JPF | Volume 13 | Number 1 | 2025 | 90 - 102

p - ISSN: 2302-8939
e - ISSN: 2527-4015

Jurnal Pendidikan Fisika

https://journal.unismuh.ac.id/index.php/jpf

DOI: 10.26618/jpf.v13i1.15654

An Assessment to Measure Problem-Solving Ability and Science Process Skills: A Feasibility Test

Rohaya Evatri Faila[1)*], Supahar [2)], Irvany Nurita Pebriana[3)]

Department of Physics Education, Universitas Negeri Yogyakarta, Yogyakarta, 55281, Indonesia

**Corresponding author: rohayaevatri.2020@student.uny.ac.id*

Received: July 31, 2024; Accepted: December 26, 2024; Published: January 24, 2025

Abstract – *The ability to solve problems and apply scientific processes is a critical component of 21st-century education, especially within the context of the Indonesian "Merdeka" curriculum, which promotes independent and adaptive learning. However, traditional assessment methods fail to capture these competencies adequately. This study aims to determine the feasibility of an assessment instrument designed to measure high school students' problem-solving ability and science process skills. The instrument was developed using Istiyono's test development model and focused on the test trial stage, which involves expert validation, empirical analysis, reliability testing, and item difficulty analysis. Using Istiyono's model, the development focused on expert validation, empirical testing, and item analysis. A total of 252 students from six high schools in Yogyakarta and surrounding regions participated in the study. The content validity was assessed using Aiken's V, showing all items to be valid. Content validity, evaluated using Aiken's V, indicated all items were valid. Empirical validation using the QUEST program yielded INFIT MNSQ values of 0.99 ± 0.81 for problem-solving and 1.01 ± 0.21 for science process skills, confirming item validity. Reliability coefficients ranged from 0.61 to 0.97, indicating high reliability. The difficulty level of items was distributed across easy, medium, and difficult categories. These findings support the feasibility of the assessment instrument in evaluating students' higher-order thinking skills. The findings demonstrated the instrument's suitability for assessing complex cognitive and inquiry-based skills. The validated instrument offers practical implications for educators to improve learning through formative assessment aligned with curriculum goals.*

Keywords: *assessment; feasibility test; problem-solving ability; science process skills*

© 2025 Physics Education Department, Universitas Muhammadiyah Makassar, Indonesia.

I. INTRODUCTION

The skills required for the 21st century, known as the 6Cs, have played a crucial role in the development of an independent learning curriculum in Indonesia, which includes critical thinking and problem-solving, collaboration, communication, creativity, citizenship/culture, civil society, and character building (Rismorlita et al., 2021; Masoleh et al., 2023). Measuring problem-solving ability and science process skills is aligned with these goals by assessing students' ability to think critically and apply the scientific method. Developing valid and reliable instruments ensured these skills were effectively evaluated, supporting the curriculum's emphasis on real-world applications and student-centered learning.

The *Merdeka* Curriculum was introduced as a solution to respond to comprehensive transformation in all aspects, especially in facing the demands of human resources (Indarta et al., 2022; Simarmata & Mayuni, 2023; Langoday et al., 2024). The *Merdeka* Curriculum aims to improve independent thinking skills, enthusiasm for learning, self-confidence, and optimism, as well as provide broad freedom of thought. In addition, it encourages students to actively learn, care about the environment, and improve skills and adaptation in facing future challenges in 21st-century skills (Sartini & Mulyono, 2022; González-Salamanca et al., 2020). The *Merdeka* Curriculum emphasizes competency-based learning and the development of learners' potential with an inclusive and creative approach.

Based on observations at SMA Negeri 4 Yogyakarta, Physics learning has still applied conventional methods, where teachers predominantly deliver material unidirectionally and rely solely on summative assessments. Students were classroom learning activities in class, which hindered teachers' ability to identify students' learning needs, obstacles, or difficulties faced and monitor their learning development. This situation particularly affects students who struggle with understanding or are hesitant to ask questions due to the absence of immediate feedback in summative assessment (Gamage et al., 2019). In the learning process, the main focus was on the cognitive domain, leading to traditional assessments through written exams that only evaluated end-of-topic knowledge (Sari et al., 2019; Malhotra et al., 2023). Teachers had not integrated formative assessment due to the large class sizes, which made it challenging.

The application of assessment in learning involves several strategies implemented by educators to achieve instructional goals and promote meaningful learning. According to its function, an assessment can be categorized as an assessment of learning (assessment as a learning process), an assessment for learning (assessment of the learning process), and an assessment of learning (assessment at the end of the learning process) (Sufyadi et al., 2021). The current trend in the implementation of assessment has often prioritized summative assessment (assessment of learning) primarily for reporting learning outcomes. However, the results were not optimally used as feedback to enhance learning.

The implementation of assessment should prioritize formative assessment over summative assessment and use the results as a basis for refining instruction (Sufyadi et al., 2021). By implementing formative assessment (assessment as learning and assessment for learning) more widely, educators could enhance the learning process and gain a more precise understanding of learners' needs and progress. Assessment offers valuable data for designing effective instruction and monitoring learning outcomes (Anggraena et al., 2022; Zeng et al., 2018).

The continued use of conventional methods, where students are passive participants, has limited teachers' ability to identify students' needs and learning progress. Pedagogy supports an inquiry-based approach that emphasizes the active involvement of learners in learning (Muslim, 2016). Through this approach, learners engage in scientific processes, including observing, hypothesizing, experimenting, and applying critical and creative thinking. Students' problem-solving ability and science process skills are essential components to be mastered in science education (Tanti et al., 2020). Both of these skills involve the process of scientific inquiry and problem solving related to science concepts relevant to real-world contexts.

Problem-solving is a critical skill students need to be able to face the challenges of the 21st century. While numerous studies have aimed to improve students' problem-solving skills, most have focused solely on the application of instructional models. Few studies have examined students' skill profiles and response quality (Fadilah et al., 2024). One of the ways to assess problem-solving in physics is through procedural task analysis. Procedural task analysis involves classifying and organizing tasks to identify relationships and reach valid conclusions (Kadir et al., 2020). The study of global warming issues involves both understanding the phenomenon and analyzing its causes and consequences (Alika et al., 2018). Thus, learners are encouraged to develop skills to solve complex, real-world problems such as global warming.

In addition to problem-solving, science process skills are equally essential. Science process skills refer to the abilities students demonstrate in scientific activities that mirror scientists' approaches (Supahar et al., 2017). They encompass cognitive and psychomotor skills used to generate knowledge, solve problems, and draw conclusions (Mabsutsah & Yushardi, 2022). Science process skills are cognitive and psychomotor skills that involve problem-solving. Many students tend to memorize formulas without grasping their practical relevance, reducing physics to mere equation application. Therefore, science process skills must be cultivated to enable students to conduct inquiries, collect and interpret data, and communicate results effectively in physics education.

Research by Hardiansyah et al. (2022) revealed that existing instructional practices have not effectively promoted students' science process skills. Assessments used by teachers were often inadequate for helping students address contextual problems, underscoring the need for better assessment instruments. Therefore, to address these issues, a solution was proposed, which was to develop a 4D-based assessment instrument involving expert validation and field testing. However, this study had limitations, including a small sample size of 30 students, 33% of test items being invalid, and an overrepresentation of easy-level questions. Therefore, it is necessary to conduct a feasibility study of an assessment on the theme of global warming to measure high school students' problem-solving ability and science process skills.

II. METHODS

This research employed the test instrument development model proposed by Istiyono (2020). The model consists of three stages: test design, test testing, and measurement. However, researchers focused only on the test testing stage. In the test planning stage, nine steps were followed, namely: (1) determination of test objectives; (2) determination of competencies and materials to be tested; (3) preparation of item distribution matrix; (4) preparation of grids; (5) writing and assembling test items; (6) preparation of scoring rubrics; (7) validation of test items; (8) revision for item improvement; and (9) assembly of test instruments. In the test testing stage, four steps were conducted, namely: (1) determination of test subjects; (2) implementation of the test; (3) scoring and data analysis of test items; and (4) revision of test items that do not meet the criteria.

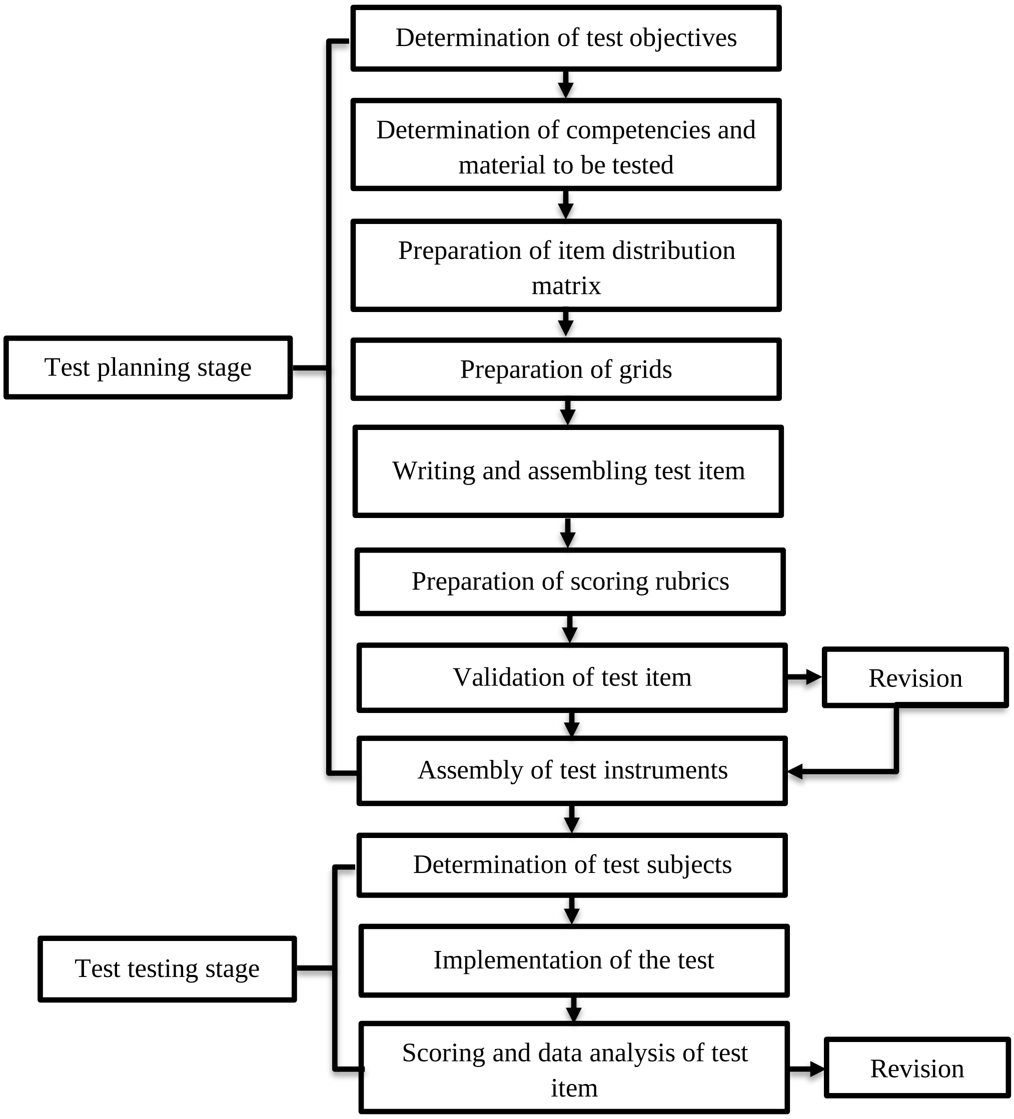

Figure 1. Flowchart of the data collection procedure

The sample selection used a cluster random sampling technique. The selection of test subjects was determined by school rankings based on Computer-Based Written Test scores in 2022. The selected high schools represented the top, middle, and bottom-ranked schools. Since rank-based clusters reflect the overall diversity of schools in terms of demographics, resources and student backgrounds, random selection of clusters helped prevent systematic bias. In addition, the sampling within the selected clusters was randomized and ensured sufficient variation to approximate the characteristics of the population.

The trial subjects consisted of students from Class X Phase E and Class IX Phase F of the Independent Curriculum who had studied the topic of global warming. The test involved 252 students from Phase E and Phase F of the Independent Curriculum at SMA Negeri 4 Yogyakarta, SMA Negeri 7 Yogyakarta, SMA Negeri 1 Ngaglik, SMA Negeri 1 Pundong, SMA Negeri 2 Klaten, and SMA Negeri 1 Wonosari. The research instruments aimed to assess high school students' problem-solving abilities and science process skills. The pretest and posttest employed the same set of questions for consistency. The questions given consisted of 5 items of problem-solving ability descriptions with categories C1, C3, C4, C5, and C6 and 21 multiple choice items of science process skills with categories C1, C2, C3, C4, and C5. Each item was validated by two physics education experts and two practitioners. The feasibility of the assessment was evaluated through expert judgment (content validity), empirical validity, reliability testing, and item analysis.

Content validity refers to the extent to which an instrument's content reflects the construct it intends to measure. In the context of research, the Aiken index was employed to measure the content validity of an instrument. Based on the assessment results from a panel of experts, the Aiken index calculates quantified item relevance to the measured construct. The formula (1) for calculating Aiken's V validity coefficient is:

$$V = \frac{\sum s}{n(c-1)} \qquad (1)$$

Remarks:
$s = r - l_0$
$V =$ index of expert agreement on validation
$r =$ number given by an expert
$l_0 =$ lowest validity score
$n =$ number of experts
$c =$ highest validity score

Based on the calculated Aiken’s V index, each item was classified according to its validity level. The V Aiken index categories can be seen in Table 1 below.

Table 1. V Aiken index category

V Aiken index	Category
< 0.4	Less
0.4 – 0.8	Medium
> 0.8	High

The empirical validity of both problem-solving and science process skill items in the pretest and posttest was evaluated using the QUEST program during a limited trial. Item validity was analyzed using the Infit Mean Square (INFIT MNSQ) values generated in QUEST. The item validity criteria based on INFIT MNSQ values are listed in Table 2.

Table 2. INFIT MNSQ value criteria (Shepardson & Adams, 1996)

INFIT MNSQ	Criteria
> 1.33	Not Valid
0.77 – 1.33	Valid
< 0.77	Not Valid

Reliability was examined through the internal consistency of the instrument. The analysis was carried out using the QUEST software. The reliability coefficient was obtained from the "Summary of Item Estimate" section in QUEST. Higher reliability values indicate a greater consistency of the items with the measurement model. The classification of reliability coefficients is presented in Table 3.

Table 3. Reliability coefficient category

Reliability coefficient	Category
0.80 – 1.00	Very high
0.60 – 0.79	High
0.40 – 0.59	Medium
0.20 – 0.39	Low
0.00 – 0.19	Very low

Item analysis was conducted to determine the difficulty level of each question. The difficulty index reflects whether an item is too easy or too difficult, based on students' ability to answer correctly. It was calculated by the proportion of students answering each item correctly. Ideally, high-quality items fall within the moderate difficulty range. In this study, item difficulty was assessed through the Threshold values obtained from QUEST output. Table 4 presents the classification of item difficulty levels.

Table 4. Level of difficulty category (Setyawarno, 2017)

Threshold	Category
$b > 2$	Very Difficult
$1 < b \leq 2$	Difficult
$-1 \leq b \leq 1$	Medium
$-2 \leq b < -1$	Easy
$b < -2$	Very Easy

III. RESULTS AND DISCUSSION

1. Validity

An assessment is considered feasible as a tool to measure students' abilities if it meets the criteria of validity reliability, and demonstrates an acceptable or good level of item difficulty. According to Sari et al. (2017), instruments that have high validity ensure that the assessment is relevant according to the predetermined objectives, and reliable instruments provide consistent and reliable results every time they are used. The results of content validation using V Aiken show that all items on the problem-solving ability and science process skills were valid.

Table 5. Empirical validation categories of problem-solving ability and science process skill

	Average value	**Category**
Problem-solving ability	0.99 ± 0.81	High
Science process skill	1.01 ± 0.21	High

The results of empirical validation that have been carried out for problem-solving ability questions state that all items were valid. This is shown in the analysis results using the Quest program with the average INFIT MNSQ value of 0.99 ± 0.81. Meanwhile, for the question of science process skills, there is one item that is invalid and needs to be eliminated, namely item number 20. Thus, there are only 21 items that met the validity criteria. However, the 21 items are declared valid, as shown in the analysis results using the Quest program with the average INFIT MNSQ value at a value of 1.01 ± 0.21.

2. Reliability

The problem-solving ability and science process skills instruments were found to be reliable. The reliability category of the problem-solving ability test was classified as high.

Table 6. Reliability categories of problem-solving ability

Reliability	**Value**	**Category**
Summary of item estimates	0.75	High
Summary of case estimate	0.61	High

The reliability of the item estimates for the problem-solving ability test was 0.75, indicating high reliability. A higher reliability value suggests that more items fit the tested model. Meanwhile, the case estimate reliability was 0.61, also categorized as high, indicating stable and consistent results upon retesting.

Table 7. Reliability categories of science process skill

Reliability	**Value**	**Category**
Summary of item estimates	0.97	High
Summary of case estimate	0.72	High

Similarly, the science process skills test also demonstrated high reliability. The item estimate reliability was 0.97, indicating very high internal consistency, which demonstrates strong alignment between the items and the measurement model. The case estimate reliability was 0.72, classified as high, which implies a strong potential for consistent results. Similar research was conducted by Abidin et al. (2019) to develop a computerized adaptive test for physics critical thinking skills (CAT-PhysCriTS) that met established feasibility standards. The findings indicate that higher item reliability enhances confidence in the alignment between the sample and the tested items, while higher personal reliability strengthens confidence in the measurement's ability to yield consistent results. In addition, research conducted by Istiyono et al. (2020) obtained high reliability on the instrument developed. This shows that the instrument is very good in terms of accuracy and tolerance to failure. Therefore, higher reliability reflects better measurement accuracy and instrument suitability. Reliable instruments ensure that measurements yield consistent results across different administrations.

3. Level of Difficulty

The difficulty level of the problem-solving items ranged from moderate to difficult, while science process skills items ranged from easy to difficult. The distribution of item difficulty levels is presented in Figure 2 and Figure 3.

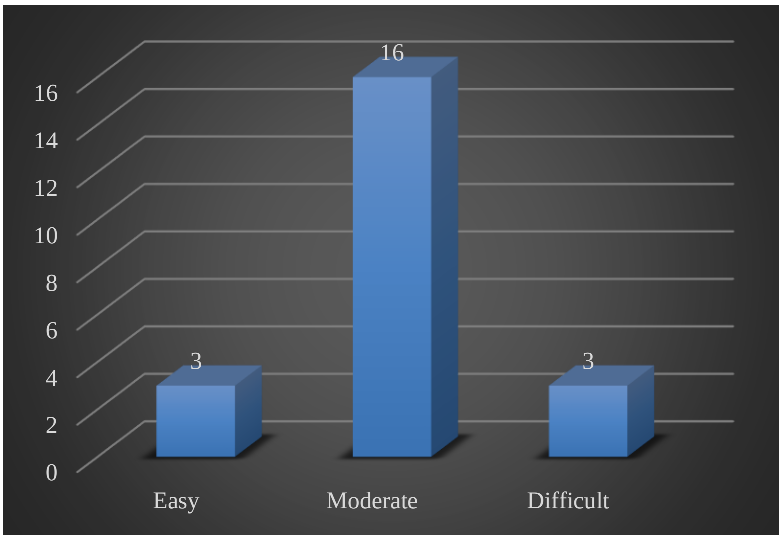

Figure 2. Distribution chart of the level of difficulty of science process skill items

Based on Figure 2, the majority of items (80%) fell into the medium difficulty category, ensuring that the test provided a balanced challenge that effectively differentiated between participants' skills. Easy items, comprising 15% of the test, were included to enable all test takers, including those of basic skill, to answer some questions correctly. The difficult questions, which also comprise 15%, were intended to identify high-skill individuals and evaluate advanced science process skills.

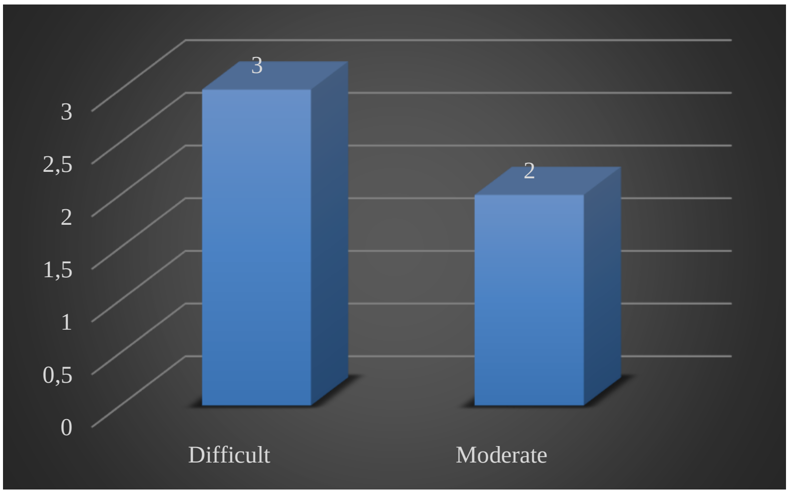

Figure 3. Distribution chart of the level of difficulty of problem-solving ability items

This distribution indicates that the test is designed for learners with a strong foundation in problem-solving, encouraging them to demonstrate more profound understanding and application of skills in complex scenarios. The 40% of items categorized as moderate served as a benchmark to assess general problem-solving ability, while the larger proportion (60%) of difficult items emphasized higher-order thinking, aiming to distinguish top-performing students.

Following the validation, reliability, and difficulty analyses, the qualified test items were compiled into a final assessment instrument. The test items developed in this study consisted of 5 items to measure problem-solving ability in the form of a description test. The selection of this test form is in accordance with Arikunto opinion in Wibowo & Faizah (2021) that the description test has significant advantages because it allows students to show their way of thinking in depth in solving problems. The description test provides an opportunity for learners to explain the steps of problem-solving in detail, showing analysis, synthesis, and evaluation skills that cannot be fully expressed through multiple-choice tests. Therefore, the descriptive test is very effective and appropriate to measure learners' problem-solving ability.

Additionally, the study produced 21 multiple-choice items to measure science process skills, each with five answer options. The multiple-choice format was chosen based on its practicality and suitability for measuring multiple indicators, a broad range of materials, and a large sample size (Ramli et al., 2018). Multiple-choice tests are ideal for large-scale testing due to efficient scoring and comprehensive content coverage. The resulting test items can accurately assess both problem-solving ability and science process skills. The development of these test items was based on the learning outcomes and learning objectives flow, with clearly defined indicators aligned to measure the targeted abilities.

An assessment can be considered suitable for the measurement of students' abilities if it meets the criteria of validity, reliability and at least has a sufficient or good level of item difficulty.

According to Sari et al. (2017) and Agnezi & Festiyed (2023), an instrument that has high validity ensures that the assessment is relevant according to the predetermined objectives and a reliable instrument provides consistent and reliable results every time it is used. In physics, a valid assessment ensures that test items are fit for purpose. Physics tests should evaluate not only students' factual knowledge but also their ability to apply principles, analyze problems, and synthesize solutions in real-world contexts. Reliability in physics assessment allows teachers to accurately track student progress and make informed instructional decisions. Furthermore, an appropriate level of item difficulty ensures that the test challenges students at different ability levels. By following these criteria, physics educators can design assessments that effectively measure student understanding, guide instructional strategies, and foster meaningful learning experiences.

IV. CONCLUSION AND SUGGESTION

This study developed five items to measure students' problem-solving skills and 21 items to assess their science process skills. The instrument was found to be feasible for assessing high school students' problem-solving and science process skills, as it fulfilled the requirements for content validity, empirical validity, reliability, and appropriate item difficulty level.

Based on the findings of this study, future research is recommended to involve a broader and more diverse sample of students to enhance the generalizability of the instrument. Further development may include integrating the instrument into digital assessment platforms to support efficient formative assessment. Longitudinal studies are also needed to evaluate the long-term impact on students' higher-order thinking skills. Additionally, exploring teachers' implementation practices and identifying challenges in classroom use would provide valuable insights. Finally, incorporating construct validity analysis and adapting the instrument for other science subjects can contribute to more comprehensive and versatile assessment development.

ACKNOWLEDGMENTS

Thanks to physics learning lecturers of Universitas Negeri Yogyakarta, physics teachers of SMAN 4 Yogyakarta, and physics teachers of SMAN 1 Banguntapan who have helped the validation process of experts and practitioners. I would also like to thank Dr. Drs. Supahar, M.Si. as the main supervisor and Mrs. Irvany Nurita Pebriana, M.Pd. as the co-supervisor who have helped and guided me in completing this research. My deepest gratitude also goes to all those who have helped in the completion of this article.

REFERENCES

Abidin, A. Z., Istiyono, E., Fadilah, N., & Dwandaru, W. S. B. (2019). A computerized adaptive test for measuring the physics critical thinking skills. *International Journal of Evaluation and Research in Education*, *8*(3), 376–383. https://doi.org/10.11591/ijere.v8i3.19642

Agnezi, L. A., & Festiyed. (2023). Development of the online assessment instrument for fluid, temperature, and heat to measure the problem solving skills of high school students. *Journal of Physics: Conference Series, 2582,* 1-10. Doi. 10.1088/1742-6596/2582/1/012053

Alika, M. F., Darsono, T., Linuwih, S. (2018). Pengembangan soal model pisa untuk mengukur kemampuan bepikir tingkat tinggi siswa smp pada materi pemanasan global. *Upej: Unnes Physics Education Journal*, *7*(3), 58-65. https://doi.org/10.15294/upej.v7i3.27678

Anggraena, Y., Ginanto, D., Felicia, N., Andiarti, A., Herutami, I., Alhapip, L., Iswoyo, S., Hartini, Y., Mahardika, R. L. (2022). *Panduan pembelajaran dan asesmen*. Badan Standar, Kurikulum, Dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia.

Fadilah, J., Ainirohmah, N., Marwaziah, N., Hafizah, N., Hayu, S. R., Andrian, D., Nofriyandi, N., & Indriati, M. (2024). Students' creative thinking ability profile in mathematics lessons of Riau Province. *Mathematics Research and Education Journal*, *8*(1), 29-41. https://doi.org/10.25299/mrej.2024.vol8(1).16823

Gamage, S. H. P. W., Ayres, J. R., Behrend, M. B., & Smith, E. J. (2019). Optimising moodle quizzes for online assessments. *International Journal of STEM Education*, *6*(1), 1-14. https://doi.org/10.1186/s40594-019-0181-4

González-Salamanca, J. C., Agudelo, O. L., & Salinas, J. (2020). Key competences, education for sustainable development and strategies for the development of 21st century skills. A systematic literature review. *Sustainability*, 12(24), 1-17. https://doi.org/10.3390/su122410366

Hardiansyah, F., Misbahudholam, A. R. M., & Hidayatillah, Y. (2022). IPAS learning assessment to measure science process skill in elementary school. *International Journal of Elementary Education*, *6*(4), 612–623. https://doi.org/10.23887/ijee.v6i4.54217

Indarta, Y., Jalinus, N., Waskito, W., Samala, A. D., Riyanda, A. R., & Adi, N. H. (2022). Relevansi kurikulum merdeka belajar dengan model pembelajaran abad 21 dalam perkembangan era society 5.0. *Edukatif: Jurnal Ilmu Pendidikan*, *4*(2), 3011–3024. https://doi.org/10.31004/edukatif.v4i2.2589

Istiyono, E., Dwandaru, W. S. B., Permatasari, A. K., & Aristiawan. (2020). Developing computer based test to assess students' problem-solving in physics learning. *Journal of Physics: Conference Series*, *1440*, 1-8. https://doi.org/10.1088/1742-6596/1440/1/012060

Istiyono, E. (2020). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan teori tes klasik dan modern*. UNY Press.

Kadir, H. D., Arsyad, M., & Marisda, D. H. (2020). Implementation of problem solving methods in elasticity course implementasi metode pemecahan masalah pada materi elastisitas. *Jurnal Pendidikan Fisika*, *8*(3), 279–285. https://doi.org/10.26618/jpf.v8i3.3803

Langoday, Y. R., Nurrahma, N., & Rijal, Syamsul. (2024). Policy reflection: Kurikulum merdeka as educational innovation in the era of society 5.0. *Edunesia: Jurnal Ilmiah Pendidikan*, *5*(2), 957–978. https://doi.org/10.51276/edu.v5i2.915

Mabsutsah, N., & Yushardi, Y. (2022). Analisis kebutuhan guru terhadap e module berbasis steam dan kurikulum merdeka pada materi pemanasan global. *Jurnal Pendidikan Mipa*, *12*(2), 205–213. https://doi.org/10.37630/jpm.v12i2.588

Malhotra, R., Massoudi, M., Jindal, R. (2023). Shifting from traditional engineering education towards competency-based approach: The most recommended approach-review. *Education and Information Technologies, 28*, 9081-9111. https://doi.org/10.1007/s10639-022-11568-6

Masoleh, S. M. T., Poushaneh, K., Babadi, A. K., & Zaheh, H. R. R. (2023). Citizenship education curriculum for students with special needs. *International Journal of Learning Spaces Studies (IJLSS), 1*(2), 44-54. https://doi.org/10.22034/lss.2023.175772

Muslim, S. S. (2016). *Penggunaan asesmen autentik dalam discovery learning untuk meningkatkan penguasaan konsep dan keterampilan proses sains siswa pada materi pemanasan global*. Thesis. Universitas Pendidikan Indonesia.

Rismorlita, C. E., Philiyanti, F., Prasetio, V. M., & Sari, L. P. (2021). Relevansi kebutuhan stakeholder terhadap pengembangan kurikulum berbasis keterampilan abad 21. *Kagami: Jurnal Pendidikan Dan Bahasa Jepang*, *12*(2), 12-20.

Sari, D. P., Darhim., & Rosjanuardi, R. (2017). Pengembangan instrumen penilaian untuk mengukur kemampuan representasi matematis siswa SMP. *Indonesian Digital Journal of Mathematics and Education*, *4*(7), 421–429.

Sari, I. P., Mustikasari, V. R., & Pratiwi, N. (2019). Pengintegrasian penilaian formatif dalam pembelajaran IPA berbasis saintifik terhadap pemahaman konsep peserta didik. *Jipva (Jurnal Pendidikan IPA Veteran), 3*(1), 52-62. https://doi.org/10.31331/jipva.v3i1.778

Sartini., & Mulyono, R. (2022). Analisis implementasi kurikulum merdeka belajar untuk mempersiapkan pembelajaran abad 21. *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, *8*(2), 1348–1363. https://doi.org/10.36989/didaktik.v8i2.392

Setyawarno, D. (2017). Penggunaan aplikasi software iteman (item and test analysis) untuk analisis butir soal pilihan ganda berdasarkan teori tes klasik. *Jurnal Ilmu Fisika dan Pembelajarannya (JIFP)*, *1*(1), 11–21. https://doi.org/10.19109/jifp.v1i1.866

Shepardson, D. P., & Adams, P. E. (1996). Coming to know and understand alternative assessment in science. *Journal of Science Teacher Education*, *7*(4), 267–282. Retrieved from http://www.jstor.org/stable/43156075

Simarmata, H. A., & Mayuni, I. (2023). Curriculum reform in Indonesia: From competency-based to freedom of learning. *International Journal Of Pedagogical Novelty, 2*(2), 01-13.

Sufyadi, S., Lambas., Rosdiana, T., Rochim, F. A. N., Novrika, S., Isyowo, S., Hartini, Y., Primadona, M., Mahardhika, R. L. (2021). P*anduan pembelajaran dan asesmen jenjang pendidikan dasar dan menengah (SD/MI, SMP/MTs, SMA/SMK/MA)*. Kepala Pusat Asesmen Dan Pembelajaran Badan Penelitian Dan Pengembangan Dan Perbukuan Kementerian Pendidikan, Kebudayaan, Riset, Dan Teknologi.

Supahar., Rosana, D., Ramadani, M., & Dewi, D. K. (2017). The instrument for assessing the performance of science process skills based on nature of science (Nos). *Cakrawala*

Pendidikan, *XXXVII*(3), 435–445. https://doi.org/10.21831/cp.v36i3.14731

Ramli, W., Arafah, K., & Sari, S. S. (2018). Test instrument development for science process skills in mechanic wave topic for grade XI SMA student. *Proceeding Book of 1st International Conference on Educational Assessment and Policy*, *1*, 37–43.

Tanti., Kurniawan, D. A., Kuswanto., Utami, W., & Wardhana, I. (2020). Science process skills and critical thinking in science: Urban and rural disparity. *Jurnal Pendidikan IPA Indonesia*, *9*(4), 489–498. https://doi.org/10.15294/jpii.v9i4.24139

Wibowo, T. E., & Faizah, S. (2021). Pengembangan soal tes untuk mengukur kemampuan pemecahan masalah siswa pada materi bentuk aljabar. *Alifmatika: Jurnal Pendidikan dan Pembelajaran Matematika*, *3*(2), 145–158. https://doi.org/10.35316/alifmatika.2021.v3i2.145-158

Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning a critical review of literature. *Educational Assessment, Evaluation and Accountability, 30*, 211-250. https://doi.org/10.1007/s11092-018-9281-9