**Jurnal Materi dan Pembelajaran Fisika**
Volume 15, Number 2, pp. 56-65 2025
Website : https://jurnal.uns.ac.id/jmpf

# Performance Assessment in STEM Projects: Publication Trends

**Annisa Salsabila[1*], Nurul Fitriyah Sulaeman[2], Zulkarnaen[3], Abdul Hakim[4]**
[1,2,3,4] *Physics Education Study Program, Mulawarman University, Samarinda, 75124*
*Corresponding Author Email : salsabilaaannisa99@gmail.com

**Article's Info**
Received: 21st, May, 2025
Accepted: 9th, October, 2025
Published: 30th, November, 2025

**DOI**:
https://doi.org/10.20961/jmpf.v15i2.102633

**How to Cite :** Salsabila, A., Sulaeman, N. F., & Zulkarnaen, Z., Hakim, A. (2025). Performance Assessment in STEM Projects: Publication Trends, *Jurnal Materi dan Pembelajaran Fisika.* 15(2), 56-65

***Abstract***. In line with the global spread of Science-Technology-Engineering-Mathematics (STEM) education, increasing attention has been given to appropriate forms of assessment. Among various approaches, performance assessment is argued to be particularly suitable for evaluating STEM learning processes. This study aims to analyze publication trends and assessment approaches in performance assessment within STEM projects between 2015 and 2025. Using a Systematic Literature Review (SLR) method, 50 peer-reviewed articles retrieved from Scopus and Google Scholar were examined through thematic analysis, based on defined inclusion and exclusion criteria. The review identified key themes, methodologies, geographic distributions, and gaps in the literature. The findings reveal a significant increase in publications in 2021, reflecting growing academic interest in authentic evaluations of 21st-century skills. Authentic assessment approaches such as performance rubrics, digital portfolios, and direct observation dominate the field. Geographically, South Asia shows the highest number of publications on this topic. However, explicit theoretical integration, validation and reliability of instruments, and teacher involvement in assessment design remain limited. These findings highlight the need to strengthen theoretical foundations, participatory approaches, and methodological rigor in developing contextual and sustainable performance assessments for STEM education. By addressing these gaps, this study contributes to guiding educators, policymakers, and researchers in designing assessment practices that effectively align with the planning, implementation, and evaluation stages of STEM learning..

**Keywords:** Performance assessment, STEM, Project-based learning, Publication trends, Authentic assessment



## INTRODUCTION

STEM (Science, Technology, Engineering, and Mathematics) education has emerged as a strategic approach to addressing the challenges of the Fourth Industrial Revolution, as it fosters essential 21st-century skills such as critical thinking, creativity, collaboration, and contextual problem-solving (Hartini et al., 2020). In the context of physics education in Indonesia, STEM-based teaching materials have been shown to improve students' climate literacy by integrating scientific, technological, engineering, and mathematical perspectives in addressing global issues such as climate change (Hakim et al., 2023) . This example illustrates how STEM education not only develops cognitive skills but also equips learners to respond to pressing societal and environmental challenges. Moreover, student engagement in STEM learning can be enhanced through the Engineering Design Process (EDP), which emphasizes design-oriented, hands-on, and collaborative tasks (Sulaeman et al.,

2021) . However, the successful implementation of STEM education largely depends on teacher readiness, which includes alignment with STEM values, pedagogical competence, and active classroom engagement (Sulaeman et al., 2022).

At the same time, assessing the outcomes of STEM learning requires approaches that go beyond traditional testing, as conventional examinations often fail to capture the complex, process-oriented, and collaborative nature of STEM activities. In this regard, performance-based assessment emerges as a particularly appropriate alternative to holistically evaluate both learning processes and outcomes. However, little is known about how performance assessment has been conceptualized and applied in STEM project contexts, particularly across diverse educational systems and regions. Addressing this gap is crucial to advancing both the theory and practice of STEM education.

Performance assessment is one of the most relevant strategies in STEM education because it can measure complex skills such as decision-making, critical reasoning, and problem-solving. This form of assessment not only evaluates outcomes but also emphasizes the thought processes, participation, creativity, and reflection that emerge during learning (Windyariani & Setiono, 2024). In this context, tools such as rubrics, portfolios, and direct observation play a critical role in ensuring authenticity and fairness.

Evidence from various studies highlights the benefits of performance assessment. In Indonesia, project-based assessment tools developed in elementary science education have proven effective in measuring students' science process skills (Pryantini et al., 2016). Similarly, performance assessments have been found more capable of identifying students' scientific competencies compared to conventional methods (Septiani & Rustaman, 2017). In secondary education, validated performance assessment instruments have demonstrated reliability and effectiveness in evaluating students' critical thinking skills in physics (F. S. Putri & Istiyono, 2017). Additionally, research indicates that such assessments enhance students' conceptual understanding and scientific reasoning (Miyati, 2019).

International perspectives further reinforce these findings. The study by Estapa & Tank (2017) highlights how teacher training in engineering challenge-based assessments can strengthen STEM learning at the basic level. More recently, Azizah et al. (2022) and Sulaeman et al. (2021) demonstrated that Engineering Design Process (EDP)-based assessments are effective in capturing a holistic range of student skills, including collaboration, innovation, and scientific accuracy.

In addition to these benefits, challenges remain in practice. Teachers often struggle to design rubrics that proportionally integrate cognitive, affective, and psychomotor dimensions. This difficulty stems from balancing diverse criteria, ensuring fairness, and maintaining inter-rater reliability (Wahyuni et al., 2018; Wulan, 2018) . At the same time, technological advances have opened opportunities for integrating digital assessment systems, which can improve efficiency, transparency, and accountability (Mellyzar et al., 2024; Retnawati et al., 2024)

Nevertheless, despite a growing body of studies demonstrating the effectiveness of performance assessment, there is still limited comprehensive mapping of global publication trends, methodological approaches, and thematic focuses in this area. Addressing this gap is essential for advancing both theory and practice. Therefore, this study aims to systematically map publication trends in performance assessment within STEM projects from 2015 to March 2025, analyzing assessment methods, geographical contexts, and dominant research priorities.

## METHOD

With the growing number of studies focusing on performance assessment in STEM education, this research employed a Systematic Literature Review (SLR) approach to systematically map publication trends, assessment strategies, and methodological patterns. Although the initial draft referred to meta-analysis, no statistical synthesis such as effect size calculation was conducted; therefore, this study is best categorized as an SLR. In conducting the SLR, a set of predefined inclusion and exclusion criteria was applied to ensure that only studies meeting the conceptual and methodological relevance to the topic were retained. The details of these criteria are presented in Table 1.

**Table 1.** Inclusion and Exclusion Criteria for Article Selection in the Systematic Literature Review

| Criteria Type | Description |
|---|---|
| Inclusion | - Peer-reviewed journal articles indexed in Scopus or Google Scholar.<br>- Published between 2015 and March 2025 in English or Bahasa Indonesia.<br>- Focused on performance or authentic assessment in STEM learning contexts.<br>- Empirical studies employing quantitative, qualitative, or mixed-method designs.<br>- Contained complete metadata (title, abstract, keywords, and references) |
| Exlusion | - Conference proceedings, book chapters, or dissertations.<br>- Studies not directly related to assessment in STEM education.<br>- Duplicated records or incomplete data.<br>- Full-text articles not accessible. |

The review process followed an adaptation of the four-stage model proposed by Putri et al. (2022) hich was streamlined into three stages to fit the scope of this study: (1) Identification and Screening, (2) Eligibility, and (3) Inclusion. The decision to merge the feasibility stage with eligibility was based on overlapping criteria, allowing for a clearer and more efficient review process.

Relevant articles were retrieved from Scopus and Google Scholar using the Publish or Perish (PoP) application. Scopus was chosen for its rigorous indexing and credibility in peer-reviewed research, while Google Scholar was included to capture additional relevant studies not indexed in Scopus, thereby enhancing comprehensiveness. The search used the keywords *"performance assessment"*, *"authentic assessment"*, and *"STEM project"*, limited to publications between 2015 to March 2025, and restricted to the education field.

**Table 2**. Stages of Investigation Papers

| Selection Stage | Selection Criteria | Number of Articles Obtained |
|---|---|---|
| Identification and Screening | Keywords: "performance assessment", "authentic assessment", "STEM project"; timeframe: 2015–2025; databases: Scopus & Google Scholar | 1.206 (Initial Total) |
| Eligibility | Peer-reviewed, complete metadata, topic relevance Indexed in Scopus, Google Scholar | 99 |
| Inclusion | Empirical studies, accessible full-text, quality-checked | 50 (final) |

After the filtering and duplication process, a total of 50 final articles were obtained, consisting of 31 articles from Scopus and 19 articles from Google Scholar. Each article was analyzed across four parameters: 1) Assessment approaches and instruments used, 2) Level of education and number of participants, 3) Geographical distribution of research, and 4) Data analysis methods.

Quantitative data were presented using publication trend graphs and frequency distribution tables. Qualitative insights were identified through thematic analysis, with themes generated inductively from the reviewed studies. The coding process was conducted by two independent reviewers, and any discrepancies were discussed and resolved to ensure inter-rater reliability and validity of interpretation.

## RESULT AND DISCUSSION

### 1. This Publication Trends By Year

As shown in Figure 1 and Table 2, the number of publications on performance assessments in STEM projects has shown a steady upward trend since 2015, reaching a peak in 2021. This trajectory illustrates both the dynamic interest and the fluctuating research priorities surrounding STEM performance assessments.

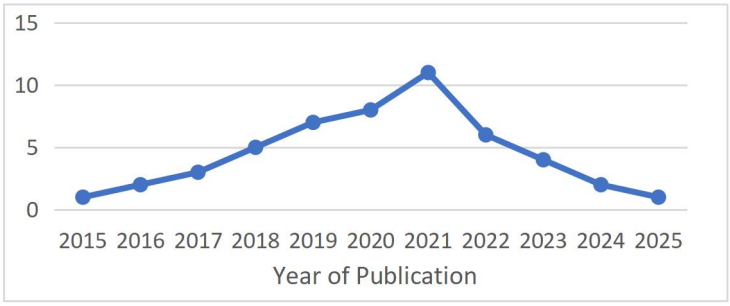


**Figure 1**. Number of publications in the previous 10 years

The peak in 2021 may be explained by the increasing recognition of the need for alternative assessment strategies capable of capturing critical thinking, collaboration, and problem-solving skills, which coincided with the growing adoption of project-based learning approaches (Diana et al., 2021). This trend aligns with Dolin et al. (2018), who emphasized the importance of authentic assessments for fostering 21st-century skills in contextual learning.

Interestingly, the data also reveal a noticeable decline in publications during 2024–2025. This decline may be associated with shifts in research focus, delays in publication processes, or broader educational priorities in the post-pandemic context. While this pattern contrasts with the global momentum of STEM education that increasingly highlights authentic assessment (Dolin et al., 2018), it also underscores the novelty of our findings, pointing to potential gaps between educational policy emphasis and current research outputs. This prompted a growing number of researchers to investigate performance assessments in STEM projects, thereby reinforcing the upward trend in publications observed during the past decade.

### 2. Assessment Approaches and Theories Used

Based on the analysis, the majority of articles employed authentic assessment approaches such as rubrics, portfolios, and direct observations. As summarized in Table 3, 60% of the studies applied authentic assessments, 24% used performance-based approaches, and 16% fell into other categories. This distribution reflects a paradigm shift from outcome-based assessment toward process-oriented and contextually grounded evaluation.

Authentic assessment is a form of evaluation that allows students to apply their knowledge and skills in real-life situations. This creates a strong connection between classroom learning and everyday contexts (Marwa et al., 2024). Such approaches emphasize critical thinking, creativity, collaboration, and contextual problem-solving, core competencies needed for 21st-century STEM learning. Importantly, these practices align with constructivist learning theories and the principles of formative assessment, which position assessment as an integral part of the learning process (Black & Wiliam, 2009; Gulikers et al., 2004).

Although theoretical frameworks are not always explicitly articulated, in practice, many researchers have applied performance-based learning principles consistent with constructivist and formative perspectives. This limited theoretical integration may be due to the fact that most studies prioritize the development and application of practical assessment instruments rather than engaging in deeper theoretical discourse.

**Table 3**. Distribution of Assessment Approaches and Instruments Used in STEM Projects

| Approach | Commonly Used Instruments | n | Percentage |
|---|---|---|---|
| Authentic Assessment | Rubrics, Portfolios, Observations | 30 | 60% |
| Performative Assessment | Presentation, Physical Products | 12 | 24% |
| Other Approaches | Reflections, Project Reports | 8 | 16% |

Beyond these dominant approaches, several studies explored alternative strategies. One example is the use of ePortfolios to document students' learning progress and promote reflective practices in STEM teacher development programs (French et al., 2023). This approach is consistent with constructivist and formative assessment principles. Similarly, unguided student reflections have been shown to foster metacognitive awareness; however, their lack of structure limits their effectiveness as formal evaluation tools (Matheson et al., 2017). These examples illustrate the diversity of assessment practices in STEM and highlight the importance of selecting strategies aligned with specific learning goals.

### 3. Geographical Context and Characteristics of Participants

As summarized in Table 3, more than half of the research on performance assessment in STEM projects was conducted in Southeast Asia (56%). This dominance can be attributed not only to national policies that mandate the integration of STEM into school curricula (Arlinwibowo et al., 2021), but also to factors such as government-driven funding schemes, curriculum reforms aligned with 21st-century skills, and strong incentives for publication in international journals. These conditions have encouraged researchers in developing countries to adopt performance-based assessments as a strategy for preparing students to face the challenges of the Industrial Revolution 4.0.

In contrast, Europe (18%) and North America (12%) contributed fewer studies, but with distinctive characteristics. European research often integrates cross-disciplinary and technology-enhanced approaches. For example, Van Hoe et al. (2024) reported the use of digital portfolios in secondary schools to assess collaborative inquiry and problem-solving skills. Meanwhile, North American studies are primarily situated in higher education, emphasizing alignment with professional practice. One example is the implementation of a two-stage exam in undergraduate engineering courses, designed to evaluate conceptual understanding and higher-order thinking while mirroring authentic engineering practices (Koretsky et al., 2022) ,many examples illustrate how mature research infrastructures in these regions enable complex, technology-driven assessment models (Bybee, 2013; Honey et al., 2014).

Regarding educational levels, the majority of studies were conducted with junior high school, senior high school, and early college students. This concentration reflects the global importance of equipping adolescents with foundational STEM competencies at a transitional stage of education. Earlier work by Darling-Hammond et al. (2017) similarly noted that performance assessments are particularly impactful when applied during the transition to higher education or the workforce. However, participant demographics beyond student levels, such as teacher involvement, gender representation, or disciplinary focus, remain underexplored, limiting a deeper understanding of contextual differences in implementation.

**Table 4.** Geographical Distribution and Student Levels in STEM Performance Assessment Research

| Territory | Student Level | Number of Articles | Presents |
|---|---|---|---|
| Southeast Asia | Junior High School – College | 28 | 56% |
| Europe | High School and College Students | 9 | 18% |
| North America | University Students | 6 | 12% |
| Multi-Country | Mix of levels and regions | 7 | 14% |

Overall, the geographical distribution of STEM performance assessment research reveals both concentration and divergence. Southeast Asia leads in volume, reflecting policy-driven educational reforms and strong research incentives. Europe and North America, while contributing fewer studies, demonstrate greater emphasis on digital integration and professional alignment. These regional variations suggest differing priorities and levels of readiness in implementing authentic assessments. For the future, this diversity highlights the need for more cross-regional collaboration to balance contextual innovations with globally transferable frameworks for STEM education.

**4. Data Collection and Analysis Methods**

As shown in Table 5, the majority of STEM performance assessment studies employed quantitative approaches (68%), primarily using presentations and rubric-based evaluations. Qualitative methods, such as classroom observations and teacher reflections, accounted for 16%, while another 16% adopted mixed-method designs. This distribution reflects researchers' preference for measurable outcomes but also shows emerging interest in capturing contextual and process-oriented dimensions of STEM learning.

**Table 5.** Data Collection Methods and Analytical Approaches in STEM Performance Assessment

| Data Collection Methods | Analytical Approach | Number of Articles | Presents |
|---|---|---|---|
| Presentations & Rubric | Quantitative | 34 | 68% |
| Observation & Reflection | Qualitative | 8 | 16% |
| Mixed-method | Mixture | 8 | 16% |

Quantitative studies typically relied on structured rubrics and descriptive statistics. For example, (Yakob et al., 2021). implemented STEM-based science learning with performance assessments using observation sheets, rubrics, and questionnaires, analyzing the data quantitatively to evaluate students' habits of mind.

Although fewer in number, qualitative approaches provided valuable insights into classroom dynamics and teacher perspectives. ElSayary (2021) used reflective observation to capture how teachers adapted performance assessments to foster student engagement and problem-solving, offering a richer understanding of the learning process beyond numerical scores.

Mixed-methods studies combined numerical results with narrative evidence. Zhumabay et al. (2024) assessed the impact of a STEM teacher training program by integrating survey scores with qualitative feedback, thereby revealing both shifts in self-efficacy and the contextual challenges teachers faced.

Although these studies collectively highlight the practical application of performance-based learning principles, theoretical integration remains limited. This may be due to the fact that many researchers focus on developing and validating assessment instruments rather than anchoring them in explicit theoretical frameworks. As a result, while practice demonstrates alignment with constructivist and formative assessment perspectives, the underlying conceptual discourse is less frequently articulated.

**5. Research Gaps and Contributions**

The analysis of 50 reviewed articles reveals several key gaps in the practice and study of performance assessment in STEM projects. First, only about 40% of studies explicitly mentioned an assessment theory as the foundation for their approach. This limited integration suggests that many researchers prioritize practical implementation over theoretical grounding. One possible explanation is that assessment theory in STEM contexts is less developed compared to other areas, making researchers focus more on instrument application than on conceptual alignment. However, a clear theoretical basis is essential for ensuring that assessments are valid, reliable, and aligned with learning objectives (Xu & Brown, 2016).

Second, only 24% of studies conducted systematic validation and reliability checks of their instruments. Without validation, rubrics and other performance tools risk producing biased or

inconsistent outcomes. For example, Dwi Masinta (2016) assessed problem-solving skills with a rubric that lacked sufficient validation, potentially compromising fairness and accuracy. Compared to multiple-choice tests, rubrics are more complex to validate because they require clearly defined criteria and inter-rater consistency (Taherdoost, 2018).

In addition to these concerns, a third significant gap is the limited involvement of teachers or practitioners in assessment design, which is found in only 20% of the reviewed studies. This is problematic because teacher participation is vital to ensure that assessments remain contextual, relevant, and sustainable in classroom practice. Collaborative design-based research approaches emphasize this partnership, yet many studies still treat teachers as implementers rather than co-designers (Anderson & Shattuck, 2012).

Fourth, longitudinal studies remain rare, with only 16% of articles taking a long-term perspective. This lack of extended evaluation prevents researchers from fully understanding how performance assessments impact students' skills over time. For example, Mutia Munifatuzzahroh & Linuwih (2024) designed a STEM-based performance assessment to evaluate creativity within a single project. However, it did not explore longer-term outcomes across multiple projects or educational stages. Longitudinal research could provide more substantial evidence of how assessments shape students' development across semesters or schooling transitions (Bryk, 2020).

**Table 6.** Research Gaps in STEM Project Performance Assessment Practices

| Research Gaps | N | Presents |
|---|---|---|
| Not explicitly mentioning assessment theory | 20 | 40% |
| Minimal instrument validation/reliability | 12 | 24% |
| Minimal involvement of teachers/assessment designers | 10 | 20% |
| Limited longitudinal studies | 8 | 16% |

Taken together, these findings highlight the need to strengthen the theoretical, methodological, and participatory foundations of STEM performance assessments. As summarized in Table 6, 40% of studies did not explicitly mention theory, 24% lacked proper validation, and 20% minimally involved teachers. Moving forward, research contributions should focus on: (1) articulating stronger theoretical frameworks, (2) establishing rigorous standards, (3) involving teachers more actively in the design process, and (4) conducting longitudinal studies to evaluate long-term learning impacts.

## CONCLUSION

This study demonstrates that performance assessments in STEM projects have experienced a notable growth in scholarly attention over the past decade, peaking in 2021. Most studies employed authentic assessment strategies such as rubrics, portfolios, and observations that reflect constructivist principles and the notion of assessment for learning, although explicit theoretical integration remains limited. Geographically, research is concentrated in Southeast Asia, particularly Indonesia and Malaysia, with a primary focus on junior high to early higher education levels. Methodologically, quantitative approaches predominate, while mixed-method studies offer more comprehensive insights.

Despite these advances, several gaps persist. First, only a limited number of studies explicitly grounded their approaches in assessment theory, suggesting a tendency to prioritize practical application over conceptual development. Second, the validation and reliability of assessment instruments remain underdeveloped, raising concerns about the accuracy and fairness of the results. Third, teacher involvement in assessment design and implementation remains minimal, which may compromise the contextual relevance and sustainability of the assessment. Finally, longitudinal studies are scarce, making it challenging to assess the long-term effects of performance assessments on students' skills and learning trajectories.

The study also carries limitations. The review was restricted to 50 articles, which may not capture the full diversity of global practices. The dominance of Southeast Asian studies limits generalizability, while the thematic scope primarily focuses on classroom-based assessments, leaving out informal learning contexts and emerging digital tools.

Future research should therefore address these limitations in several ways. Priority should be given to the development and rigorous validation of assessment instruments that ensure reliability and adaptability. Implementation studies are needed to examine how performance assessments operate in diverse classroom settings, while longitudinal research could capture their long-term effects across multiple projects and stages of education. Expanding research into advanced physics, interdisciplinary STEM, and AI-supported assessments will enrich the field. Moreover, strengthening teacher professional development and policy support is crucial for embedding authentic, contextual, and sustainable assessment practices.

Overall, performance assessments hold strong potential as a key framework for evaluating 21st-century skills in STEM education. By addressing theoretical, methodological, and participatory gaps, future research can enhance both the academic rigor and practical application of STEM performance assessments.

## ACKNOWLEDGMENTS (OPTIONAL)

The researchers would like to thank the Faculty of Teacher Training and Education at Mulawarman University, Indonesia, for their support in funding and publishing this research.

## REFERENCES

Anderson, T., & Shattuck, Julie. (2012). Design-Based Research: A Decade of Progress in Education Research? *Educational Researcher*, 41(1), 16–25. https://doi.org/10.3102/0013189X11428813

Arlinwibowo, J., Retnawati, H., & Kartowagiran, B. (2021). *How to Integrate STEM Education in The Indonesian Curriculum? A Systematic Review*. https://doi.org/10.31643/2021.03

Azizah, R. D. A. F. Z., Sulaeman, N. F., & Komariyah, L. (2022). Performance Assessment in STEM-EDP Project for High School Students: Validity Aspect. *December 2022*, *6*(2), 130–139. http://e-journal.hamzanwadi.ac.id/index.php/kpj/index

Black, P., & Wiliam, D. (2009). Developing The Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Bryk, A. S. (2020). Learning to Improve. *Stanford Social Innovation Review*. https://doi.org/10.48558/4802-QT84

Bybee, R. W. (2013). *Challenges and Opportunities: The Case for Education*. www.nsta.org/permissions.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*.

Diana, N., Yohannes, & Sukma, Y. (2021). The Effectiveness of Implementing Project-Based Learning (PjBL) Model in STEM Education: A Literature Review. *Journal of Physics: Conference Series*, *1882*(1). https://doi.org/10.1088/1742-6596/1882/1/012146

Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring Relations Between Formative and Summative Assessment. In J. Dolin & R. Evans (Eds.), Transforming Assessment: Through an Interplay Between Practice, Research and Policy (pp. 53–80). *Springer International Publishing*. https://doi.org/10.1007/978-3-319-63248-3_3

Dwi Masinta, A. (2016). Validasi Rubrik Sebagai Pedoman Untuk Menilai Kemampuan Pemecahan Masalah Siswa SMA Berbasis Inkuiri dengan Menggunakan Tes Uraian. *Unnes Physics Education Journa*l, *5(2)*. http://journal.unnes.ac.id/sju/index.php/upej

ElSayary, A. (2021). Using a Reflective Practice Model to Teach STEM Education in a Blended Learning Environment. *Eurasia Journal of Mathematics, Science and Technology Education*, *17(2)*. https://doi.org/10.29333/ejmste/9699

Estapa, A. T., & Tank, K. M. (2017). Supporting Integrated STEM in The Elementary Classroom: A Professional Development Approach Centered on An Engineering Design Challenge. *International Journal of STEM Education*, *4*(1). https://doi.org/10.1186/s40594-017-0058-3

French, D. , Rhew, B. , H. S., & Dorsey, S. (2023). Using ePortfolios as An Authentic Assessment Tool for Integrated STEM Professional Development. *Contemporary Issues in Technology and Teacher Education*, *23*(1).

Gulikers, J. T. M., Bastiaens, T. J., Kirschner, P. A., & Bastioens, T. J. (2004). A Five-Dimensional Framework for Authentic Assessment. *Technology Research and Development*, *52*(3), 67–86.

Hakim, A., Efwinda, S., Damayanti, P., Syam, M., Mutmainah, O., & Norsaputra, A. (2023). Improving Climate Literacy of High School Students Through Physics Teaching Materials Based on STEM. *Jurnal Pendidikan IPA Indonesia*, *12*(4), 526–536. https://doi.org/10.15294/jpii.v12i4.43188

Hartini, S., Mariani, I., Misbah, & Sulaeman, N. F. (2020). Developing of Students Worksheets Through STEM Approach to Train Critical Thinking Skills. *Journal of Physics: Conference Series*, *1567*(4). https://doi.org/10.1088/1742-6596/1567/4/042029

Honey, M. A., Pearson, G., & Schweingruber, H. (2014). STEM Integration in K-12 Education: Status, Prospects, and An Agenda for Research. In *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research*. National Academies Press. https://doi.org/10.17226/18612

Koretsky, M. D., McColley, C. J., Gugel, J. L., & Ekstedt, T. W. (2022). Aligning Classroom Assessment with Engineering Practice: A Design-Based Research Study of A Two-stage Exam with Authentic Assessment. *Journal of Engineering Education*, *111*(1), 185–213. https://doi.org/10.1002/jee.20436

Marwa, N. W. S., Pitria, P. R., & Madani, F. (2024). Development of Authentic Assessment of 21st-century Skills in Kurikulum Merdeka. *Inovasi Kurikulum*, *21*(2), 635–646. https://doi.org/10.17509/jik.v21i2.67273

Matheson, A., Wood, L., & Franklin, S. V. (2017). *Guided and Unguided Student Reflections*. http://arxiv.org/abs/1710.05259

Mellyzar, M., Andriani, R., Lukman, I. R., Muttakin, M., Wati, E., Pasaribu, A. I., & Fadli, Mhd. R. (2024). Digital Assessment Tools: As a Media for Assessing High School Science Learning. *Jurnal Penelitian Pendidikan IPA*, *10*(3), 1362–1374. https://doi.org/10.29303/jppipa.v10i3.6416

Miyati, A. A. (2019). Penerapan Model Pembelajaran Generatif Dengan Setting Kelas Kooperatif STAD (Student Team Achievement Division) Untuk Meningkatan Pemahaman Konsep Kimia dan Kinerja Ilmiah Siswa Kelas X MIPA 7 SMA Negeri 5 Denpasar Tahun Pelajaran 2016/2017. *Jurnal Santiaji Pendidikan (JSP)*, *9*, 189–199. https://doi.org/10.36733/jsp.v9i2.400

Mutia Munifatuzzahroh, U., & Linuwih, S. (2024). Development of STEM-Based Performance Assessments to Analyze Students' Creativity Physics Communication. *Physics Communication*, *8*(2), 77–86. http://journal.unnes.ac.id/nju/index.php/pc

Pryantini, P. A., Sumantri, M., & Widiana, I. W. (2016). Pengembangan Asesmen Proyek Pada Pembelajaran IPA Kelas V Semester Genap di SD Negeri 1 Sangsit Tahun Pelajaran 2015/2016. *Mimbar PGSD Undiksha*, *4*(1). https://doi.org/10.23887/jjpgsd.v4i1.7386

Putri, F. S., & Istiyono, E. (2017). The Development of Performance Assessment of STEM- Based Critical Thinking Skill in the High School Physics Lessons. *International journal of environmental & science education*, *12*, 1269–1281.

Putri, S. A., Sulaeman, N. F., & Putra, P. D. A. (2022). Trend of Technological Pedagogical Content Knowledge (TPACK) for Pre-service Science Teachers: A Historical Review. *Jurnal Pendidikan Fisika*, *10*(2), 165–175. https://doi.org/10.26618/jpf.v10i2.7801

Retnawati, H., Kardanova, E., Sumaryanto, S., Prasojo, L. D., Jailani, J., Arliani, E., Hidayati, K., Susanti, M., Lestari, H. P., Apino, E., Rafi, I., Rosyada, M. N., Tuanaya, R., Dewanti, S. R., Sotlikova, R., & Kassymova, G. K. (2024). A Systematic Review of the Use of Technology in Educational Assessment Practices: Lesson Learned and Direction for Future Studies. *International Journal of Robotics and Control Systems*, *4*(4), 1656–1693. https://doi.org/10.31763/ijrcs.v4i4.1572

Septiani, A., & Rustaman, N. Y. (2017). Implementation of Performance Assessment in STEM (Science, Technology, Engineering, Mathematics) Education to Detect Science Process Skill. *Journal of Physics: Conference Series*, *812*(1). https://doi.org/10.1088/1742-6596/812/1/012052

Sulaeman, N., Efwinda, S., & Putra, P. D. A. (2022). Teacher Readiness in STEM Education: Voices of Indonesian Physics Teachers. *Journal of Technology and Science Education*, *12*(1), 68–82. https://doi.org/10.3926/jotse.1191

Sulaeman, N. F., Putra, P. D. A., Mineta, I., Hakamada, H., Takahashi, M., Ide, Y., & Kumano, Y. (2021a). Exploring Student Engagement in STEM Education through the Engineering Design Process. *Jurnal Penelitian Dan Pembelajaran IPA*, *7*(1), 1. https://doi.org/10.30870/jppi.v7i1.10455

Taherdoost, H. (2018). Validity and Reliability of the Research Instrument : How to Test the Validation of a Questionnaire/Survey in Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3205040

Van Hoe, A., Wiebe, J., Rotsaert, T., & Schellens, T. (2024). The Implementation of Peer Assessment as A Scaffold During Computer-supported Collaborative Inquiry Learning in Secondary STEM Education. *International Journal of STEM Education*, *11*(1). https://doi.org/10.1186/s40594-024-00465-8

Wahyuni, N. L. P. W., Wibawa, M. C., & Renda, N. T. (2018). Pengaruh Model Pembelajaran Kooperatif Tipe Group Investigation Berbantuan Asesmen Kinerja Terhadap Keterampilan Proses Sains. *International Journal of Elementary Education*, *2*(3), 202–210. https://ejournal.undiksha.ac.id/index.php/IJEE

Windyariani, S., & Setiono, S. (2024). Performance Assessment Creative Thinking Rubric in Science, Technology, Engineering, and Mathematics (STEM) Learning. *JISE*, *13*(2), 64–73. https://journal.unnes.ac.id/journals/jise/

Wulan, A. R. (2018). *Menggunakan Asesmen Kinerja: Untuk Pembelajaran Sains dan Penelitian* (Vol. 3). UPI Press.

Xu, Y., & Brown, G. T. L. (2016). Teacher Assessment Literacy in Practice: A Reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/https://doi.org/10.1016/j.tate.2016.05.010

Yakob, M., Hamdani, H., Sari, R. P., Haji, A. G., & Nahadi, N. (2021). Implementation of Performance Assessment in STEM-Based Science Learning to Improve Students' Habits of Mind. *International Journal of Evaluation and Research in Education*, *10*(2), 624–631. https://doi.org/10.11591/ijere.v10i2.21084

Zhumabay, N., Yelemessova, Z., Balta, N., Abylkassymova, A., Bakytkazy, T., & Marynowski, R. (2024). Designing effective STEM courses: a mixed-methods study of the impact of a STEM education course on teachers' self-efficacy and course experiences. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1276828