

The Development and Application of Rubrics in Education: A Scopus-Based Systematic Literature Review (2016–2026)

Ahmad Hanafi Nurmawansyah ^{1*} , Aji Joko Budi Pramono ¹ 

¹ UIN Raden Mas Said Surakarta, Indonesia

* Author Correspondence

Article History

Received: 29 May 2026;

Revised : 10 June 2026;

Accepted: 14 July 2026.

Keywords

Rubric;
Educational Assessment;
Systematic Literature Review;
Evaluation;
Scopus.



Abstract

Rubric is an assessment instrument that plays a central role in supporting transparency and consistency of evaluation in education. This study aims to map the development, application, and effectiveness of rubrics in education through the Systematic Literature Review (SLR) approach following PRISMA 2020 guidelines. A literature search was conducted using keywords ("rubric" AND "education") in the Scopus database selected for its broad coverage of high-quality international journals and its concentration of Q1 and Q2 publications in education yielding 186 initial records. After selection stages based on the 2016–2026 range, Q1–Q4 journal rankings, abstract availability, and full-text feasibility screening using the Watase Uake automated SLR tool, a total of 13 articles were selected as the corpus of analysis. Given the single-database approach, these findings should be understood as representative of high-impact Scopus-indexed literature rather than an exhaustive mapping of the entire rubric research field. The results show that the development and use of rubrics in education has increased significantly in the last decade, with diverse focuses including: (1) competency-based formative assessment; (2) self-assessment and self-directed learning; (3) integration of artificial intelligence in automated assessment; and (4) cross-disciplinary competency development covering the fields of STEAM, bioinformatics, and language education. Methodologically, the research analyzed was dominated by qualitative approaches and instrument development, followed by quantitative and mixed methods, with most research conducted in higher education contexts. These findings imply the need for more systematic standardization in rubric development, strengthening educators' capacity to use rubrics as learning tools, and further exploration of the role of technology in educational evaluation.

Contact : Corresponding author  e-mail: ahmad25.hanafi@gmail.com

How to Cite : Nurmawansyah, A. H., & Pramono, A. J. B. The Development and Application of Rubrics in Education: A Scopus-Based Systematic Literature Review (2016–2026). *Pedagogi : Jurnal Pendidikan Dan Pembelajaran*, 6(2), 66–81. <https://doi.org/10.56393/pedagogi.v6i2.4440>



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). Allows readers to read, download, copy, distribute, print, search, or link to the full texts of its articles and allow readers to use them for any other lawful purpose. The journal hold the copyright.

Introduction

Rubric is one of the most essential assessment instruments in the modern education assessment ecosystem due to its ability to provide clear, structured, and standardized guidance in evaluating student performance. In general, rubrics are defined as a set of interrelated criteria and standards used to assess or evaluate a specific task or learning outcome (Brookhart, 2013). With rubrics, both educators and students share a common understanding of learning outcome expectations, making the assessment process more transparent, fair, and consistent. Recent research confirms that rubrics not only improve the reliability of assessments but also significantly contribute to improving the quality of feedback that learners receive (Panadero & Jonsson, 2020). In the context of competency-based education, rubrics serve as a bridge between established learning objectives and empirically observable evidence of student achievement, ensuring that evaluations are both valid and purposeful (Andrade, 2019). Furthermore, rubrics facilitate effective communication among educators, learners, and other stakeholders about the expected quality standards in any given assignment or project. Thus, rubrics have developed into an integral component of education quality assurance systems that prioritize meaningful, measurable, and equitable learning outcomes.

In its development, the use of rubrics has undergone a significant transformation, shifting from a tool primarily employed for summative grading to a holistic instrument embedded throughout the formative learning process. Rubrics are now integrated into the teaching-learning cycle as ongoing guides that help students monitor and evaluate their own progress against clearly defined standards. The Assessment for Learning (AfL) paradigm developed by Black and Wiliam (1998) explicitly positions rubrics as a key instrument in creating responsive, student-centered learning environments that prioritize continuous feedback. Formative rubrics support the development of self-assessment and peer assessment skills, which in turn strengthen students' metacognitive capacity and learning independence (Andrade, 2019). Recent international studies from contexts in the Americas, Europe, and Asia confirm that the integrated use of rubrics in the learning cycle results in consistent increases in student motivation and active involvement in their own learning (Karaman, 2024; Phuong et al., 2025). Moreover, rubrics align with authentic assessment objectives by providing a framework that allows learners to relate their activities to real-world competency standards in a tangible and meaningful way. The evolution of this rubric function reflects a broader paradigmatic shift in education from models of knowledge transmission toward collaborative models of shared knowledge construction between educators and learners.

The rapid development of digital education technology has opened new and significant dimensions in the development and application of rubrics across diverse learning environments. The integration of rubrics with technology-based platforms allows for faster, more consistent, and more accessible feedback for learners at any time and in any location (Radovic & Seidel, 2025). Advances in artificial intelligence (AI) and natural language processing have further enabled the development of automated scoring systems that use rubrics as an algorithmic basis to generate assessments that are both explainable and transparent to learners and educators alike (Kaldaras et al., 2022; Xu et al., 2025). Within the Explainable AI (XAI) framework, rubrics serve as a transparency guide that bridges human assessment standards with machine-based

capabilities, making automated feedback more interpretable and educationally meaningful. This rubric-guided AI assessment approach has the potential to address human resource limitations in large-scale assessments, such as national examinations or high-enrollment classroom contexts. However, the integration of rubrics with AI technology also raises fundamental questions regarding assessment validity, algorithmic fairness, and the ethical implications of automating educational evaluation. These technological developments position rubrics at the strategic intersection between traditional assessment practices and the emerging future of technology-driven education.

Despite the growing adoption and recognition of rubrics in education, their development and implementation practices still face a number of significant and persistent challenges. First, striking inconsistencies in rubric design approaches across institutions and disciplines lead to wide variation in instrument quality, undermining the comparability of assessments (Reddy & Andrade, 2010). Second, inter-assessor reliability in rubric-based assessment is often inadequate when criteria are not sufficiently specific, clearly operationalized, or accompanied by calibration training for assessors. Third, the phenomenon of "rubric washing" wherein rubrics are employed as a mere formality rather than as a meaningful guide integrated into the learning process remains a common problem across diverse educational contexts (Phuong et al., 2025). Fourth, most rubrics currently available were developed within specific institutional or disciplinary contexts and cannot be directly generalized or transferred to other settings without substantial modification. Fifth, the professional capacity of educators to develop high-quality rubrics and deploy them effectively as pedagogical tools still requires significant strengthening through structured, ongoing professional development programs. Sixth, the persistent gap between the demonstrated pedagogical potential of rubrics and the realities of their implementation in the field underscores the urgent need for more systematic and evidence-based research into the factors that mediate successful rubric adoption.

Although the rubric literature has expanded considerably over the last two decades, significant methodological gaps remain in the way existing research has been synthesized and disseminated. The majority of available literature reviews on rubrics are narrative in nature and do not employ rigorous, transparently documented search protocols, leaving them vulnerable to selection bias and interpretive subjectivity (Gough et al., 2017). Most existing reviews are also narrow in scope, often limited to a single discipline or educational level, without accounting for the broader cross-contextual patterns that characterize rubric use globally. These limitations make it difficult for education decision-makers including curriculum designers and assessment policy developers to obtain a comprehensive, evidence-based picture of rubric development and implementation across contexts. A systematic review that employs standardized academic databases and strict inclusion criteria would provide a far more objective and replicable account of the rubric research landscape than narrative approaches can offer. Furthermore, the rapid integration of AI and educational technology into assessment practices has introduced entirely new dimensions of rubric use that existing reviews have not yet adequately addressed. Therefore, a systematic literature review using strict protocols, broad analytical scope, and up-to-date coverage is urgently needed to fill this gap and provide an actionable evidence base for educators, researchers, and policymakers.

Research on rubrics carries substantial theoretical and practical significance for a wide range of education stakeholders across all levels of the system. Theoretically, a deeper and more systematic understanding of how rubrics function across different learning contexts can enrich educational assessment theory, particularly in relation to the constructs of validity, reliability, and the consequential basis of evaluation instruments (Messick, 1995). From a practical standpoint, the findings of rubric research provide an empirical foundation for educators seeking to design assessment instruments that are more effective, equitable, and responsive to diverse learner needs. For curriculum developers, understanding best practices in rubric development enables the design of educational programs that are more coherent, competency-aligned, and standardized across institutional settings. For policymakers, a comprehensive synthesis of rubric research offers a solid empirical foundation for crafting assessment policies that support systemic quality improvement in education (Popham, 2011). In the context of AI's rapid development, rubric research has become increasingly strategic, as rubrics now serve as the interface between human grading standards and machine-based automated scoring systems. Investment in rubric research and development therefore contributes directly to improving the fairness, transparency, and overall quality of educational assessment at both institutional and systemic levels.

Based on the background, identification of research gaps, and significance described above, this study aims to conduct a comprehensive systematic literature review on the development and application of rubrics in education using the PRISMA 2020 protocol, with Scopus as the primary database. Three specific objectives guide this review: (1) mapping the trend of research development regarding rubrics in education from 2016 to 2026 based on Scopus-indexed literature; (2) identifying the thematic focus and context of rubric application across various levels and fields of education; and (3) analyzing the diversity of methodological approaches employed in previous studies related to the development and evaluation of rubrics. This study contributes to the field by providing a rigorously systematic and up-to-date synthesis of high-impact rubric research, going beyond the scope and methodological limitations of existing narrative reviews. The findings are expected to be particularly useful for educators, researchers, curriculum designers, and assessment policy developers who seek a solid evidence base for rubric development and implementation practices. By explicitly situating the review within the current landscape of AI integration and educational technology, this study also opens a forward-looking research agenda at the intersection of rubrics, assessment validity, and algorithmic fairness. In addition, the review is expected to inform future longitudinal studies that can track the long-term impact of rubric use on learner competency development and educational equity. Ultimately, this review aims to advance the field toward a more standardized, evidence-informed, and technologically responsive approach to rubric use in education.

Method

This study uses the Systematic Literature Review (SLR) approach following the PRISMA 2020 guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analysis). This approach was chosen because it allows systematic and structured identification, selection, and

synthesis of literature, resulting in scientifically accountable findings. Every stage of the SLR process from the formulation of research questions to the search for literature to the extraction and analysis of data was carried out following a pre-designed protocol. The literature search was carried out systematically using the Scopus academic database, which is one of the largest and most comprehensive scientific databases in the world. The selection of Scopus was deliberately based on its concentration of high-impact publications: Scopus consistently indexes a higher proportion of journals ranked Q1 and Q2 by Scimago Journal Ranking compared to multidisciplinary repositories, making it particularly suitable for a review oriented toward identifying rigorous, quality-assured contributions to rubric research. The keywords used were ("*rubric*" AND "*education*"), designed to capture all articles discussing the use of rubrics in the context of education in general. It is acknowledged that restricting the search to a single database represents a methodological limitation: studies indexed exclusively in databases such as ERIC, Web of Science, or PsycINFO may not be captured. Accordingly, the findings of this review should be interpreted as representative of high-impact Scopus-indexed rubric research rather than an exhaustive synthesis of the entire field. Future studies are encouraged to replicate and extend this review using multiple databases to enhance comprehensiveness.

Results and Discussion

Results

Based on the results of a systematic search on the Scopus database using this combination of keywords yielded a total of 186 initial records. The distribution of publications shows a significant upward trend throughout the search period, with a sharp spike in 2025 reaching 28 articles, followed by 2021 with 18 articles and 2024 with 19 articles, indicating that research interest in rubric topics in education has continued to grow in recent years.

Result from Keyword Search

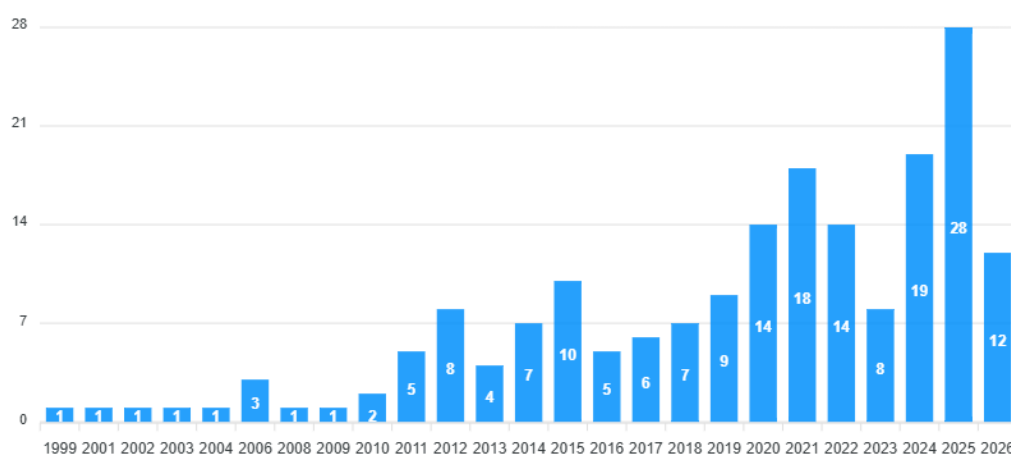


Figure 1. Distribution of Publications of Keyword Search Results by Year (1999–2026)

Source: Scopus search results using keywords ("*rubric*" AND "*education*")

To ensure the relevance and quality of the articles analyzed, this study established clear inclusion and exclusion criteria. Inclusion criteria: (1) articles published within the 2016–2026 time frame; (2) articles indexed in journals with Q1, Q2, Q3, or Q4 ratings based on the Scimago

Journal Ranking; (3) articles that have abstracts; and (4) articles that substantially discuss the development, application, or evaluation of rubrics in the context of education. Exclusion criteria: (1) duplicate articles; (2) irrelevant articles based on the automated system's assessment; (3) articles without available abstracts; and (4) articles not meeting the required journal-level category. The literature selection process was carried out through four main stages in accordance with the PRISMA 2020 flowchart: identification, screening, eligibility, and inclusion. The selection and management of records were supported by Watase Uake Tools, an automated software platform for SLR management that assists researchers in systematic deduplication, relevance filtering, and data extraction. The use of this tool introduces a degree of automation in the screening stage particularly in identifying irrelevant records which may differ from fully manual screening. This is acknowledged as a feature of the methodology that future reviewers should be aware of when replicating the study. At the identification stage, a search using keywords ("rubric" AND "education") in the Scopus database yielded 186 records. Before entering the screening stage, some records were eliminated: no duplicate records were found (n=0); 46 records were identified as irrelevant by the automated system based on the 2016–2026 range filter; 17 records were excluded because they did not meet the journal-level criteria (Q1–Q4); and 1 record was excluded because it did not have an abstract. The total number of records entering the screening stage was 122. At the screening stage, out of 122 records screened, 51 were excluded based on abstract review, leaving 71 reports submitted for full document retrieval. However, 58 reports could not have their full text accessed, so only 13 reports were successfully accessed and assessed for eligibility. At the eligibility level, all 13 reports were comprehensively assessed based on the inclusion and exclusion criteria. No reports were excluded at this stage, so all 13 reports were declared eligible for inclusion. This study thus analyzed 13 scientific articles as the final data corpus.

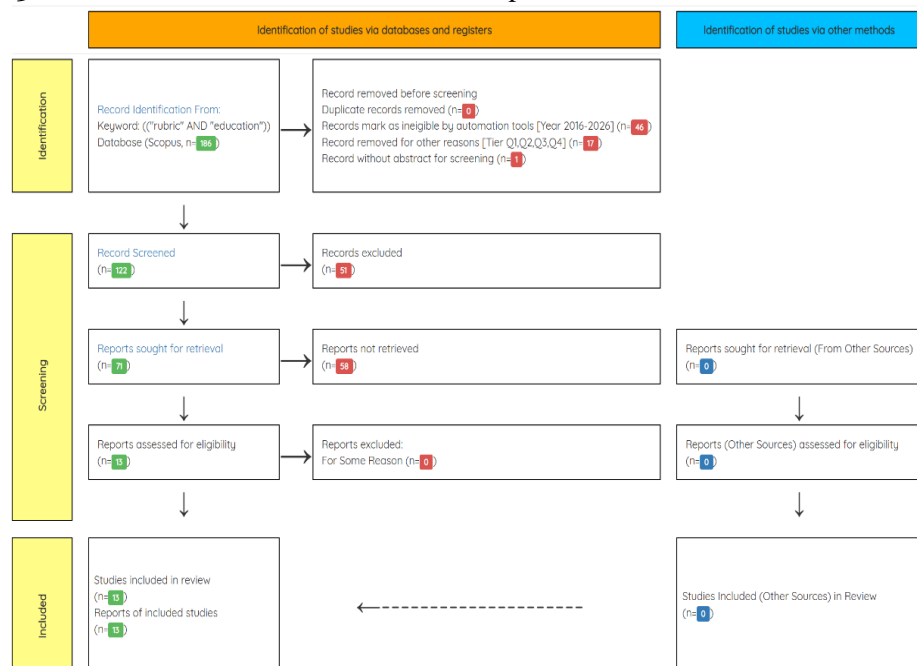


Figure 2. PRISMA Flowchart 2020 – Literature Selection Proses

Source: Adapted from PRISMA 2020 Reporting, processed using Watase Uake Tools

The data extraction process was carried out systematically using an extraction form that included information about the author, year of publication, article title, journal name, journal ranking (Q1–Q4), and the main research focus. The data were extracted independently and then verified to ensure accuracy and completeness. The analysis was conducted descriptively and thematically to identify the dominant research patterns, trends, and focuses in the literature analyzed. Table 1 below presents a complete list of articles included in this study along with their bibliographic information.

Table 1. List of Articles Included in the Systematic Literature Review

No	Author	Year	Title	Journal	Tier
1	Radovic, S.; Seidel, N.	2025	Introduction to the SRL-S Rubric for Evaluation of Innovative Higher Educational Technology for Self-Regulated Learning	Innovative Higher Education	Q2
2	Brufau Alvira, N.; Bannister, P.; Santamara Urbieto, A.	2025	Validating the PANDORA GenAI Susceptibility Rubric for Higher Education Assessment: A Field Test of All Translation and Interpreting BA Assignments	Higher Education Quarterly	Q1
3	Xu, W.; Kassim, M. S. S.; Hoo, W. L.; Yang, W.; Xu, T.	2025	Explainable AI for education: Enhancing essay scoring via rubric-aligned chain-of-thought prompting	International Journal of Modern Physics C	Q3
4	Phuong, H. Y.; Le, T. T.; Pham, T. T.	2025	Assessment for learning or assessment for scoring: Washback of analytic rubrics in Vietnamese EFL classrooms	Language Testing in Asia	Q1
5	Karaman, P.	2024	Effects of using rubrics in self-assessment with instructor feedback on pre-service teachers academic performance, self-regulated learning and perceptions of self-assessment	European Journal of Psychology of Education	Q1
6	Mang, H. M. A.; Chu, H. E.; Martin, S. N.; Kim, C. J.	2024	Correction: Developing an Evaluation Rubric for Planning and Assessing SSI-Based STEAM Programs in Science Classrooms	Research in Science Education	Q1
7	Kaldaras, L.; Yoshida, N. R.; Haudek, K. C.	2022	Rubric development for AI-enabled scoring of three-dimensional constructed-response assessment aligned to NGSS learning progression	Frontiers in Education	Q2
8	Freitas, D. de; Calafell, G.; Pierson, A. H. C.	2022	A Rubric to Evaluate Critical Science Education for Sustainability	Sustainability	Q1
9	Nadolski, R. J.; Hummel, H. G. K.; Rusman, E.; Ackermans, K.	2021	Rubric formats for the formative assessment of oral presentation skills acquisition in secondary education	Educational Technology Research and Development	Q1
10	Tractenberg, R. E.	2021	The Assessment Evaluation Rubric: Promoting Learning and Learner-Centered Teaching through Assessment in Face-to-Face or Distanced Higher Education	Education Sciences	Q2
11	Habig, B.	2020	Practical Rubrics for Informal Science Education Studies: (1) a STEM Research Design Rubric for Assessing Study Design and a (2) STEM Impact Rubric for Measuring Evidence of Impact	Frontiers in Education	Q2
12	Scott, R.; García-Aracil, A.	2020	Assessing Master Students Competencies Using Rubrics: Lessons Learned from Future Secondary Education Teachers	Sustainability	Q1
13	Tractenberg, R. E.; Lindvall, J. M.; Attwood, T. K.; Via, A.	2019	The Mastery Rubric for Bioinformatics: A tool to support design and evaluation of career-spanning education and training	PLOS ONE	Q1

Source: Data extraction results from Scopus, processed using Watase Uake Tools (2026)

After passing the stages of identification, screening, eligibility, and inclusion, a total of 13 articles were designated as the final analysis corpus. The temporal distribution of these articles reflects a consistent growth trend in rubric research in education, particularly in the period 2019 to 2025. In terms of journal quality, the majority of articles (eight out of 13) were indexed in Q1-ranked journals, four in Q2, and one in Q3. The concentration of Q1 and Q2 articles reflects a deliberate design feature of this review: the selection of Scopus as the source database ensures that the literature analyzed represents high-impact, internationally peer-reviewed work. While this enhances the rigor of the evidence base, it also means that findings should not be generalized as a complete picture of all rubric research contributions, published in lower-indexed or non-Scopus databases are not represented. The distribution of countries of origin is diverse, including the United States, Spain, Vietnam, Germany, Turkey, the Netherlands, Brazil, South Korea, Australia, Malaysia, and China, demonstrating that rubric use in education is a global phenomenon studied from various cultural perspectives and education systems.

Methodologically, the 13 articles analyzed consisted of six qualitative studies or instrument development, four quantitative studies, and three studies with a mixed-method approach. This diversity of methodologies reflects the complexity of rubric studies in education that cannot be approached through just one research paradigm. The findings of the 13 articles are synthesized into four main themes that emerged inductively from the thematic analysis process: (1) the development of rubrics as an assessment instrument; (2) rubrics in formative assessment and competency development; (3) rubrics in self-assessment and self-directed learning; and (4) integration of rubrics with artificial intelligence technology. Several studies in this corpus focus on the development of rubrics as evaluation instruments for specific educational programs and curricula. Tractenberg et al. (2019) introduced the Mastery Rubric for Bioinformatics (MR-Bi), an innovative tool designed to support the design and evaluation of cross-career bioinformatics education and training. MR-Bi was developed through an intensive two-year collaborative process involving an international workshop in Stockholm and weekly online meetings with bioinformatics experts from the United States, Sweden, and the United Kingdom. This rubric integrates 12 Knowledge, Skills, and Abilities (KSA) based on scientific methods with explicit developmental trajectories, allowing the mapping of student competencies from the level of new students to independent practitioners. The advantage of MR-Bi lies in its approach based on Bloom's Taxonomy and Knowles' Andragogy, ensuring alignment between cognitive level and learning readiness at each stage of competency development.

In line with this approach, Tractenberg (2021) developed an Assessment Evaluation Rubric (AER) specifically designed to evaluate the quality of assessment in higher education, both in face-to-face and distance learning modalities. The AER integrates four comprehensive assessment dimensions: Alignment (with learning objectives), Utility (usefulness for instructional decision-making), Complexity (cognitive complexity based on Bloom's Taxonomy), and Clarity (of instruction and assessment expectations). The strong theoretical foundations of the AER, including Messick's validity theory, Wiggins and McTighe's Backward Design, and Degree of Freedom Analysis, make it an instrument that can be used across

disciplines and instructional contexts. Although the AER had not been empirically validated at the time of publication, the study provides a strong conceptual foundation for the development of a more standardized assessment evaluation system in higher education. Mang et al. (2024) introduced an evaluation rubric specifically designed to plan and assess STEAM programs based on Socioscientific Issues (SSI) in science classes. Developed in South Korea and Australia, this rubric fills a significant gap in the literature as no previously available evaluation instruments specifically accommodated the integration of socio-scientific issues in STEAM learning. The SSI-STEAM rubric builds on the SSI Framework that emphasizes ethical, social, and scientific dimensions in an integrated manner, allowing educators to assess STEAM program quality not only in terms of scientific content but also social relevance and students' critical thinking development.

Freitas et al. (2022) developed the FACTS rubric using the Delphi method to assess the quality of science activities in the context of Critical Science Education for Sustainability (CSES). The development process involved three rounds of questionnaires with 37 international experts from Latin America and the Iberian Peninsula, guided by the STSE (Science-Technology-Society-Environment) perspective and the Morin Complexity Paradigm. The study confirms that valid rubric development requires an iterative consensus process involving multiple stakeholders. Scott (2020) offers an interesting perspective through the development of two practical rubrics for Informal Science Education (ISE) research: the STEM Research Design Rubric and the STEM Impact Rubric. These two rubrics were developed through focus groups with ISE researchers, practitioners, and evaluators, then piloted on 25 existing ISE studies. The inter-assessor reliability test showed satisfactory Cohen's Kappa scores of 0.89 for the research design rubric and 0.74 for the impact rubric, indicating a good level of assessment consistency. The main contribution of Scott's (2020) research lies in offering user-friendly and practical tools for ISE researchers who previously lacked standardized instruments to assess research quality outside formal education contexts. Overall, the findings on the rubric development theme show that a quality rubric development process requires at least three main components: (1) a strong theoretical foundation relevant to the context of application; (2) the involvement of experts and stakeholders in the validation process; and (3) systematic reliability testing. There is a clear tendency that rubrics developed specifically for a context or discipline tend to be more effective than generic rubrics.

Nadolski et al. (2021) conducted a comprehensive experimental study to compare the effectiveness of three rubric formats in the formative assessment of oral presentation skills in Dutch secondary education. The study involved 158 students divided into three conditions: a group using a text-based analytic rubric, a video-enhanced analytic rubric (via the Viewbrics tool), and a control group using a conventional rubric. Analysis using a mixed ANOVA repeated measures design with Bonferroni post-hoc yielded significant findings: the first hypothesis (H₁) was accepted as students using the analytic rubric significantly outperformed the control group ($p < 0.01$). However, the second hypothesis (H₂) was rejected because no significant differences were found between the text-based and video-enhanced rubric groups. This H₂ rejection has important theoretical implications. Although the video-enhanced rubric intuitively appears richer because it includes visual performance examples, the addition of multimedia elements

does not automatically increase rubric effectiveness as a formative assessment tool. This is consistent with Mayer's multimedia learning theory, which holds that format complexity is not always directly proportional to learning effectiveness. More important than presentation format is the clarity of criteria, the relevance of indicators to the given task, and how the rubric is integrated into the learning process. Isusi-Fagoaga and García-Aracil (2020) examined the use of rubrics as a competency-based formative assessment tool in master's education in Spain, particularly in music teacher education programs. Using an action research design with 74 master's students, this study found that consistent rubric use in the assessment process resulted in a significant increase in students' perception of mastery of transversal and specific competencies. These findings are relevant in the context of the Bologna Process and SDG Goal 4, which emphasize competency-based quality education. However, the study also identified an important limitation: rubrics alone are not sufficient to increase students' learning awareness, and need to be combined with other complementary assessment methods.

A very interesting perspective is offered by Phuong et al. (2025) through a qualitative study on the influence of analytical rubric washback on teaching and assessment practices of EFL (English as a Foreign Language) lecturers in Vietnam. Using semi-structured interviews with nine EFL lecturers from two higher education institutions, the study applied Braun and Clarke's Thematic Analysis to uncover patterns. The main finding is that the washback effect of analytical rubrics is highly variable and strongly influenced by the lecturer's career stage. Lecturers at the beginning of their careers tend to interpret rubrics as mechanistic assessment tools, while those in the middle and end of their careers are better able to use rubrics as tools to encourage learning. Karaman (2024) conducted a comprehensive quasi-experimental study to examine the influence of rubric use in self-assessment combined with instructor feedback on academic achievement, self-directed learning (SRL), and perceptions of prospective teachers in Turkey. This study involved 79 prospective teacher students divided into an experimental group ($n=44$, using a self-assessment rubric with 10 criteria and 4 levels) and a control group ($n=35$, without self-assessment). The results of ANCOVA and MANCOVA analyses showed that the experimental group significantly outperformed the control group both in academic achievement ($F = 4.36$; $p < 0.05$) and in overall self-directed learning ($F = 3.75$; $p = 0.015$). Qualitative analysis of students' reflective journals using the Grounded Theory approach further revealed that the majority of participants found the rubric very useful for goal setting, obtaining meaningful feedback, and conducting self-reflection.

Radović and Seidel (2025) offer a different perspective by developing the SRL-S Rubric, an innovative instrument designed not to assess learners' performance, but to evaluate the extent to which technology-based learning environments support self-directed learning. The SRL-S rubric consists of 14 items that operationalize three SRL phases from the Zimmerman (2000) model: the Forward Thinking phase (planning and goal setting), the Performance phase (monitoring and control), and the Self-Reflection phase (evaluation and adaptation). The instrument was validated through three systematic phases: initial development based on literature synthesis, validation by four SRL experts, and inter-assessor reliability testing (Intraclass Correlation Coefficient/ICC) with four faculty members.

Overall, research with the theme of self-assessment and SRL emphasized that rubrics have great potential as catalysts for the development of student learning independence. When rubrics are used not only as external assessment tools but also as guides for self-assessment, they can encourage the development of metacognitive skills that are the foundation of self-directed learning. One of the most significant developments recorded in the literature analyzed is the integration of rubrics with AI technology for automated assessment. Kaldaras et al. (2022) conducted pioneering research in the development of analytical rubrics specifically designed to facilitate AI-based (machine learning) assessments of three-dimensional constructive responses aligned with the Next Generation Science Standards (NGSS). This study involved 1,252 written responses of grade 9 students used to train and validate machine learning models using the Constructed Response Classifier (CRC). The results of inter-rater reliability testing showed very high Cohen's Kappa values (0.82–0.97), and the comparison between holistic and analytical rubrics reached a 99% agreement rate. A properly structured analytical rubric can thus serve as a bridge between human assessment and machine assessment.

Xu et al. (2025) went a step further by developing an Automated Essay Scoring (AES) system based on Large Language Models (LLMs) that explicitly integrate IELTS rubrics through a Chain-of-Thought (CoT) prompting mechanism reinforced with Reinforcement Learning from Human Feedback (RLHF). The study used a dataset of 5,088 IELTS essays from real candidates annotated by two certified examiners on four traits: Task Response (TR), Coherence and Cohesion (CC), Lexical Resources (LR), and Grammatical Range and Accuracy (GRA). The system showed highly competitive results with Quadratic Weighted Kappa (QWK) values of 0.6814, BLEU 0.2873, ROUGE-L 0.4588, and Cosine Similarity 0.6856. This approach increases assessment transparency because the model not only generates scores but also generates reasoning that explicitly reflects rubric criteria fundamentally different from conventional AES systems, which are black-box in nature. Brufau et al. (2025) introduced the PANDORA GenAI Vulnerability Rubric, an innovative instrument developed in Spain to assess the vulnerability of assessments to generative AI manipulation. This study analyzed 151 assessments from 29 lecturers of the Translation and Interpretation study program. The study's key findings reveal a significant gap between institutional guidance on AI and practical frameworks that lecturers can use to evaluate and modify their assessments to be more resilient to AI manipulation. Overall, the theme of rubric integration with AI reflects a significant paradigmatic shift: rubrics are no longer seen solely as instruments used by human educators, but have evolved into a fundamental component in a technology-based assessment ecosystem. This development opens vast research opportunities while presenting new ethical and methodological challenges.

Discussion

The findings of this systematic review, interpreted through the lens of relevant educational assessment theories, reveal significant conceptual shifts in the way rubrics are understood and used. From the perspective of the formative assessment theory of Black and Wiliam (1998), the majority of studies analyzed confirm that rubrics are most effective when they function as interactive learning tools, not just final evaluation instruments. Nadolski et al.'s (2021) findings on the superiority of analytical rubrics in formative assessment and Phuong et

al.'s (2025) findings on the rubric-based washback effect consistently support the Afl's central proposition that effective feedback should be specific, criterion-referential, and integrated into the learning cycle. Within the framework of Wiggins' (1998) authentic assessment theory, the findings also show that the most successful rubrics are those developed in the context of tasks that reflect real-world challenges, such as the SSI-based STEAM rubric by Mang et al. (2024) and the MR-Bi rubric for bioinformatics by Tractenberg et al. (2019). This implies that rubric development is inseparable from the design of meaningful authentic assessment tasks. The analysis of methodological diversity from the 13 articles reveals patterns reflecting the developmental stage of rubric research as a scientific field. Of the 13 articles included, there were six studies based on qualitative approaches or instrument development (Tractenberg et al., 2019; Tractenberg, 2021; Freitas et al., 2022; Phuong et al., 2025; Brufau et al., 2025; Radović & Seidel, 2025), four quantitative studies (Nadolski et al., 2021; Karaman, 2024; Xu et al., 2025; Mang et al., 2024), and three studies with a mixed approach (Scott, 2020; Isusi-Fagoaga & García-Aracil, 2020; Kaldaras et al., 2022). The dominance of qualitative approaches and instrument development reflects a research community still actively building the foundations of the instrument before moving on to large-scale effectiveness studies.

The absence of longitudinal research in this corpus is the most striking gap. The long-term impact of rubric use on the development of students' assessment competencies and literacy has never been systematically quantified within this set of Scopus-indexed literature. A longitudinal research agenda that can track the development of students' self-assessment skills and metacognition throughout education programs needs to be a top priority in future rubric research. A cross-thematic synthesis of the four study areas reveals several convergent patterns. First, rubrics were most effective when integrated into the overall teaching-learning cycle rather than used only at the end of the process. In the context of formative assessment, rubrics encourage meaningful feedback (Nadolski et al., 2021; Phuong et al., 2025). In the context of SRL, rubrics serve as metacognitive guides (Karaman, 2024; Radović & Seidel, 2025). In the context of AI, rubrics become a bridge between human standards and machine algorithms (Kaldaras et al., 2022; Xu et al., 2025). This cross-contextual unity of function suggests that rubrics have a high degree of conceptual flexibility while requiring careful adaptation to specific application goals. Second, the quality of rubric design significantly determines its effectiveness. All research reviewed confirms that rubrics with specific, well-operationalized criteria are more effective than those with ambiguous or generic descriptors. This is consistent with the principles of construct validity emphasized by Messick (1995), where the validity of an assessment instrument lies not only in its content but also in the accuracy of construct representation. Third, social, cultural, and institutional contexts play an important role in determining rubric effectiveness. The findings of Phuong et al. (2025) on the variation of washback effects by lecturer career stage, as well as Brufau et al.'s (2025) findings on the gap between institutional policies and practice in the field, remind us that rubrics do not operate in a vacuum. Successful rubric implementation requires a deep understanding of contextual dynamics, including the prevailing assessment culture, educators' professional capacity, and institutional support available. Fourth, there is a clear trajectory toward the integration of rubrics with increasingly sophisticated digital technology from simple digital tools such as Viewbrics (Nadolski et al.,

2021) to LLM-based AI systems that use rubrics as assessment guides (Xu et al., 2025; Kaldaras et al., 2022). The fundamental question that future research needs to answer is: to what extent can rubrics developed for human assessment be adapted without losing construct validity for AI-based assessments?

The integration of rubrics with artificial intelligence reflected in the research of Kaldaras et al. (2022) and Xu et al. (2025) brings profound implications for the future of educational assessment systems. From the perspective of assessment literacy theory (Popham, 2011), this development requires expanding the definition of educator assessment competencies to include an understanding of how AI-based assessment systems work and the ability to evaluate the validity and fairness of their outputs. Brufau et al.'s (2025) findings on the vulnerability of traditional assessments to generative AI manipulation add a critical new dimension: rubrics not only need to be adaptive to technology but also need to be designed to maintain academic integrity in an era of AI proliferation. From an education policy perspective, the results of this study imply the need to develop regulatory frameworks and implementation guidelines that assist educational institutions in integrating AI-based rubrics responsibly and ethically. The research in this corpus also revealed several research gaps that are important to address in the future. First, most research is still limited to the context of higher education and secondary education in developed countries, so generalization of findings to primary education and developing countries requires further investigation. Second, longitudinal research measuring the long-term impact of rubric use on the development of learning competencies and student independence is still very limited. Third, research on students' perceptions not only educators' about the effectiveness and experience of using different types of rubrics needs to be expanded so that the perspectives of the main beneficiaries of rubrics can be better represented in the literature.

Conclusion

This systematic review succeeded in identifying and analyzing 13 relevant scientific articles regarding the use of rubrics in education through a strict selection process based on the PRISMA 2020 protocol, using the Scopus database and supported by Watase Uake Tools for automated record management. Referring to the three objectives set, the following conclusions are drawn. First, with respect to the goal of mapping research development trends: rubric research in education has experienced significant growth in the last decade, with a clear surge in publications in the period 2021–2025. Most of the research was published in Q1 journals from different countries a reflection of the deliberate focus on high-impact Scopus-indexed literature marking rubrics as a topic that has received serious attention from the international scientific community. The latest trend is characterized by increasing study on the integration of rubrics with AI technology and the importance of maintaining academic integrity in the era of generative AI. Second, in relation to the purpose of identifying the focus and context of rubric implementation, four main focuses are identified: (a) the development of rubrics as cross-disciplinary assessment instruments including bioinformatics, STEAM, continuing science education, and informal education; (b) rubrics in formative assessment and transversal competency development in secondary and higher education; (c) rubrics as catalysts for self-

assessment and self-directed learning; and (d) rubric integration with AI technology for automated assessment and evaluation of academic integrity. The implementation context covers various levels of education from secondary to postgraduate, with the dominance of studies in the higher education context from American, European, and East Asian countries. Third, the studies analyzed were dominated by qualitative approaches and instrument development (six articles), followed by quantitative (four articles) and mixed methods (three articles). This pattern shows that the research community is still actively building and validating rubric instruments. The most striking gap is the absence of longitudinal research, indicating the urgent need for long-term studies to measure the ongoing impact of rubric use on learners' competency development. It is important to note that these findings are bounded by the scope of this review: a single database (Scopus), 13 articles derived from automated SLR tooling, and a concentration on high-ranking journals. Future studies are therefore strongly encouraged to (1) expand to multiple databases (e.g., ERIC, Web of Science, PsycINFO), (2) cover the context of primary education and developing countries, (3) further explore the effectiveness of AI-based rubrics, and (4) investigate the long-term impact of rubric use on learners' learning independence through longitudinal research designs.

Acknowledgments

The authors would like to express their sincere gratitude to Universitas Islam Negeri Raden Mas Said for providing the necessary research facilities throughout the course of this study. The institution's support, including dedicated workspace and stable internet access, greatly facilitated the data retrieval process during the writing of this article. The authors personally funded this research.

Authors' Note

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

References

- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, Article 87. <https://doi.org/10.3389/feduc.2019.00087>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Brufau Alvira, N., Bannister, P., & Santamaría Urbieta, A. (2025). Validating the PANDORA GenAI vulnerability rubric for higher education assessment: A field test of all BA translation and interpreting assignments. *Higher Education Quarterly*, 79(2), 145–163. <https://doi.org/10.1111/hequ.12500>
- Freitas, D. de, Calafell, G., & Pierson, A. H. C. (2022). Rubric to evaluate critical science education for sustainability. *Sustainability*, 14(3), Article 1697. <https://doi.org/10.3390/sui4031697>
- Gough, D., Oliver, S., & Thomas, J. (Eds.). (2017). *An introduction to systematic reviews* (2nd ed.). SAGE.

-
- Isusi-Fagoaga, R., & García-Aracil, A. (2020). Assessing master students' competencies using rubrics: Lessons learned from future secondary education teachers. *Sustainability*, 12(22), Article 9561. <https://doi.org/10.3390/su12229561>
- Kaldaras, L., Yoshida, N. R., & Haudek, K. C. (2022). Rubric development for AI-supported three-dimensional constructed response assessment aligned with NGSS learning progressions. *Frontiers in Education*, 7, Article 875678. <https://doi.org/10.3389/feduc.2022.875678>
- Karaman, P. (2024). Effects of rubric use in self-assessment with instructor feedback on pre-service teachers' academic performance, self-regulated learning and self-assessment perceptions. *European Journal of Educational Psychology*, 17(1), 23–41. <https://doi.org/10.1007/s10212-023-00720-6>
- Mang, H. M. A., Chu, H. E., Martin, S. N., & Kim, C. J. (2024). Developing an evaluation rubric for planning and assessing SSI-based STEAM programs in science classrooms. *Research in Science Education*, 54(1), 89–108. <https://doi.org/10.1007/s11165-023-10125-x>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Nadolski, R. J., Hummel, H. G. K., Rusman, E., & Ackermans, K. (2021). Rubric formats for formative assessment of oral presentation skill acquisition in secondary education. *Educational Technology Research and Development*, 69(4), 1965–1986. <https://doi.org/10.1007/s11423-021-10021-9>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30, Article 100329. <https://doi.org/10.1016/j.edurev.2020.100329>
- Phuong, H. Y., The, T. T., & Pham, T. T. (2025). Assessment for learning or assessment for grading: Washback of analytic rubrics in Vietnamese EFL classrooms. *Language Testing in Asia*, 15(1), Article 12. <https://doi.org/10.1186/s40468-025-00310-5>
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265–273. <https://doi.org/10.1080/08878730.2011.605048>
- Radovic, S., & Seidel, N. (2025). Introducing the SRL-S rubric for evaluating innovative higher educational technology for self-regulated learning. *Innovative Higher Education*, 50(1), 55–74. <https://doi.org/10.1007/s10755-024-09720-8>
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. <https://doi.org/10.1080/02602930902862859>
- Scott, B. (2020). Practical rubrics for informal science education studies: (1) STEM research design rubric for assessing study design and (2) STEM impact rubric for measuring evidence of impact. *Frontiers in Education*, 5, Article 54. <https://doi.org/10.3389/feduc.2020.00054>
- Tractenberg, R. E. (2021). Assessment evaluation rubric: Promoting learner-centered teaching and learning through assessment in face-to-face or distance higher education. *Education Sciences*, 11(8), Article 390. <https://doi.org/10.3390/educsci11080390>
-

-
- Tractenberg, R. E., Lindvall, J. M., Attwood, T. K., & Via, A. (2019). Guidelines for curriculum and course development in higher education and training. *PLOS ONE*, 14(3), Article e0212703. <https://doi.org/10.1371/journal.pone.0212703>
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.
- Xu, W., Kassim, M. S. S., Hoo, W. L., Yang, W., & Xu, T. (2025). Explainable AI for education: Improving essay scoring through rubric-aligned chain-of-thought prompting. *International Journal of Modern Physics C*, 36(4), Article 2550038. <https://doi.org/10.1142/S0129183125500385>