik pengambilan sampel menggunakan total sampling dengan jumlah sampel 107 lembar jawaban siswa. Data dianalisis menggunakan ANATES ver. 4.0.2 program. Berdasarkan hasil penelitian diperoleh: 1) reliabilitas sebesar 0,50 dengan kriteria cukup; 2) daya pembeda kategori baik dan buruk yang mendominasi yaitu 30% sedangkan 25% sedang dan 7,5% sangat baik; 3) tingkat kesukaran yang diperoleh dengan kriteria sedang sebesar 62,5%, soal mudah sebesar 15%, soal sulit sebesar 12,5% dan soal sangat sulit sebesar 2,5%; 4) keefektifan pengecoh, ada 35 butir soal atau 87,5% keefektifan pengecoh yang berfungsi dengan baik. Sedangkan untuk 5 butir soal kualitas pengecoh tidak berfungsi atau efektivitasnya sangat buruk. Dengan demikian, dapat disimpulkan bahwa pertanyaan-pertanyaan tersebut telah memenuhi kriteria pertanyaan baik, namun beberapa pertanyaan perlu direvisi.

ABSTRACT

The questions used as evaluations test the memory aspects of the cognitive, affective, and psychomotor domains in Bloom's taxonomy which are not optimal. Some science teachers still have difficulty formulating questions, so the instruments developed by teachers are not valid. The research objective is to analyze the quality quantitatively (reliability, discriminatory power, difficulty level, and distractor effectiveness). This study aims to This research is a descriptive study. The population in this study were all student answer sheets on the Science Assessment Exam. The sampling technique used was total sampling with a sample of 107 student answer sheets. Data analysis using ANATES ver. 4.0.2 programs. The results based on the research obtained: 1) reliability of 0.50 with criteria; 2) the distinguishing power of good and bad categories that dominates is 30% while 25% as moderate and 7.5% is very good; 3) the level of difficulty obtained with moderate criteria is 62.5%, easy questions are 15%, difficult questions are 12.5% , and challenging questions are 2.5%; 4) the effectiveness of distractors, there are 35 items or 87.5% of effectiveness distractors that work well. As for the five items, the quality of the distractors does not work, or their effectiveness is terrible. Thus, it can be said that the questions met the criteria well, but some need to be revised.

1. INTRODUCTION

Learning is a process of student interaction with learning resources. In the learning process many factors influence how the learning process is meaningful. These factors include teachers, students, infrastructure, media, environment and so on (Binali et al., 2021; Mohd Zain et al., 2018; Purwanita et al., 2019). These factors must support each other so that quality and meaningful learning for students is realized (Marchand Martella et al., 2021; Silberman et al., 2021; Van Alten et al., 2020). With quality learning, it is expected that the learning objectives will be achieved by a student. The achievement of these learning objectives can be measured with measuring instruments in the form of tests, observation sheets and so on (Kruit et al., 2018; Leutner et al., 2017; Siddiq et al., 2019). The learning process includes activities that teachers do in planning, implementing evaluation and monitoring programs that occur in educational situations to achieve certain goals, namely teaching every day other than holidays, students carrying out learning activities at school (Rohita et al., 2018; Suryaningsih & Rimpiati, 2018). Learning is not non-stop, there is time for students to take a break from learning activities at school to find out what they have achieved from the learning process they are doing (Agung et al., 2017; Patria & Heswari, 2021). All students call it a test or exam and the teachers usually call it evaluation. In this case, educational institutions such as

*Corresponding author.
E-mail addresses: eriknandaputra18@gmail.com (Erik Nanda Putra)

schools play an important role in improving student achievement and the quality of their graduates (Dhika et al., 2019; Manikowati & Iskandar, 2018). Student learning achievement can be determined through measurement activities. Measuring learning achievement is an activity to quantify symptoms or objects in the form of achievement expressed in numbers (Heong et al., 2011; Puspitasari et al., 2019). Furthermore, the measurement results are processed into information used in determining policy. In other words, evaluation of learning outcomes requires data obtained through measurement activities (Basri, 2017; Yektyastuti & Ikhsan, 2016).

Ideally, the learning result instruments prepared by the teacher can provide a lot of information about students' abilities (Heong et al., 2011; Nurmala et al., 2014). However, many problems that arose when the teacher prepared this test instrument, it turned out that there were still many weaknesses in the preparation process, resulting in tests that did not have valid measuring power (Arif, 2016; Kuntoro & Wardani, 2020). Test instruments that do not have a valid measuring power will not be able to provide any information about the test taker's ability (Kuantum et al., 2018; Sudirman et al., 2020). The teacher as an evaluator must pay attention to the evaluation requirements in order to obtain the expected results in an effort to measure the level of mastery of each student on the material that has been taught. In conducting the evaluation, the teacher gives a test in accordance with the material that has been taught. Therefore, it is necessary to create a measuring tool to determine student success, that measurement tool is often called a test (teacher-made test) (Candra et al., 2018; Wulandari, Abadi, & Ganing, 2018). After making observations at SMP Negeri 2 Koto XI Tarusan, it was found that the questions used as evaluation tended to test more aspects of memory while the cognitive, affective, and psychomotor domains in Bloom's taxonomy were not optimal. The reason the researcher wanted to analyze the questions made by the biology teacher at SMP Negeri 2 Koto XI Tarusan was because based on the analysis of the questions made by the teacher, some science teachers at SMP Negeri 2 Koto XI Tarusan still had difficulty formulating questions, such as the questions made were the result of modified questions that had been read from the source. certain. Modifications made here in terms of reading text, images, and data or graphs supporting the questions, while the main formulation of the questions remains the same. Thus, it means that the teacher has not really formulated the questions on their own.

Teachers need an assessment instrument in the form of questions to carry out assessments, both for testing cognitive, affective, and psychomotor aspects (Maison et al., 2020; Simarmata et al., 2019). The assessment instrument used by the teacher to test student learning outcomes is usually taken from various books or a collection of exam questions (Imania & Bariah, 2019; Wulandari, Abadi, & Suniasih, 2018). Questions can be in the form of multiple choice or descriptions. But the reality in the field after observation, the questions tend to test more aspects of memory. Students must begin to be trained to think higher order (Anton et al., 2018; Sukmasari & Rosana, 2017). Training students to be skilled can be done by the teacher by practicing questions that invite students to think in the level of analysis, synthesis and evaluation (Arif, 2016; Pratiwi et al., 2017). Evaluation is the determination of the extent to which the ability that can be achieved by students in the learning process which has previously been determined by a benchmark value, so that an assessment can be carried out. Assessment of learning outcomes is carried out by the teacher on an ongoing basis, aiming to monitor the learning process and progress learners are invited to increase the effectiveness of learning activities (Kuantum et al., 2018; Kuntoro & Wardani, 2020). The results of the assessment will be used by teachers and students as feedback to determine the achievement of learning objectives. Middle Semester Assessment (MSA) is one of the benchmarks to determine the ability of students in understanding the material of a subject. The achievement test of learning outcomes is a test used to reveal the level of achievement of students (Sudirman et al., 2020; Wulandari, Abadi, & Ganing, 2018). Through MSA questions, the teacher will find it easier to know or measure the level of ability of their students. The quality of MSA items will greatly affect the information obtained by the teacher about the abilities of their students, because good quality questions will provide more accurate information to the teacher. A good test, will be able to reveal the true state of the student's ability (Candra et al., 2018; Maison et al., 2020). Previous findings stated that teachers have necessary instruments in measuring students' abilities (Simarmata et al., 2019; Wulandari, Abadi, & Suniasih, 2018). Other findings also state that the assessment instrument can measure students' abilities (Imania & Bariah, 2019; Wicaksono et al., 2016). It makes the instrument very important for teachers to have. The purpose of this study was to determine the quality of multiple-choice tests made by teachers for science subjects quantitatively in terms of reliability, discriminating power, difficulty level, and distractor effectiveness.

2. METHODS

Type of research is descriptive which systematically describes the analysis of the quality of the items. The approach used is a quantitative approach because the results of the item quality analysis are

carried out systematically, planned, and clearly structured and in this study, numbers are obtained from statistical calculations (Sudijono, 2008). As for the subjects in this study were students of class VIII SMP Negeri 2 Koto XI Tarusan academic year 2020/2021, totaling 40 students. This sample selection technique uses saturated sampling technique. Another term saturated sample is census, where all members of the population are sampled. Samples are also often interpreted as being the maximum sample, plus no amount will change the representativeness (Sudijono, 2008). The data analysis technique used is descriptive statistics, namely statistics that have the task of organizing and analyzing numerical data in order to provide a regular, concise and clear picture of a symptom, event or situation, so that certain definitions or meanings can be drawn. Descriptive statistics to analyze the items were carried out using the Anates program developed by (Adjii, 2019).

3. RESULT AND DISCUSSION

Results

Reliability comes from the word Rely which means trust and reliable, which means trustworthy. Trustworthiness is about constancy and consistency. Reliability relates to the accuracy of the instrument in measuring what is measured, the accuracy of the measurement results and how accurate if the measurement is repeated. Reliability relates to the ability of the measuring instrument to take measurements carefully. Reliability is the accuracy and precision produced by measuring instruments in making measurements. The results of the analysis show that the reliability value of all 40 items is 0.50 with sufficient criteria. Sufficient criteria indicate that the questions have met the reliability that is good enough or can be trusted in measuring the consistency of student abilities. The high reliability in question includes the accuracy or accuracy of the measurement results and the consistency or stability of the measurement results. The test will be said to have reliability if the test is made to have consistent results in measuring what it wants to measure. This shows that the test can be used. However, after reducing the 7 omitted items, the reliability value increased to 0.69 so that it was included in the high category. The high reliability referred to in this case includes the accuracy or accuracy of the measurement results and the consistency or stability of the measurement results.

Disciminating Power is the ability of items to discriminating high-ability students (understand the material) from students with low abilities (lack of understanding of the material). The results of the analysis of the distinguishing power of MSA II questions in science subjects class VIII SMP 2 Koto XI Tarusan in the 2020/2021 school year as a whole resulted in the bad and good distinguishing power that dominates, namely 30% while 25% is moderate and 7.5% is very good. The results of the analysis of the full distinguishing power of the questions can be seen in Table 1.

Table 1. Distribution of MSA II Questions for Natural Science Subjects class VIII SMP Negeri 2 Koto XI Tarusan Academic Year 2020/2021.

No	Discriminating Power	No Item	Amount	Percentage (%)
1.	<0 (Minus)	18, 19, 28.	3	7,5
2.	0,00-0,19 (Bad)	2, 3, 5, 10, 12, 13, 16, 27, 29, 33, 34, 35.	12	30
3.	0,20-0,29 (Average)	1, 17, 20, 21, 22, 23, 25, 36, 37, 40.	10	25
4.	0,30-0,49 (Good)	4, 6, 7, 8, 11, 14, 15, 24, 26, 30, 31, 32.	12	30
5.	0,50-1,00 (Very Good)	9, 38, 39.	3	7,5

In a calculation problem, the difficulty level of the problem is a measure of how much difficulty the problem is. If a problem has a balanced (proportional) level of difficulty, it can be said that the question is good. The level of difficulty of the questions is the opportunity to answer correctly a question at a certain ability level which is usually expressed in the form of an index. The difficulty index can be defined as the proportion of test takers who answered correctly. The quality or not of the items in the learning outcome test items can first be seen from the degree of difficulty or level of difficulty possessed by each of these items. The items of the learning outcome test items can be stated as good items, if the items are not too difficult and not too easy, in other words, the degree of difficulty of the item is moderate or sufficient. The results of the analysis of the difficulty level of MSA II questions in science subjects class VIII SMP Negeri 2 Koto XI Tarusan Academic Year 2020/2021 as a whole, it was obtained that the dominating questions were questions with moderate criteria with 62.5%, while easy questions were 15%, difficult questions 12.5 % and very difficult questions 2.5%. Effectiveness Based on the effectiveness of the distractor, it can be determined whether the distractorr can function properly or not. Distractors function well when at least

5% of all test takers are selected alternative answers. Based on the results of calculations and interpretations, there are 35 items or 87.5% of the effectiveness of distractors functioning properly. Meanwhile, for 5 items, the quality of the distractor does not function or has very poor effectiveness. The cheater that does not function properly indicates that the cheater is too conspicuous and heterogeneous, so that the cheater does not have the appeal of being selected by test takers who do not understand the material. The full quality analysis results can be seen in Table 2.

Table 2. Distribution of MSA II Questions for Natural Science Class VIII SMP Negeri 2 Koto XI Tarusan Academic Year 2020/2021.

No	Distractor Effectiveness Item	No Item	Amount	Percentage (%)
1.	Works	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 22, 24, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40.	35	87,5
2.	Doesn't Work	17, 21, 23, 25, 28	5	12,5

Discussion

The learning evaluation process certainly cannot be separated from the assessment instrument as an evaluation tool to measure the extent to which students understand the material that has been delivered by the teacher (Harta et al., 2020; Sagala & Andriani, 2019). Assessment instruments can take the form of tests, questionnaires, interviews and so on. The assessment process in schools generally uses tests, both written tests and oral tests (Praslova, 2010; Siddiq et al., 2019). Assessment in learning has a close relationship with the objectives set in the curriculum which can be used as a tool by the teacher to know and understand the success rate of students in absorbing the material provided and as a material for consideration for determining further educational programs (Ardianti et al., 2019; Leutner et al., 2017). To assess the value of out comes learning as student achievement, teachers can use tests. The test is a measuring tool, usually used to measure the success of students in gaining abilities. Seeing the importance of a test in an effort to measure student learning outcomes, the teacher should design a test of learning outcomes well and in accordance with the learning objectives (Maulida et al., 2020; Nurtanto et al., 2020). Thus the test items must be parallel with the formulation of learning outcomes indicators. With good tests, it is hoped that students will also get good learning results and can be used as a measure of success in the world of education (Siddiq et al., 2019; Wei et al., 2021). Evaluation Process aims to assess the effectiveness and efficiency of learning activities as materials for improvement and perfection of the implementation program (Leutner et al., 2017; Suratmi et al., 2020). The main purpose of conducting learning evaluations is to obtain accurate information about the level of achievement of learning objectives by students (Anton et al., 2018; Heru Purnomo & Wilujeng, 2016). A good evaluation must be based on the learning objectives set by the teacher which then really strive for achievement by the teacher and students (Candra et al., 2018; Sukmasari & Rosana, 2017). Learning evaluation related to education with learning objectives, as well as educational goals. Evaluation cannot be separated from the process of teaching and learning activities because evaluation is part of the process of teaching and learning activities (Maison et al., 2020; Simarmata et al., 2019).

The low reliability of a test can occur due to several factors. Factors that may affect the level of reliability of the questions are the number of students, and the ability of each student. These two factors indicate that for reliability, the more students and the higher the student's ability will support the reliability value of the question itself. There are several factors that affect reliability and one of them is the one stated by the researcher above, namely the higher the student's ability, the higher the reliability and the greater the number of test participants the greater the reliability (Khaerunnisa & Pamungkas, 2018; Lestari & Harjono, 2021). In addition, there are also several factors that affect reliability, namely the length of the test related to the number of questions and subjective evaluations that will reduce the reliability value, related to the implementation of the test and the evaluation of the teacher himself (Subagia & Wiratma, 2016; Utami & Wardani, 2020). Items with excellent, good and sufficient discriminating power must be maintained by including them in the question bank, while items with very poor and poor discriminating power must be completely repaired by tracing the causes of the failure. One of the efforts that can be done is to fix questions that are not clear in meaning so that they confuse high-ability students in answering them (Dessiane & Kristin, 2021; Srirahayu & Arty, 2018). The items must be able to reflect the difference in ability between students who understand the material and students who do not understand the material (H Purnomo & Wilujeng, 2016; Subagia & Wiratma, 2016).

4. CONCLUSION

Based on the results of the discussion, it can be concluded that the quality of the multiple choice test made by the teacher for science subjects in terms of the reliability value, discriminating power, difficulty level, and distractor effectiveness, namely for the reliability value of the questions used is included in the sufficient criteria, the discriminating power of the questions is included good criteria, the difficulty level of the questions with good criteria, and the distractor effectiveness is functioning properly. However, thus there are still questions that have to be revised again to function properly and questions that cannot be used must be discarded and replaced with better questions so that the student assessment instrument functions properly and the tests given to students really do. can measure students' ability to understand what has been previously learned.

5. REFERENCES

Adjii, K. (2019). Instrumen penilaian kedisiplinan siswa sekolah menengah kejuruan. *Assessment and Research on Education*, *1*(1). https://doi.org/10.33292/arisen.v1i1.19.

Agung, A. A. G., Widiana, I. W., & Indrasuari, N. K. S. (2017). Pengembangan Aktivitas Pembelajaran Mengasosiasi Berbasis Media Gambar Berseri Dalam Meningkatkan Proses Kognitif Siswa. *Jurnal Ilmiah Sekolah Dasar*, *1*(3), 138. https://doi.org/10.23887/jisd.v1i3.10323.

Anton, K., Sahri, J., & Tohidin, D. (2018). Pengembangan Instrumen Tes Kelentukan Statis Berbasis Teknologi Sensor. *Jurnal Sosioteknologi*, *17*(3), 343–349. https://doi.org/10.5614/sostek.itbj.2018.17.3.1.

Ardianti, S. D., Wanabuliandari, S., Saptono, S., & Alimah, S. (2019). A needs assessment of edutainment module with ethnoscience approach oriented to the love of the country. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 153–161. https://doi.org/10.15294/jpii.v8i2.13285.

Arif, M. (2016). Pengembangan Instrumen Penilaian Mapel Sains melalui Pendekatan Keterampilan Proses Sains SD/MI. *Ta'allum: Jurnal Pendidikan Islam*, *4*(1). https://doi.org/10.21274/taalum.2016.4.1.123-148.

Basri, I. (2017). Evaluasi Pembelajaran Sekolah Dasar (SD) Berbasis Pendidikan Karakter dan Multikultural. *Jurnal Ilmiah Sekolah Dasar*, *1*(4), 247. https://doi.org/10.23887/jisd.v1i4.12593.

Binali, T., Tsai, C.-C., & Chang, H.-Y. (2021). University students' profiles of online learning and their relation to online metacognitive regulation and internet-specific epistemic justification. *Computers & Education*, *175*. https://doi.org/10.1016/j.compedu.2021.104315.

Candra, I., Sulistya, N., & Prasetyo, T. (2018). Pengembangan Instrumen Sikap Sosial Tematik Siswa SD Kelas IV. *Jurnal Ilmiah Sekolah Dasar*, *2*(4), 455. https://doi.org/10.23887/jisd.v2i4.16167.

Dessiane, S. T., & Kristin, F. (2021). Pengembangan Instrumen Penilaian Sikap Sosialpembelajaran Tematik Kelas 4 SD. *Jurnal Pendidikan Ilmu Pengetahuan Sosial Indonesia*, *6*(1), 21–26. https://doi.org/10.26737/jpipsi.v6i1.2310.

Dhika, H., Destiawati, F., Sonny, M., & Surajiyo. (2019). Study of the use and application of the moodle e-learning platform in high school. *Journal of Physics: Conference Series*, *1175*, 012219. https://doi.org/10.1088/1742-6596/1175/1/012219.

Harta, J., Rasuh, N. T., & Seriang, A. (2020). Using HOTS-Based Chemistry National Exam Questions to Map the Analytical Abilities of Senior High School Students. *Journal of Science Learning*, *3*(3), 143–148. https://doi.org/10.17509/jsl.v3i3.22387.

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. Bin, & Mohamad, M. M. B. (2011). The Level of Marzano Higher Order Thinking Skills among Technical Education Students. *International Journal of Social Science and Humanity*, *1*(2), 121–125. https://doi.org/10.7763/ijssh.2011.v1.20.

Imania, K. A., & Bariah, S. K. (2019). Rancangan Pengembangan Instrumen Penilaian Pembelajaran Berbasis Daring. *Jurnal Petik*, *5*(1), 31–47. https://doi.org/10.31980/jpetik.v5i1.445.

Khaerunnisa, E., & Pamungkas, A. S. (2018). Pengembangan Instrumen Kecakapan Matematis Dalam Konteks Kearifan Lokal Budaya Banten Pada Materi Bangun Ruang Sisi Datar. *Kreano: Jurnal Matematika Kreatif-Inovatif*, *9*(1), 17–27. https://doi.org/10.15294/kreano.v9i1.11210.

Kruit, P. M., Oostdam, R. J., van den Berg, E., & Schuitema, J. A. (2018). Assessing students' ability in performing scientific inquiry: instruments for measuring science skills in primary education. *Research in Science and Technological Education*, *36*(4). https://doi.org/10.1080/02635143.2017.1421530.

Kuantum, Cilacap, K., Ramadhan, G., Dwijananti, P., & Wahyuni, S. (2018). Analisis Kemampuan Berpikir Tingkat Tinggi (High Order Thinking Skills) Menggunakan Instrumen Two Tier Multiple Choice Materi Konsep Dan Fenomena Kuantum Siswa Sma Di Kabupaten Cilacap. *UPEJ Unnes Physics*

Education Journal, *7*(3), 85–90. https://doi.org/10.15294/upej.v7i3.27682.
Kuntoro, B. T., & Wardani, N. S. (2020). Pengembangan Instrumen Penilaian Sikap Sosial Pembelajaran Tematik Kelas III SD. *Jurnal Ilmiah Wahana Pendidikan*, *6*(3), 295–307. https://doi.org/10.5281/zenodo.3737983.
Lestari, N., & Harjono, N. (2021). Pengembangan Instrumen Penilaian Karakter PPK Aspek Kemandirian Pada Siswa SD Kelas 4. *MIMBAR PGSD Undiksha*, *9*(1). https://doi.org/10.23887/jjpgsd.v9i1.33379.
Leutner, D., Fleischer, J., Grunkorn, J., & Klieme, E. (2017). *Competence Assessment in Education Research, Models and Instruments*. Springer Nature. https://doi.org/10.1007/978-3-319-50030-0_16.
Maison, Darmaji, Astalini, Agus Kurniawan, D., Sumaryanti, & Perdana, R. (2020). Supporting Assessment in Education: E-Assessment Interest in Physics. *Universal Journal of Educational Research*, *8*(1), 89–97. https://doi.org/10.13189/ujer.2020.080110.
Manikowati, Nf., & Iskandar, D. (2018). Pengembangan Model Mobile Virtual Laboratorium Untuk Pembelajaran Praktikum Siswa Sma. *Jurnal Kwangsan*, *6*(1), 23. https://doi.org/10.31800/jtp.kw.v6n1.p23--42.
Marchand Martella, A., Yatcilla, J. K., Park, H., Marchand-Martella, N. E., & Martella, R. C. (2021). Investigating the active learning research landscape through a bibliometric analysis of an influential meta-analysis on active learning. *SN Social Sciences*, *1*(9), 1–24. https://doi.org/10.1007/s43545-021-00235-1.
Maulida, I., Dibia, I. K., & Astawan, I. G. (2020). The Development of Social Attitude Assessment Instrument and Social Studies Learning Outcomes Grade IV on Theme of Indahnya Keragaman di Negeriku. *Indonesian Journal Of Educational Research and Review*, *3*(1), 12. https://doi.org/10.23887/ijerr.v3i2.25823.
Mohd Zain, N., Mohd Fadil, N. F., & Abdul Hadi, A. (2018). Learning Management System: An Experience and Perception Study from Medical Imaging Lecturers and Scholars in a Private University. *International Journal of Interactive Mobile Technologies (IJIM)*, *12*(7), 174. https://doi.org/10.3991/ijim.v12i7.9638.
Nurmala, D. A., Tripalupi, L. E., & Suharsono, N. (2014). Pengaruh motivasi belajar dan aktivitas belajar terhadap hasil belajar akuntansi siswa kelas X akuntansi SMK Negeri 1 Singaraja Tahun Ajaran 2013/2014. *Jurnal Pendidikan Ekonomi*, *4*(1), 86–95. https://doi.org/10.23887/jjpe.v4i1.3046.
Nurtanto, M., Arifin, Z., Sofyan, H., Warju, W., & Nurhaji, S. (2020). Development of model for professional competency assessment (Pca) in vocational education: Study of the engine tune-up injection system assessment scheme. *Journal of Technical Education and Training*, *12*(2), 34–45. https://doi.org/10.30880/jtet.2020.12.02.004.
Patria, S. F. D., & Heswari, S. (2021). Development of ethnomathematic-based on mathematics e-module to improve students' logical thinking skills. *AIP Conference Proceedings 233*, 040005–1–040005–040007. https://doi.org/10.1063/5.0043250.
Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in Higher Education. *Educational Assessment, Evaluation and Accountability*, *22*(3), 215–225. https://doi.org/10.1007/s11092-010-9098-7.
Pratiwi, P. H., Nur, H., & Martiana, A. (2017). Pengembangan Modul Mata Kuliah Penilaian Pembelajaran Sosiologi Berorientasi Hots. *Jurnal Cakrawala Pendidikan*, *36*(2), 201–209. https://doi.org/10.21831/cp.v36i2.13123.
Purnomo, H, & Wilujeng, I. (2016). Pengembangan Bahan Ajar dan Instrumen Penilaian IPA Tema Indahnya Negeriku Penyempurnaan Buku Guru dan Siswa Kurikulum 2013. *Jurnal Prima Edukasia*, *4*(1), 67–78. https://doi.org/10.21831/jpe.v4i1.7697.
Purnomo, Heru, & Wilujeng, I. (2016). Pengembangan Bahan Ajar dan Instrumen Penilaian IPA Tema Indahnya Negeriku Penyempurnaan Buku Guru dan Siswa Kurikulum 2013. *Jurnal Prima Edukasia*, *4*(1), 67–68. https://doi.org/10.21831/jpe.v4i1.7697.
Purwanita, Y., Riyanto, Y., & Suyanto, T. (2019). The Influence of Multimedia Assisted Inquiry Learning Methods on My Heroes Theme of Critical Thinking Skills and Learning Outcomes of Class IV Students ofElementary School. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(7), p9169. https://doi.org/10.29322/ijsrp.9.07.2019.p9169.
Puspitasari, T. O., Putri, Y. E., & Yohanes, Y. (2019). Sikap Terhadap Konsentrasi Belajar Siswa Pada Mata Pelajaran Fisika di Sekolah Menengah Atas. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*. https://doi.org/10.30599/jipfri.v3i2.537.
Rohita, R., Fitria, N., Bustan, R., & Haryadi, D. (2018). Teacher's Understanding of the Scientific Approach in the 2013 Curriculum for Early Childhood Education. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, *2*(2), 235. https://doi.org/10.31004/obsesi.v2i2.105.
Sagala, P. N., & Andriani, A. (2019). Development of Higher-Order Thinking Skills (HOTS) Questions of

Probability Theory Subject Based on Bloom's Taxonomy. *Journal of Physics: Conference Series*, *1188*(1). https://doi.org/10.1088/1742-6596/1188/1/012025.

Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2019). Taking a future perspective by learning from the past – A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review*, *16*. https://doi.org/10.1016/j.edurev.2016.05.002.

Silberman, D., Carpenter, R., Takemoto, J. K., & Coyne, L. (2021). The impact of team-based learning on the critical thinking skills of pharmacy students. *Currents in Pharmacy Teaching and Learning*, *13*(2), 116–121. https://doi.org/10.1016/j.cptl.2020.09.008.

Simarmata, N. N., Wardani, N. S., & Prasetyo, T. (2019). Pengembangan Instrumen Penilaian Sikap Toleransi Dalam Pembelajaran Tematik Kelas Iv Sd. *Jurnal Basicedu*, *3*(1), 194–199. https://doi.org/10.31004/basicedu.v3i1.122.

Srirahayu, R. R. Y., & Arty, I. S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 168–181. https://doi.org/10.21831/pep.v22i2.20270.

Subagia, I. W., & Wiratma, I. G. L. (2016). Profil Penilaian Hasil Belajar Siswa Berdasarkan Kurikulum 2013. *JPI (Jurnal Pendidikan Indonesia)*, *5*(1), 39–54. https://doi.org/10.23887/jpi-undiksha.v5i1.8293.

Sudijono, A. (2008). *Pengantar Statistik Pendidikan*. raja grafindo persada.

Sudirman, S., Kistiono, K., Akhsan, H., & Ariska, M. (2020). Pengembangan Instrumen Penilaian Pengetahuan, Sikap Dan Keterampilan Ipa Berbasis Berpikir Kritis Pada Konsep Listrik Siswa SMP. *Jurnal Inovasi Dan Pembelajaran Fisika*. https://doi.org/10.36706/jipf.v7i1.10903.

Sukmasari, V. P., & Rosana, D. (2017). Pengembangan penilaian proyek pembelajaran IPA berbasis discovery learning untuk mengukur keterampilan pemecahan masalah. *Jurnal Inovasi Pendidikan IPA*, *3*(1), 101–110. https://doi.org/10.21831/jipi.v3i1.10468.

Suratmi, Laihat, Asnimar, & Ela Okta Handini. (2020). Teachers understanding of HOTS based assessment in elementary schools. *The 2nd International Conference on Elementary Education*, *2*(23), 1157–1164.

Suryaningsih, N. M. A., & Rimpiati, N. L. (2018). Implementation of Game-Based Thematic Science Approach in Developing Early Childhood Cognitive Capabilities. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, *2*(2), 194–201. https://doi.org/10.31004/obsesi.v2i2.90.

Utami, D. A. P., & Wardani, N. S. (2020). Pengembangan Instrumen Penilaian Kognitif dalam Pembelajaran Tematik Kelas 5 SD. *Jurnal Ilmiah Kependidikan*, *20*(2), 1–18. https://doi.org/10.12345/lentera.v12i2.463.

Van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers and Education*, *158*(August), 104000. https://doi.org/10.1016/j.compedu.2020.104000.

Wei, X., Saab, N., & Admiraal, W. (2021). Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review. *Computers & Education*, *163*. https://doi.org/10.1016/j.compedu.2020.104097.

Wicaksono, T. P., Muhardjito, & Harsiati, T. (2016). Pengembangan penilaian sikap dengan teknik observasi , self assessment , dan peer assessment pada pembelajaran tematik kelas V SDN Arjowinangun 02 Malang. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, *1*(1), 45–51. https://doi.org/10.17977/jp.v1i2.5214.

Wulandari, P., Abadi, I. B. G. S., & Ganing, N. N. (2018). Pengaruh Model Pembelajaran Think Pair Share Berbasis Penilaian Portofolio Terhadap Kompetensi Pengetahuan Ipa Siswa Kelas Iv Sd Negeri Gugus Kapten Kompyang Sujana Denpasar Barat Tahun 2017/2018. *Mimbar PGSD*, *6*, 1–9. https://doi.org/10.23887/jjpgsd.v6i3.15772.

Wulandari, P., Abadi, I. B. G., & Suniasih, N. W. (2018). Pengaruh Model Pembelajaran Think Pair Share Berbasis Penilaian Portofolio Terhadap Kompetensi Pengetahuan IPA Siswa Kelas IV SS Negeri Gugus Kapten Kompyang Sujana Denpasar Barat Tahun 2017/2018. *MIMBAR PGSD Undiksha*, *6*(3), 161–168. https://doi.org/10.23887/jjpgsd.v6i3.15772.

Yektyastuti, R., & Ikhsan, J. (2016). Pengembangan media pembelajaran berbasis android pada materi kelarutan untuk meningkatkan performa akademik siswa SMA. *Jurnal Inovasi Pendidikan IPA*, *2*(1), 88. https://doi.org/10.21831/jipi.v2i1.10289.