

# Mapping Performance Assessment Tools of Practical Skills in Chemistry Education: Evidence from a Systematic Review

Faradhilla Bonita<sup>1</sup>, Nahadi<sup>1\*</sup>, Heli Siti Halimatul Munawaroh<sup>2</sup>

<sup>1</sup> Department of Chemistry Education, Indonesia University of Education, Bandung, Indonesia.

<sup>2</sup> Department of Chemistry, Indonesia University of Education, Bandung, Indonesia.

Received: January 12, 2026

Revised: April 04, 2026

Accepted: June 05, 2026

Published: June 06, 2026

Corresponding Author:

Nahadi

[nahadi@upi.edu](mailto:nahadi@upi.edu)

DOI: [10.29303/jppipa.v12i5.14258](https://doi.org/10.29303/jppipa.v12i5.14258)

 Open Access

© 2026 The Authors. This article is distributed under a (CC-BY License)



**Abstract:** Performance assessment has emerged as an important approach in chemistry education to address the limitations of conventional testing. This study aims to systematically review the implementation of performance-based assessments in chemistry, focusing on instrument typology, educational context, and targeted competencies. Guided by the PRISMA protocol, a comprehensive search was conducted in the Scopus database for peer-reviewed research articles published between 2015 and 2025. Due to the very strict inclusion criteria that focused exclusively on validated chemistry-specific instruments, nine selected articles were synthesized. Results indicate that analytical rubrics remain the most dominant format (67% of reviewed tools), with digital video-based assessments emerging as a recent trend, while 88% of implementations are skewed towards higher education. These authentic tools consistently report positive impacts on students' practical competence and procedural engagement. However, pedagogical challenges remain: current instruments often fail to capture deep conceptual reasoning due to the high cognitive load experienced by students during active tasks. This review concludes that while performance assessments effectively measure technical compliance, global standardization of the Evidence-Centered Design (ECD) framework is urgently needed to address the current fragmentation of instruments and systematically align laboratory evidence with measurable conceptual milestones.

**Keywords:** Authentic assessment; Chemistry education; Laboratory skills; Performance assessment; Scientific reasoning

## Introduction

Chemistry learning across the educational continue from secondary education to higher education still faces fundamental challenges in an assessment system dominated by conventional cognitive assessments, such as multiple-choice tests, essay exams, and written laboratory reports. While these assessment methods are effective in measuring conceptual mastery, various studies have shown that this approach has significant limitations in evaluating important dimensions of scientific competence, such as laboratory practical skills, scientific reasoning, and the ability to construct evidence-based scientific arguments (Al-Kafawein & Al-Hilal, 2025; J. Harsh et al., 2017; Veale et al., 2020). Yet, chemistry is inherently an experimental discipline,

where inquiry, empirical observation, laboratory decision-making, and data-driven reasoning are at the heart of scientific practice (Shofiyah et al., 2026). This mismatch between the scientific characteristics of chemistry and the assessment practices used has the potential to hinder the development of students' holistic and authentic scientific competence (Pahriah et al., 2025; Putri & Hasruddin, 2025).

The limitations of traditional assessment are increasingly apparent in the context of laboratory practical. Veale et al. (2020) revealed that written report-based assessments often fail to detect weaknesses in students' practical skills, as students may present seemingly perfect data without truly demonstrating technical competence or process understanding. Similar findings have been reported in pre-university education,

### How to Cite:

Bonita, F., Nahadi, & Munawaroh, H. S. H. (2026). Mapping Performance Assessment Tools of Practical Skills in Chemistry Education: Evidence from a Systematic Review. *Jurnal Penelitian Pendidikan IPA*, 12(5), 106–113. <https://doi.org/10.29303/jppipa.v12i5.14258>

where procedural and recipe-like laboratory reports tend not to encourage in-depth analysis and actually limit meaningful learning (Yusof et al., 2021). This situation suggests that assessment systems that focus solely on end-product knowledge are no longer adequate to represent true chemistry competency.

In response to these limitations, performance assessment, or authentic assessment, has developed, emphasizing the direct assessment of scientific processes, products, and skills in the context of tasks that mimic real-life scientific practice. This assessment is relevant for evaluating various chemistry learning activities, such as conducting laboratory experiments, project-based learning, beginning student research, and developing and presenting scientific arguments (Albay & Eisma, 2021; Deng & Wang, 2017; Liu et al., 2021). Various studies report that performance-based assessments can improve experimental skills, strengthen scientific thinking, and develop students' scientific communication and reflection skills more meaningfully than conventional assessments (Albay & Eisma, 2021; J. Harsh et al., 2017; Panadero et al., 2025).

Performance assessments include analytical rubrics, which have been demonstrated to be an effective approach to increasing objectivity and consistency. These analytical rubrics provide teachers with the opportunity to systematically assess each dimension of student performance, from equipment preparation and procedure implementation to collaboration and analysis of results (Asli et al., 2024; Panadero et al., 2025). Furthermore, student involvement in the assessment process, through self-assessment, using rubrics, can enhance self-reflection and learning motivation. However, the implementation of rubrics used in performance assessments is often manual, requiring time and intensive observation by teachers (Vasileiadou & Karadimitriou, 2021).

In line with these developments, while authentic performance assessment and digital tracking technologies have grown rapidly in general education, their systematic application within chemistry education remains surprisingly scarce. To date, only a limited number of empirical studies have successfully developed and validated specialized performance instruments in this discipline. The Performance Assessment of Undergraduate Research Experiences (PURE) instrument, for example, was designed to assess the development of students' scientific thinking skills through real-world problem-based research experiences (J. Harsh et al., 2017). Furthermore, Liu et al. (2021) developed a Delphi consensus-based rubric to assess the quality of chemistry experiment videos created by pre-service teachers, focusing on content organization, experimental procedures, and presentation quality.

Furthermore, the Evidence-Centered Design (ECD) framework was used to ensure alignment between learning objectives, performance evidence, and assessment instruments in chemistry lab assessments (Zhang & Wink, 2021). Panadero et al. (2025) study also highlighted the role of rubric design and feedback type in influencing student attention patterns and performance, underscoring the importance of a carefully designed assessment structure.

However, despite the growing global emphasis on authentic education, empirical literature evaluating performance assessment specifically within chemistry education remains surprisingly scarce and deeply fragmented. Existing studies operate in methodological isolation, focused predominantly on university-level contexts (undergraduate and pre-service teachers) while leaving secondary education (K-12) severely underrepresented. Furthermore, while rubric-based frameworks dominate the literature—ranging from time-intensive manual tools like the Performance Assessment of Undergraduate Research Experiences (PURE) (J. A. Harsh, 2016) to multi-indicator video questionnaires (Liu et al., 2021) and highly theoretical Evidence-Centered Design (ECD) frameworks (Zhang & Wink, 2021) there is a stark absence of comparative synthesis regarding their design, feasibility, and effectiveness. This fragmentation inflicts a severe practical impact; chemistry educators face a confusing dichotomy and are often forced to choose blindly among isolated instruments without a standardized compass to guide their selection based on practical constraints, classroom sizes, or targeted scientific skills.

Therefore, a Systematic Literature Review (SLR) is urgently required at this juncture to consolidate these scattered findings and untangle the theoretical fragmentation. This study offers a distinctive novelty by establishing a unified conceptual framework based on a rigorous synthesis of the limited existing literature. Specifically, this review comprehensively maps: (1) the predominant forms of performance-based rubrics used in chemistry learning, (2) the comparative contexts of their implementation spanning from secondary education to undergraduate settings, and (3) the explicit scientific competencies they evaluate. By highlighting previous literature gaps and filtering the current landscape, this study provides a much-needed standardized compass and a practical toolkit for chemistry educators and researchers to design and deploy realistic, validated, and authentic performance assessment instruments.

## Method

This study employed a Systematic Literature Review (SLR) design to systematically identify, analyze, and synthesize studies addressing performance assessment in chemistry education. The SLR approach was chosen because it allows researchers to obtain a comprehensive overview of the assessment methods, application contexts, and scientific skills focused on across studies, while simultaneously identifying existing research gaps. The review process strictly followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, through systematic steps: formulating the research question, conducting a transparent literature search, applying strict inclusion and exclusion criteria, conducting a quality assessment, extracting data, and synthesizing the findings using a qualitative-descriptive approach.

The primary literature search was conducted using the Scopus database to ensure a comprehensive, multidisciplinary coverage of high-quality, peer-reviewed articles across multiple international publishers (such as ACS, RSC, Wiley, and Elsevier). To ensure no relevant studies were overlooked given the specialized nature of the topic, cross-referencing and supplementary searches were performed using Google Scholar and ERIC. To eliminate search string ambiguity and ensure reproducibility, the final syntax combined relevant terms using Boolean operators (AND/OR) and was executed as follows:

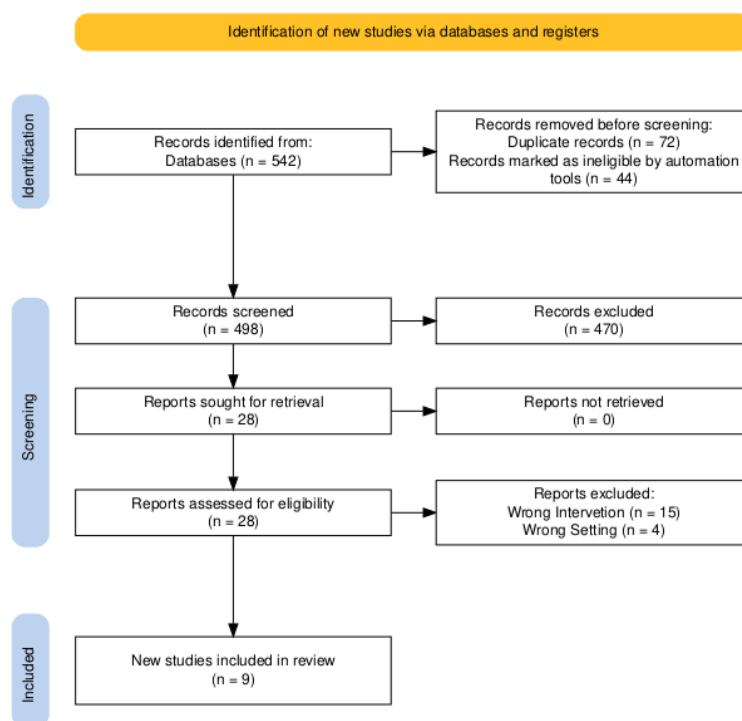
("performance assessment" OR "performance-based assessment" OR "authentic assessment") AND ("chemistry education" OR "chemistry learning") AND ("scientific skills" OR "practical skills" OR "rubric\*").

The initial search phase yielded 542 records from the databases. The complete screening process, including duplication removal and exclusion phases based on abstract and full-text accessibility, is visually mapped in the PRISMA Flow Diagram (Figure 1).

To refine the initial pool of articles, explicit operational boundaries were defined. Studies were selected if they were original, peer-reviewed research articles published between 2015 and 2025, written in English, and explicitly focused on designing, validating, or implementing performance-based assessment tools within a chemistry context. Conversely, review papers, non-research publications, and studies outside the field of chemistry education were excluded. The formalized criteria are detailed in Table 1.

**Table 1.** Inclusion & exclusion criteria

Inclusion	Exclusion
Scopus-indexed research articles (2015–2025)	Non-Scopus or published outside 2015–2025
Original research articles	Review articles and non-research publications
Performance assessment	Non-performance assessment
Chemistry content	Studies not related to chemistry content



**Figure 1.** PRISMA Flow Diagram

To ensure methodological rigor beyond indexing compliance, the selected articles underwent a dedicated quality appraisal stage. Each study was evaluated against three core qualitative benchmarks adapted from standard appraisal frameworks: (1) clarity and relevance of the stated research objectives, (2) the structural validity and appropriateness of the described chemistry performance instrument, and (3) the logical alignment between the empirical data collected and the final conclusions drawn. Only studies that fully satisfied these quality thresholds were approved for final synthesis, ensuring that the resulting pool of nine articles possessed high methodological credibility.

Data from the final nine qualified articles were systematically extracted into a standardized matrix to facilitate thematic analysis. The extraction process was highly specific, capturing the following metadata and structural components from each study: (1) Article Identity (Author, Year, and Journal), (2) Assessment Form or Format (e.g., rubrics, video assessment questionnaires, Evidence-Centered Design frameworks), (3) Targeted Scientific Competencies/Skills, (4) Educational Level and Context (e.g., Secondary Education/K-12 or Higher Education/Undergraduate), and (5) Key Empirical

Findings. Following extraction, a qualitative-descriptive synthesis was performed, grouping the data into thematic categories to map the current landscape of performance assessment in chemistry education and reveal overarching research gaps.

## Result and Discussion

Based on a systematic synthesis of the nine reviewed articles, performance-based assessment in chemistry education is applied in a variety of forms and contexts. Overall, the literature confirms that performance-based assessment provides a more comprehensive, authentic, and holistic representation of student competency than conventional cognitive measures. However, a critical examination of these findings reveals significant thematic patterns, structural imbalances, and theoretical gaps that define the current landscape.

Table 2 presents a summary of the key findings from the nine reviewed studies, highlighting the form of performance assessment used and the scientific skills assessed in chemistry learning.

**Table 2.** Literature review results

Article identity	Assessment form	Focus skills	Educational level	Main findings
Liu et al. (2021)	Video assessment questionnaire: developed using the Delphi method, it was used to assess chemistry experiment videos created by prospective teachers. The rubric includes 20 indicators across three aspects	Chemistry experiment teaching skills, including: <ul style="list-style-type: none"> <li>• content organization</li> <li>• experimental operations</li> <li>• video quality.</li> </ul>	Undergraduate	The use of the video rubric significantly improved the laboratory teaching performance of prospective teachers, particularly in the experimental operations aspect. However, the rubric did not significantly improve content organization or video quality. The rubric was validated as an easy to use and effective tool for self-reflection.
Al-Kafawein & Al-Hilal (2025)	<ul style="list-style-type: none"> <li>• Multiple-choice questions (MCQs) (formative &amp; summative).</li> <li>• Performance-based assessment (practical project). Graded using a special rubric.</li> </ul>	<ol style="list-style-type: none"> <li>1. Conceptual understanding (via MCQs).</li> <li>2. Practical application skills: experimental design, data analysis, problem-solving, critical thinking, and scientific communication (via project).</li> </ol>	Undergraduate	Significant improvements were seen in conceptual understanding (MCQ scores) and application skills (project scores) throughout the quarter. Performance-based assessment consistently yielded higher average scores than MCQs, demonstrating its strength in assessing knowledge application. The two methods complement each other.
Yusof et al. (2021)	Project-based learning (chemistry group project) with assessment through proposals, final reports, laboratory observations, and peer assessment.	Practical skills, scientific writing, literature searches, collaboration, and project management were developed.	Undergraduate	The average practical grade was 84.9% and scientific writing was 84.4%. This approach improved motivation, conceptual understanding, and social and

Article identity	Assessment form	Focus skills	Educational level	Main findings
				scientific reasoning skills in pre-university students.
Harsh et al. (2017)	PURE (performance assessment of undergraduate research experiences) instrument: A performance test with open-response tasks contextualized in real-world chemistry problems from primary literature.	Scientific thinking	Undergraduate	Participation in undergraduate research (URE) contributed to significant improvements in 10 of the 11 targeted skills. Weak correlations were found between student self-reports and actual measured performance, indicating the need for direct performance data for valid evaluation.
Zhang & Wink (2021)	Evidence-centered design (ECD) in acid-base titration lab, with assessment through lab reports, graphs, and post-lab questions.	Lab report structure	Undergraduate	Students are able to demonstrate procedural skills and the use of mathematics as a tool, but lack reasoning skills related to curve patterns and in-depth conceptual understanding of chemical species during titration.
Albay & Eisma (2021)	Performance assignment assessment (based on the design thinking process) in the form of a teaching demonstration. Assessment was conducted using a rubric covering various teaching performance indicators and video recordings.	Skills in designing and implementing creative, interactive, and student-centered learning were assessed. Integration of 21 <sup>st</sup> century skills: collaboration, empathy, non-linear problem-solving, and presentation.	Undergraduate	The experimental group (using the design thinking process) significantly outperformed the control group in all aspects of the teaching demonstration. The design thinking process effectively facilitated the creation of a creative and interactive classroom.
Veale et al. (2020)	Direct laboratory observation rubric, assessed by instructors based on students actual performance during the lab.	Organic lab skills (reaction setup, anhydrous technique, follow up work, distillation), experimental planning, and result interpretation are assessed.	Undergraduate	Direct rubric-based assessment increases student engagement, especially in low-performing groups (~10% improvement). Provides immediate feedback and reduces the bias of written report-based assessment.
Panadero et al. (2025)	Rubric with manipulation of the order of performance levels combined with various types of feedback (process-based, product-based, rubric-based, no feedback). Monitored using eye-tracking and think-aloud protocols.	Engagement with the rubric (visual attention patterns & cognitive strategies) and task performance landscape analysis.	Undergraduate	<ol style="list-style-type: none"> <li>1. The order of performance levels influences visual engagement. The highest level placed first attracts the most attention.</li> <li>2. Process- and product-based feedback yields the greatest performance improvements.</li> <li>3. Rubric-based feedback increases gaze transitions between the rubric and the task, indicating deeper cognitive engagement.</li> </ol>
Deng & Wang (2017)	Written scientific argumentation Tasks (5 chemistry-context tasks). Assessed using a criteria rubric developed based on the dimensions of structure (claim, evidence, warrant,	Written scientific argumentation competence: the ability to formulate claims, present evidence, provide rationale (warrant), and refute opposing arguments (rebuttal) in writing.	Secondary education	Chinese students' written scientific argumentation competence is generally weak. Students find it easier to state claims and evidence than to provide warrants and rebuttals (rebuttal). Competence is influenced by grade level and school level, but not by gender.

Article identity	Assessment form	Focus skills	Educational level	Main findings
	rebuttal), content quality, logic, and language.			Different tasks (topics) affect student performance.

### *Methodological Bias in Target Competencies and Educational Level*

A critical analysis of target competencies and sample populations reveals a clear bias toward higher education settings (undergraduate research, preservice teachers, and university foundation years), significantly underrepresenting secondary education (K-12) (Deng & Wang, 2017). This gap is particularly problematic because existing evidence from secondary school contexts suggests that performance assessment can be successfully implemented and yield significant learning gains. For example, a study involving eighth-grade students (SMP) in real-laboratory science learning demonstrated that the use of performance assessment instruments significantly improved both cognitive and psychomotor outcomes, with N-gain values in the medium category and average psychomotor scores exceeding the minimum mastery criterion (Hartina et al., 2020). However, such studies remain exceptional rather than representative, highlighting the urgent need for more accessible, classroom-ready instruments for K-12 settings. This demographic imbalance significantly impacts the types of scientific competencies targeted by current instruments.

In university laboratory settings, direct observation rubrics and ECD frameworks excel at capturing physical procedural skills, numerical proficiency, and experimental data handling far more effectively than traditional written reports (Veale et al., 2020; Zhang & Wink, 2021). Furthermore, when integrated into authentic, long-term undergraduate research, frameworks like PURE demonstrate significant positive impacts across nearly all domains of advanced scientific thinking (J. Harsh et al., 2017).

However, when performance tasks are driven to evaluate complex cognitive and communicative dimensions, systemic weaknesses emerge. For example, in written scientific argumentation tasks, student performance remains stagnant at the lower structural level, characterized by the ability to formulate basic claims and identify surface-level evidence, but shows serious deficiencies in providing logical warrants and strong rebuttals (Deng & Wang, 2017).

### *The Performance vs. Deep Conceptual Reasoning Paradox: A Cognitive Load Perspective*

The most striking gap identified across empirical results is that while performance assessments successfully measure procedural execution and application of knowledge, higher-order conceptual

reasoning skills such as interpreting complex curves, molecular visualization, and understanding the specific behavior of chemical species remain consistently weak (Zhang & Wink, 2021). This phenomenon creates a critical educational paradox: why do authentic performance tools still fail to foster and capture deep conceptual understanding?

This limitation can be comprehensively explained through Cognitive Load Theory (CLT) (Yusof et al., 2021). During active chemistry experiments or complex performance tasks, students face a tremendous intrinsic cognitive load. Managing unfamiliar glassware, executing precise physical techniques (such as titrations), and maintaining safety protocols occupy a significant portion of their limited working memory capacity (Veale et al., 2020). Because the physical execution of performance tasks demands high cognitive expenditure, students experience a “split attention” effect (Yusof et al., 2021) and lack the cognitive space necessary to simultaneously process both abstract and microscopic chemical concepts behind the macroscopic changes they observe. Empirical studies applying CLT to science laboratory learning confirm this phenomenon. In an out-of-school heuristic science lesson designed for 5th to 8th grade students, researchers found that despite instructional materials being carefully structured to reduce extraneous cognitive load, a substantial proportion of students (approximately 33.5%) still failed to achieve measurable learning gains because their working memory was overwhelmed by the coordination of multiple information sources—instructions, equipment, and observation tasks—leaving insufficient capacity for deeper conceptual integration (Meissner & Bogner, 2013). As a result, students complete the tasks mechanically. The current rubric design fails to address this because it measures physical performance and conceptual reflection as isolated end-product milestones, rather than as structured and integrated steps (Seery et al., 2017; Veale et al., 2020).

### *Mechanisms of Impact on Learning and Cognitive Engagement*

Despite these conceptual limitations, the literature consistently demonstrates that performance-based tools have a significantly positive impact on student engagement and systemic equity compared to traditional testing. In comparative settings, performance-based tasks yield higher average scores for knowledge application than conventional Multiple Choice Questions (MCQs), acting as a valuable and non-

redundant complement to cognitive testing (Al-Kafawein, 2025).

Mechanistically, the transparency of direct observation rubrics promotes student equity by actively closing achievement gaps for low-performing student groups and eliminating the subjective assessment bias inherent in standard laboratory reports (Veale et al., 2020). However, the ultimate effectiveness of these benefits is largely determined by the instrument architecture; the order of performance levels, the layout of the rubric, and the method of delivering descriptive feedback determine how students allocate visual attention and manage cognitive stress during the task (Panadero et al., 2025).

#### *Comprehensive Synthesis of Research Gaps*

In conclusion, although performance assessments in chemistry education are highly effective in capturing complex, contextual, and procedural skills, three distinct research gaps remain: **Conceptual Reasoning Deficit:** Current instruments evaluate behavioral adherence and end-product data but fail to reduce procedural friction, leaving students with insufficient working memory to engage in in-depth chemical reasoning; **Secondary Education (K-12) Gaps:** There is a severe shortage of validated and realistic performance instruments designed for crowded secondary school classrooms, as 88% of the current literature is limited to well-resourced university settings; **Implementation Burden:** The current landscape offers complex, highly siloed, and time-consuming frameworks (such as PURE and ECD) without providing integrated, practical, and measurable guidance for classroom teachers, lacking extensive research support.

#### **Conclusion**

Based on a systematic review of nine selected studies, this review confirms that analytical rubrics and digital video-based assessments represent the most significant trends for objectively measuring chemistry procedural skills, particularly in higher education laboratory settings. While generally effective, these performance-based tools excel at evaluating structural appropriateness and experimental operations, while their capacity to measure in-depth conceptual reasoning and molecular-level visualization remains limited by students' high working memory demands during hands-on tasks. To address the identified fragmentation across current instrument formats, this study recommends global standardization of Evidence-Centered Design (ECD) based instruments to systematically align chemistry curricula and explicitly align laboratory evidence with conceptual milestones.

Due to the limited sample size of only nine highly indexed articles and the strong institutional bias toward undergraduate environments, these findings should be interpreted with caution to avoid overgeneralization, highlighting the urgent need for future research to validate measurable, low-cost performance frameworks specifically tailored for high school (K-12) chemistry classrooms.

#### **Acknowledgments**

The author would like to thank all the researchers whose work was referenced in this study for their valuable scientific contributions to the development of chemistry education. He also expresses his appreciation to colleagues and academics who provided constructive input and suggestions during the preparation and refinement of this manuscript.

#### **Author Contributions**

F.B performed the roles of Conceptualization, methodology, formal analysis, writing original draft preparation, writing review and editing. N & H.S performed the roles of Conceptualization, writing review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

#### **Funding**

All authors declare no external funding.

#### **Conflict of Interest**

Authors declare no conflict of interest.

#### **References**

- Al-Kafawein, J. (2025). Assessing Students' Progress in Chemistry: Using Multiple-Choice Questions and Performance-Based Assessments. *Journal of Curriculum and Teaching*, 14(1), 174–183. <https://doi.org/10.5430/jct.v14n1p174>
- Al-Kafawein, J., & Al-Hilal, M. (2025). Assessing Students' Progress in Chemistry: Using Multiple-Choice Questions and Performance-Based Assessments. *Journal of Curriculum and Teaching*, 14(1), 174–183. <https://doi.org/10.5430/jct.v14n1p174>
- Albay, E. M., & Eisma, D. V. (2021). Performance task assessment supported by the design thinking process : Results from a true experimental research. *Social Sciences & Humanities Open*, 3(January). <https://doi.org/10.1016/j.ssaho.2021.100116>
- Asli, N. F., Matore, M. E. E. M., & Yunus, M. M. (2024). Construct validity of primary trait writing rubrics based on assessment use argument (AUA) validation framework. *Heliyon*, 10(October). <https://doi.org/10.1016/j.heliyon.2024.e40053>
- Deng, Y., & Wang, H. (2017). Research on evaluation of Chinese students' competence in written scientific argumentation in the context of chemistry.

- Chemistry Education Research and Practice*, 18(1), 127–150. Retrieved from <https://pubs.rsc.org/en/content/articlehtml/2017/rp/c6rp00076b>
- Harsh, J. A. (2016). Designing performance-based measures to assess the scientific thinking skills of chemistry undergraduate researchers. *Chemistry Education Research and Practice*, 17(4), 808–817. <https://doi.org/10.1039/C6RP00057F>
- Harsh, J., Esteb, J. J., & Maltese, A. V. (2017). Evaluating the development of chemistry undergraduate researchers' scientific thinking skills using performance-data: First findings from the performance assessment of undergraduate research (PURE) instrument. *Chemistry Education Research and Practice*, 18(3), 472–485. <https://doi.org/10.1039/c6rp00222f>
- Hartina, L., Rosidin, U., & Suyatna, A. (2020). Pengaruh Penerapan Instrumen Performance Assessment pada Pembelajaran IPA Berbasis Laboratorium Real terhadap Hasil Belajar Siswa. *Jurnal Penelitian Pendidikan IPA (JPPIPA)*, 6(1), 25–31. <https://doi.org/10.29303/jppipa.v6i1.299>
- Liu, C., Chen, S., Liu, S., Liu, J., Wang, J., & Han, Y. (2021). A Delphi Study and Verification of Assessment of Chemistry Experiments Videos. *Croatian Journal of Education*, 23(3), 903–934. <https://doi.org/10.15516/cje.v23i3.4018>
- Meissner, B., & Bogner, F. X. (2013). Towards Cognitive Load Theory as Guideline for Instructional Design in Science Education. *Worlds Journal of Education*, 3(2), 24–37. <https://doi.org/10.5430/wje.v3n2p24>
- Pahriah, P., Ratnaya, I. G., Parwata, I., & Budhyani, I. (2025). Authentic Assessment in Chemistry Education at Senior High Schools and Higher Education: A Systematic Literature Review (2015–2025). *Path of Science*, 11(7), 1013–1026. <https://doi.org/10.22178/pos.120-10>
- Panadero, E., Delgado, P., Zamorano, D., Pinedo, L., & Fern, A. (2025). Putting excellence first: How rubric performance level order and feedback type influence students reading patterns and task performance. *Learning and Instruction*, 99(May). <https://doi.org/10.1016/j.learninstruc.2025.102168>
- Putri, S. R., & Hasruddin. (2025). Analysis of Science Literacy Abilities Through Authentic Assessment of High School Students. *Bioilmi: Jurnal Pendidikan*, 11(2), 81–92. <https://doi.org/10.19109/bioilmi.v11i2.30138>
- Seery, M. K., Agustian, H. Y., Doidge, E. D., Kucharski, M. M., Connor, H. M. O., & Price, A. (2017). Developing laboratory skills by incorporating peer-review and digital badges. *Chemistry Education Research and Practice*, 18(3), 403–419. <https://doi.org/10.1039/c7rp00003k>
- Shofiyah, N., Suprpto, N., & Jatmiko, B. (2026). A Data-Driven Approach to Improving Scientific Reasoning: The Development of the Data-Based Learning Cycle Model. *European Journal of Educational Research*, 15(2), 357–373. <https://doi.org/10.12973/eu-jer.15.2.357>
- Vasileiadou, D., & Karadimitriou, K. (2021). Examining the impact of self-assessment with the use of rubrics on primary school students performance. *International Journal of Educational Research Open*, 2(January). <https://doi.org/10.1016/j.ijedro.2021.100031>
- Veale, C. G. L., Jeena, V., & Sithebe, S. (2020). Prioritizing the Development of Experimental Skills and Scientific Reasoning: A Model for Authentic Evaluation of Laboratory Performance in Large Organic Chemistry Classes. *Journal of Chemical Education*, 97(3), 675–680. <https://doi.org/10.1021/acs.jchemed.9b00703>
- Yusof, W. R. W., Kimi, M., Sharifatun, W., Wan, H., Alha, M., Shahabudin, M., Al-ha, A., Mona, S., Aziz, A., Nur, I., Jay, D., & Gopal, R. (2021). Assessment of Practical and Scientific Writing Skills for Pre-University Students through Project-Based Learning. *Journal of Chemical Education*, 99(2), 715–722. <https://doi.org/10.1021/acs.jchemed.1c00843>
- Zhang, H., & Wink, D. J. (2021). Examining an Acid-Base Laboratory Practical Assessment from the Perspective of Evidence-Centered Design. *Journal of Chemical Education*, 98(6), 1898–1909. <https://doi.org/10.1021/acs.jchemed.0c01405>