JPPIPA 12(8) (2026)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Assessment in STEM Education: A Bibliometric and Systematic Review of Methods, Innovations, and Interdisciplinary Approaches

Magfira Aulia[1*], Yusman Wiyatmo[1], Putri Rose Amanda[1], Bob Widodi, Heru Kuswanto[1]

[1] Master's Program in Physics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia.

Received: April 02, 2026
Revised: July 11, 2026
Accepted: August 25, 2026
Published: August 31, 2026

Corresponding Author:
Magfira Aulia
magfiraaulia.2024@student.uny.ac.id

DOI: 10.29303/jppipa.v12i8.14898

Abstract: STEM (Science, Technology, Engineering, and Mathematics) education plays an essential role in preparing students for technological advancement. However, selecting appropriate assessment methods remains challenging across educational levels. This study employed a systematic literature review (SLR) and bibliometric analysis to examine assessment methods, technological innovations, and interdisciplinary approaches in STEM education. A total of 98 articles were retrieved from the Scopus database. Following the identification, screening, eligibility, and inclusion stages based on predefined inclusion and exclusion criteria, 36 articles published between 2019 and 2024 were selected for analysis. Bibliometric analysis examined keyword relationships and research networks, while VOS viewer visualized the bibliometric maps. The findings revealed that summative assessment was the dominant assessment method across educational levels, whereas formative assessment was also widely used to provide continuous feedback. Bibliometric analysis identified STEM education as the largest research cluster, closely related to technology integration and interdisciplinary learning. In addition, 12 of the 36 reviewed studies reported technology-enhanced and interdisciplinary assessment innovations, including eye-tracking, Bayesian inference, virtual robotics, game-based learning, and large language models. These findings highlight the potential of technology-enhanced and interdisciplinary assessment to improve STEM learning across educational levels.

Keywords: Assessment; Interdisciplinary learning; STEM education; Technology integration

Introduction

STEM (science, technology, engineering, and mathematics) education is experiencing rapid development in many developed countries, prompting an urgent need to provide well-designed instructional contexts (Christopoulos et al., 2020). STEM education equips students with essential knowledge, attitudes, and problem-solving skills to address real-world challenges (Solihah et al., 2024). Student retention and success in STEM has become critical to meet the growing demand for labor. Therefore, the development and evaluation of scientific and analytical reasoning skills are indispensable to ensure the effectiveness of programs that prepare students for STEM professions (Lehmeidi Dong et al., 2023). In addition, previous studies have consistently reported that STEM-based learning can enhance students' creativity and collaboration skills, further emphasizing its role in developing essential 21st-century competencies (Gunawan et al., 2025). STEM-integrated learning activities have also been shown to enhance students' collaboration skills and learning outcomes (Rasmi et al., 2025).

Measuring learner-centered pedagogy has attracted the attention of many instructors and researchers who seek reliable and valid ways to describe teaching practices. However, two major challenges remain: (1)

How to Cite:
Aulia, M., Wiyatmo, Y., Amanda, P. R., Widodi, B., & Kuswanto, H. (2026). Assessment in STEM Education: A Bibliometric and Systematic Review of Methods, Innovations, and Interdisciplinary Approaches. *Jurnal Penelitian Pendidikan IPA*, *12*(8), 8–20. https://doi.org/10.29303/jppipa.v12i8.14898

methods for measuring learner-centeredness are often resource-intensive and difficult to implement, and (2) commonly used observation protocols are not fully aligned with learner-centered frameworks, as active learning is often equated with student-centeredness (Donham et al., 2024).

Numerous assessment instruments and conceptual inventories have been developed to diagnose students' misconceptions and improve instruction in science education. However, appropriate and efficient assessment tools for measuring students' understanding of engineering and technology concepts remain limited. Consequently, science teachers often rely on general rubrics to assess students' technological competencies during and after instruction. Therefore, understanding students' alternative conceptions is essential for designing effective instructional strategies that address engineering and technology concepts (Asghar et al., 2019). Recent studies have introduced project-based STEAM assessment as a more comprehensive approach to evaluating students' competencies (Alhayat et al., 2024). These limitations highlight the need for assessment methods that comprehensively evaluate all STEM domains rather than focusing primarily on science.

Integrated STEM activities have been shown to improve students' understanding of statistical concepts. However, maximizing the learning potential of these activities requires a deeper understanding of how students develop conceptual knowledge throughout the learning process (Watson et al., 2023). Furthermore, the integration of digital technologies into education has transformed conventional teaching practices by creating new opportunities to engage students and enrich learning experiences. Consequently, technology-enhanced learning has become an increasingly important component of STEM education. In line with these developments, STEM-based digital assessment has been shown to effectively measure higher-order thinking skills while promoting interactive and reflective learning (Ramadhani et al., 2025).

Despite these developments, significant gaps remain in STEM assessment. Although various assessment tools are available for science education, assessment methods specifically designed for engineering and technology concepts remain limited. Furthermore, systematic reviews have emphasized the importance of developing appropriate diagnostic instruments to identify students' understanding and misconceptions in science learning (Shiddiqi et al., 2024). Recent studies have examined the relationship between students' logical reasoning skills, scientific literacy, and science learning achievement, highlighting the relevance of assessing these competencies in science education (Saputra et al., 2025). In addition, digital technologies and interdisciplinary approaches have demonstrated considerable potential to improve teaching effectiveness and support more personalized learning. However, comprehensive evidence regarding how these innovations are integrated into STEM assessment remains limited (Khalid et al., 2024). These limitations highlight the need for a comprehensive synthesis of current assessment methods and emerging innovations in STEM education. To address this gap, this study combines a systematic literature review (SLR) and bibliometric analysis to provide a comprehensive overview of assessment methods, technological innovations, and interdisciplinary approaches in STEM education. This integrated approach aims to provide a comprehensive synthesis of STEM assessment methods, map research trends, and examine emerging technological and interdisciplinary innovations in STEM assessment. Therefore, this study is expected to provide comprehensive insights for researchers and educators in developing more effective STEM assessment practices and identifying future research directions.

This article aims to explore assessment methods and models in STEM learning, including recent innovations that integrate technology, to provide more comprehensive insights into addressing learning challenges in the digital age. The research questions are: What are the most frequently used assessment methods or models in STEM learning? What are the latest innovations in STEM learning assessment methods or models, and to what extent are they able to integrate technology or interdisciplinary approaches?

Method

Identification

The first step is to identify keywords and search for similar terms using dictionaries, encyclopedias, and previous research. The search string for the Scopus database was created after all relevant phrases were selected as shown in Table 1. In the initial stage of the systematic review procedure, the current research project successfully extracted 98 articles from the Scopus database.

Table 1. The Search Strings

Database	Search strategy	Publication period	Document type
Scopus	TITLE-ABS-KEY (assessment AND STEM AND science AND education AND students) AND PUBYEAR > 2018 AND PUBYEAR < 2025	2019–2024	Journal articles

Screening

There were 98 articles using various inclusion and exclusion criteria established by the researchers, which included excluding books, chapters, book series, reviews, proceedings and conferences. In addition, this review focused only on English-language papers. In addition, some terms were restricted to searching only for relevant articles. It is important to note that the schedule was selected considering a five-year timeframe (2019-2024).

Eligibility Criteria

Eligibility Criteria describe eligibility criteria (inclusion and exclusion) that clearly allow to set boundaries for a systematic literature review. These criteria aim to create a structured framework that facilitates the inclusion of studies that fulfill important prerequisites while excluding studies that do not fit our research objectives.

Table 2. Inclusion and Exclusion Criteria

Category	Inclusion Criteria	Exclusion Criteria
Type of publication	Articles published in journals	Articles published in proceeding conference, book, website or bog, and others
Journal specifications	International journals indexed by Scopus	International journals not indexed by Scopus
Publication year	2019-Oktober 2024	Before 2019
Language	English	Non-English
Field	Science Education	Outside of science
Topic	Assessment of STEM learning	Not include assessment in stem learning

Data Abstraction and Analysis

PRISMA-P Process

This study adopted a rigorous systematic review protocol that followed the 4-step PRISMA-P process. This process has the following steps: (1) Identification of the literature relevant to this study, (2) Screening using the criteria defined in table 2, (3) Classification of the methodically screened articles and (4) Determining the articles for inclusion in this review. At this stage, the literature identified from the Scopus database search was screened to exclude articles that were: (1) published before 2019, (2) written in a language other than English, (3) published but not yet subjected to a peer review process (e.g., conference papers, book chapters, letters), and (4) not relevant to this review. The relevance of individual articles was determined by examining the title, abstract, results, and methods. Any article that did not meet these screening criteria was excluded from this study. The relevance of individual articles was determined by examining the title, abstract, results, and methods. Any article that did not meet these screening criteria was excluded from this study. Regarding the prismatic shape identified in the article, the details are shown in Figure 1.

Figure 1 illustrates the article selection process conducted in this systematic literature review. The identification stage began with 98 articles identified through the Scopus database using the predetermined search strings. Based on the document type criteria, 17 documents were excluded, resulting in 81 records for further screening. In the screening stage, duplicate records and irrelevant articles based on title and abstract assessment were removed, leaving 50 articles for eligibility assessment. Subsequently, the eligibility assessment was conducted based on the predefined inclusion and exclusion criteria, resulting in the exclusion of 14 articles due to their lack of relevance. Finally, 36 articles were included for further analysis.

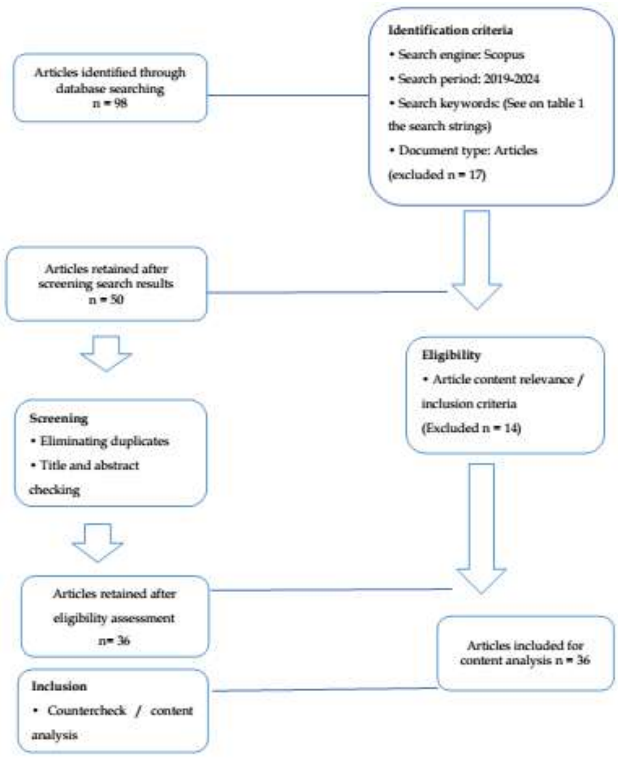

Figure 1. Article selection process

Bibliometric Analysis

Software, namely the Biblioshiny application and VOSviewer (Version 1.6.20), was applied for bibliometric analysis using Scopus database search results. The selected articles were then exported into BIB and RIS files. BIB and RIS files were used for bibliometric analysis. RIS files were loaded into Vosviewer to create a co-occurrence map of bibliographic data, and BIB files were loaded into the Biblioshiny application to collect, analyze, and visualize bibliometric data. Bibliometric analysis was performed on each article, including year of publication, keywords, number of journal publication citations, and country of publication.

Result and Discussion

Result

This systematic literature review examined articles published from 2019 - October 2024 with a focus on STEM learning assessment. Based on analysis using the bibliometrix application, it shows that the number of articles published each year has fluctuated, the data can be seen in the diagram below.

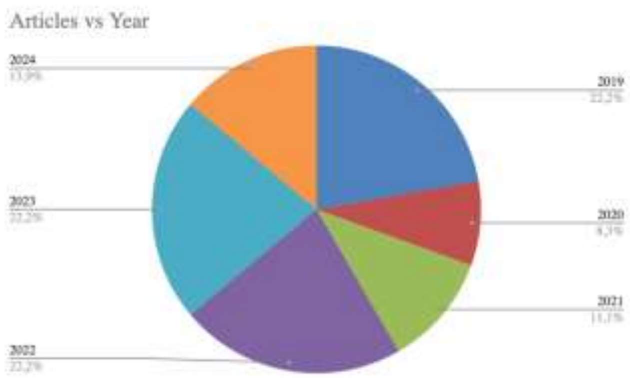

Figure 2. Number of article publications (2019-2024)

Based on the 36 articles that have been analyzed with the help of the Bibliometrix application, it is known that the journals that publish the most articles are Education Sciences and Frontiers in Education with a total of 3 articles each. The 3 articles published by Education Sciences are Exploring Secondary Students' Alternative Conceptions about Engineering Design Technology (Asghar et al., 2019), Assessing and Benchmarking Learning Outcomes of Robotics-Enabled STEM Education (Rahman, 2021), and Active Learning in STEM Education with Regard to the Development of Inquiry Skills (Ješková et al., 2022). Then 3 articles were published by Frontiers in Education yaitu Field science in the age of online learning: Dynamic instruction of techniques to assess soil physical properties (Hinckley & Fendorf, 2022), An application of Bayesian inference to examine student retention and attrition in the STEM classroom (Bertolini et al., 2023), and Linking information from multiple representations: an eye-tracking study (Susac et al., 2023). The graph can be seen in figure 2.

Table 3. Total Citations and Average Article Citations

Country	TC	Average Article Citations
Usa	151	10,10
Canada	55	27,50
Ireland	45	22,50
China	40	40,00
United Kingdom	24	24,00
Slovakia	16	16,00
Georgia	14	14,00
Malaysia	7	3,50
Australia	6	2,00
Croatia	4	4,00
New Zealand	1	1,00
Indonesia	0	0,00

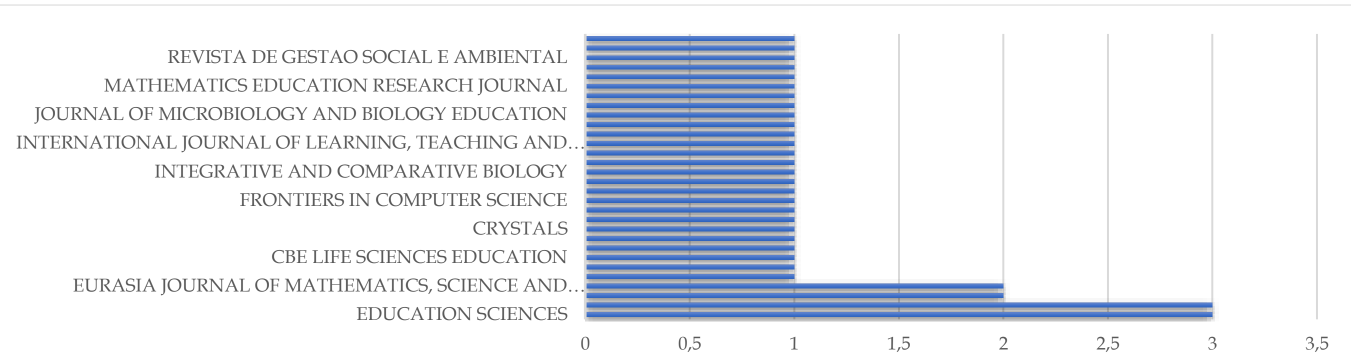

Figure 3. Journal and articles

Then in Table 3, we can observe the countries that have total citations and the average number of citations per article. The United States has the most total citations, 151 with an average of 10.10 per article. Of course this shows a large contribution to research but the distribution of quality is more even than countries like China. Countries like China, Canada, and Ireland have high article quality based on average citations per article. Meanwhile, countries with low citations (e.g. Malaysia, Indonesia, and Australia) may need to improve the quality or number of their publications to increase academic visibility globally.

This section discusses the most frequently used assessment methods or models in STEM learning at different levels of education. The analysis was conducted by reviewing the filtered articles, and the data is presented in tables based on education level (elementary, middle, high school, university, teacher and Gen Z Indonesia). The data is as follows:

Table 4. Assessment Models/Methods for University Level

Title	Year	Assessment Method/Model
Participation in voluntary re-quizzing is predictive of increased performance on cumulative assessments in introductory biology	2019	Formative assessment
Testing the test: Are exams measuring understanding?	2019	Formative assessment
A User Interface (UI) and User eXperience (UX) evaluation framework for cyberlearning environments in computer science and software engineering education	2020	Formative assessment
Evaluating STEM-based sustainability understanding: A cognitive mapping approach	2021	Summative assessment
Field science in the age of online learning: Dynamic instruction of techniques to assess soil physical properties	2022	Formative, authentic, project-based assessment
Game-Based Learning: Enhancing Student Experience, Knowledge Gain, and Usability in Higher Education Programming Courses	2022	Summative assessment
An application of Bayesian inference to examine student retention and attrition in the STEM classroom	2023	Data-driven assessment
Life Science Research Immersion Program Improves STEM-Specific Skills and Science Attitudes among Precollege Students	2023	Summative assessment
Learning and STEM identity gains from an online module on sequencing-based surveillance of antimicrobial resistance in the environment: An analysis of the PARE-Seq curriculum	2023	Summative assessment
Linking information from multiple representations: an eye-tracking study	2023	Summative assessment
If you are learner-centered and you know it, raise your hand: Perspectives on and implementation of pedagogical changes by science instructors during the COVID-19 pandemic	2024	Formative assessment
Constructively Aligned Instructional Design For Oral Presentations	2024	Formative assessment

This table presents data on the various assessment models or methods used in STEM learning at the University level. The analysis is based on articles published from 2019-2024 with a focus on assessment methods or models applied to support student learning outcomes. Formative assessment is the dominant assessment used as in evaluating comprehension through re-quizzes, user interface (UI/UX), and the use of tools such as eye-tracking to map learning representations. Formative assessment is suitable because it helps students to understand concepts deeply during the learning process. Summative assessment is more often used to measure overall student outcomes after the learning process, such as in the mastery of sustainability-based STEM materials or online modules for understanding antimicrobial resistance.

Table 5. Assessment Model/Method for Senior High School Level

Title	Year	Assessment Method/Model
Exploring secondary students, alternative conceptions about engineering design technology	2019	Formative assessment
Electronic Field Trips for Science Engagement: The Streaming Science Model	2019	Formative assessment
Question generating supported by blended learning platform: Issues of social justice for environmental education	2019	Project-based assessment and formative assessment
Assessing and benchmarking learning outcomes of robotics-enabled stem education	2021	Formative and summative assessment
Entropy of Co-Enrolment Networks Reveal Disparities in High School STEM Participation	2021	Formative assessment
Active Learning in STEM Education with Regard to the Development of Inquiry Skills	2022	Summative assessment
The Gamification of XFEL Education Using XFEL Crystal Blaster	2022	Formative assessment
Making Waves in the Classroom ,Engaging students in STEM through hands-on coastal oceanography and engineering	2022	Formative and summative assessment
Browser-based simulation for novice-friendly classroom robotics	2023	Summative assessment

Assessment methods at the high school level show a balance between process (formative) and outcome (summative) assessment. The combination with project-based approaches also reflects the need for high school students to understand practical applications of STEM and interdisciplinary skills. Most of the methods used at the high school level are formative assessment-based, such as in the exploration of students' alternative understanding of engineering design technologies, gamification-based learning, and browser-based robotics simulations.

Table 6. Assessment Model/Method for Junior High School Level

Title	Year	Assessment Method/Model
Authentic science with dissemination increases self-efficacy of middle school students	2019	Summative assessment
Development and validation of the middle grades computer science concept inventory (MG-CSCI) assessment	2020	Summative assessment
Examining high achievement in mathematics and science among post-primary students in Ireland: a multilevel binary logistic regression analysis of PISA data	2022	Summative assessment
Indonesian national assessment support: Can RE-STEM Android app improve students, scientific literacy skills?	2023	Summative assessment
Applying large language models and chain-of-thought for automatic scoring	2024	Technology-based assessment

The junior high school level presents assessment methods or models used in STEM learning. Most articles predominantly use summative assessment to measure the final or overall results of learning, such as increasing student self-efficacy through authentic scientific activities, Inventory of computer science concepts to evaluate understanding of technology materials and analysis of PISA results to identify students with high results or achievements in math and science. One of the articles discusses the use of RE_STEM-based Android applications to improve students' science literacy.

Table 7. Assessment Models/Methods for Primary School and Early Childhood Education Levels

Title	Year	Assessment Method/Model	Level
Building a Better Wall: Assessing Children, Design Technology Learning in Nature-Based Early Childhood Education	2024	authentic and formative assessment	Early Childhood Education
Exploring changes in primary students, attitudes towards science, technology, engineering and mathematics (STEM) across genders and grade levels	2019	summative assessment	Primary School
Building engineering awareness: Problem-based learning approach for STEM integration	2020	summative assessment	Primary School
Consolidating understanding of variation as part of STEM: experimenting with plant growth	2023	formative and summative assessment	Primary School

The assessment methods analyzed cover a wide range of assessment approaches that adapt to the characteristics of learning at the two levels of education, namely PAUD and SD. The first article on PAUD uses authentic and formative assessment to assess nature-based learning at the early childhood level. This approach is suitable because it prioritizes evaluation based on performance or real products produced by children. Then formative assessment serves to provide continuous feedback throughout the learning process. Elementary school combines formative, summative and problem-based assessments to provide a holistic picture of students' progress in understanding STEM learning concepts and to measure learning outcomes and students' deep understanding of STEM concepts.

Assessments in this table include formative assessments, summative assessments, project-based assessments, authentic assessments, as well as some evaluation methods used during the pandemic or to assess STEM literacy in a broader context. Assessment methods in junior and senior high schools emphasize formative, authentic and project-based assessments that allow students to engage in active, interactive and integrated learning. Summative assessments are more widely used in SD and SMP to evaluate end-of-learning achievements in specific topics, such as science literacy and climate change. Science teachers use authentic assessments to evaluate their ability and understanding in teaching STEM, which is highly relevant to curriculum adaptation during the pandemic. Some articles are related to science teachers such as research on using Rasch validation instruments to measure generation Z's application of STEM during the pandemic. The summative assessment and validation of this instrument was used to assess the effectiveness of STEM approaches in the classroom during distance learning and adaptation to pandemic conditions

Table 8. Assessment Model/Method for Mixed Levels

Title	Year	Assessment Method/Model	Level
SOLO taxonomy as EFL students, disciplinary literacy evaluation tool in technology-enhanced integrated astronomy course	2019	formative assessment	Junior high school and senior high school
The UK's Project Faraday and Secondary STEM Education	2021	project-based, authentic, and formative assessments	Junior high school and senior high school
Using empirical science education in schools to improve climate change literacy	2023	summative assessment	Elementary school and junior high school
Online teaching during the COVID-19 pandemic: exploring science/STEM teachers, curriculum and assessment practices in Canada	2022	summative assessment	Elementary School Junior High School Science Teacher
Rasch Validation of Instrument Measuring Gen-Z Science, Technology, Engineering, and Mathematics (STEM) Application in Teaching during the Pandemic	2022	summative assessment	Gen Z Indonesia
How science teachers deal with STEM education: An explorative study from the lens of curriculum ideology	2024	authentic assessment	Elementary School Science Teacher

Then for the latest innovations in STEM learning assessment methods or models include several articles able to integrate technology or interdisciplinary approaches. As for more details are in the table 8.

The table 8 summarizes the latest innovations in STEM learning models published in the 2019-2024 timeframe. These articles demonstrate the integration of modern technologies and interdisciplinary approaches in supporting the assessment of STEM learning at various levels of education, from junior high school to university. Formative assessment models are the most dominant type of assessment, with key innovations such as the use of eye tracking technology for learning evaluation and gamification approaches to support interactive learning. Other technology-based innovations, such as browser-based stimulus and large language models, also show significant potential in improving the efficiency of automated assessment.

Table 9. Innovations and Interdisciplinary Approaches in STEM Learning

Article Title	Year	Types of Evaluation Models	Innovation (Technology or Interdisciplinary Approach)	Educational Level
A User Interface (UI) and User eXperience (UX) evaluation framework for cyberlearning environments in computer science and software engineering education	2020	Formative assessment	Use of UI/UX in technology-based assessment	University
Game-Based Learning: Enchaning Student Experience, Knowledge Gain, and Usability In Higher Education Porgramming Courses	2022	Summative assessment	Interdisciplinary approach with game based learning	University
Field science in the age of online learning Dynamic instruction of techniques to assess soil physical properties	2022	Formative assessment	Incorporation of technology for scientific field-based assessment	University
An application of Bayesian inference to examine student retention and attrition in STEM Education	2023	Data-driven assessment	Use of Bayesian Inference in data-driven assessment	University
Linking information from multiple representations: an eye tracking stdudy	2023	Formative assessment	Use of eye tracking technology for formative assessment	University
Electronic Field Trips for Science Engagement: The Streaming Science Model	2019	Formative assessment	Use of technology for field trips in science learning	Senior High School
Question generating supported by blended learning platform: Issues of social justice for environmental education	2019	Project-based and formative assessment	Interdisciplinary approach in environmental learning using blended learning	Senior High School
The Gamification of XFEL Education Using XFEL Crystal Blaster	2022	Formative assessment	Interdisciplinary approach in environmental learning using blended learning	Senior High School

Article Title	Year	Types of Evaluation Models	Innovation (Technology or Interdisciplinary Approach)	Educational Level
Browser-based simulation for novice-friendly classroom robotics	2023	Summative assessment	Use of browser-based simulation in robotics learning	Senior High School
Applyng large language models and cahin-of-thought for automatic scoring	2024	Technology-based assessment	Use of big language and chain-of-thought models for automated assessment	Junior High School
SOLO taxonomy as EFL students' disciplinary literacy education tool in technology-enhanced integrated astronomy course	2019	Formative assessment	Using SOLO taxonomy in a technology-based astronomy course to evaluate EFL students' disciplinary literacy	Junior High School-Senior High School
The UK's Project Faraday and Secondary STEM Education	2021	Project-based, authentic, and formative assessment	Using project-based and authentic approaches that integrate interdisciplinary STEM learning	Junior High School-Senior High School

Project-based approaches are widely applied, especially at the high school level to integrate STEM with other disciplines. This interdisciplinary approach adds value by enhancing student engagement and the relevance of STEM learning to real life. At the tertiary level, advanced innovations such as Bayesian analysis and UI/UX-based assessment are the main focus. Although innovations are more prevalent in higher education, the implementation of technology for junior and senior high school levels shows potential for further development. The overall findings indicate that innovations in STEM assessment continue to evolve to address the need for more relevant, effective and technology-based learning.

Discussions

RQ 1: What are the Most Frequently Used Assessment Methods or Models in STEM Learning?

Based on the analysis of 36 relevant articles, assessment in STEM learning can be divided into two main categories: formative and summative assessment. Summative assessment is generally conducted after the learning process is complete and is used to assess the overall achievement of students (Adinda et al., 2021). At all levels, the most dominant assessment implemented in STEM learning is summative assessment.

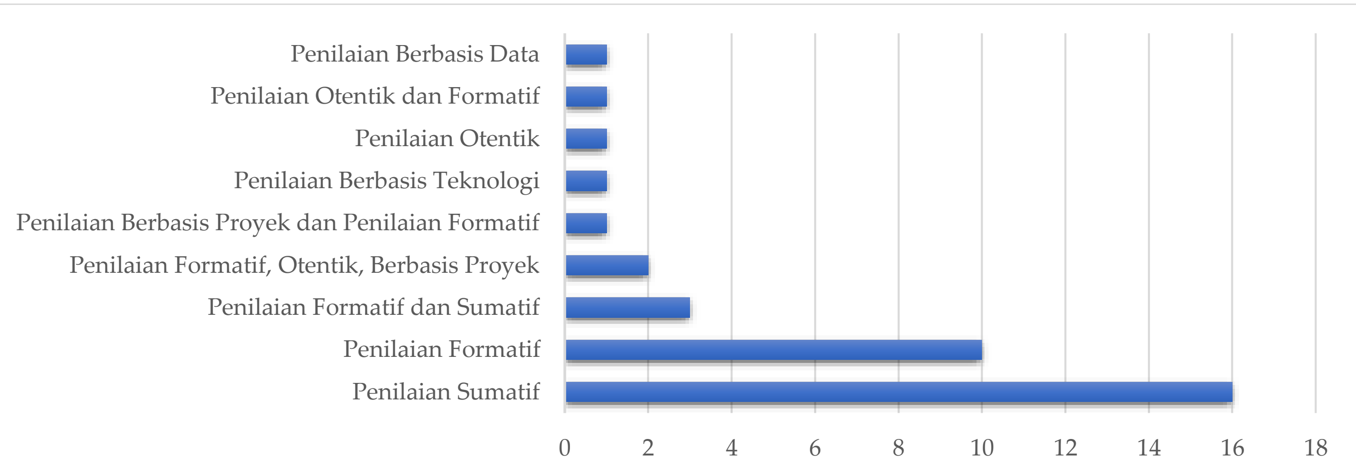

Figure 4. Assessment in STEM learning

At the university level, the most commonly used assessments are formative and summative. Research (Sato et al., 2019), shows that formative assessments use rubrics developed specifically for the previous course, which helps students articulate their understanding more clearly. Summative research at this level, such as that conducted by Susac et al. (2023), uses technology and PISA tests to assess students' understanding of math and science materials.

At the high school level, summative assessments are conducted using inquiry skills tests, such as the one by Ješková et al. (2022), which assesses students' ability to select relevant variables and the relationships between them. Meanwhile, at the junior high school level, summative assessment is conducted through surveys or questionnaires to measure students' scientific self-efficacy and career interest in STEM fields (Broder et al., 2019). At the elementary level, summative assessment can be a content knowledge test, as conducted by Rehmat et al. (2020), involving multiple-choice questions and open-ended science, math, engineering and technology questions. At the early childhood level, assessment is formative and authentic, as shown by Rehmat et al. (2020), where research often relies on observation and anecdotal recording of children's skills in technology use.

Summative assessment is also used at the high school level with a combined approach of formative and summative assessment, as done by Rahman (2021), who used questionnaire responses to assess learning outcomes in math and science lessons with traditional and robotics approaches. At the primary school level, formative and summative assessment was conducted by Watson et al. (2023), who identified answer categories to answer questions related to formative and summative aspects of STEM learning.

Formative assessment is also applied at the middle and high school levels, as described by Chubko et al. (2019), who used the SOLO taxonomy to measure students' disciplinary literacy in STEM education, particularly for students learning science in English as a foreign or second language. This assessment was conducted in a STEM integrated astronomy course with digital storytelling. Beyond assessment models, valid assessment instruments are essential for identifying students' learning needs and supporting effective instructional decisions (Asnawi et al., 2023).

In addition, research Kumar et al. (2023) demonstrated the use of summative assessments that measured students' interest in STEM learning through pre- and post-quizzes. Summative assessments were also found in DeCoito et al. (2022), which focused on STEM teachers' experiences with online teaching during the COVID-19 pandemic, and Qudratuddarsi et al. (2022), which tested the validity of STEM application instruments using Rasch analysis.

In general, summative assessment is more widely used because it provides an overall picture of student achievement or learning outcomes. However, formative assessment is also important in providing feedback and facilitating student learning development, as described by Sirianansopa (2024) and Alshakhi (2024), who emphasize the difference between formative and summative assessment to measure student understanding.

From the 36 articles analyzed using the VOS viewer application, a network map was obtained that shows how the relationship between the main keywords in the research related to STEM learning. The largest cluster in this map is STEM education, which includes keywords such as assessment, education, and students, indicating a major focus on assessment in STEM learning. Another cluster depicts the trend of technology integration, seen in the strong relationship between technology and learning, illustrating how technology is used in research. In addition, there is an emphasis on interdisciplinary learning, reflecting innovations in assessment and more active, real-life relevant approaches to learning. The integration of technology in STEM assessment is also supported by previous studies showing that digital assessment tools can provide alternative approaches for evaluating students' learning outcomes more effectively and flexibly (Mellyzar et al., 2024).

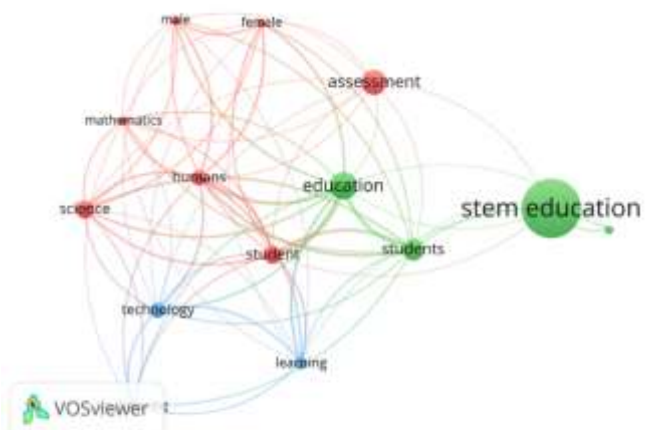

Figure 5. STEM education network map

Having discussed the main trends in models or methods in STEM assessment in RQ 1, this research will also examine further innovations in assessment methods or models that can integrate technology and interdisciplinary approaches.

RQ 2: What are the latest innovations in STEM learning assessment methods or models, and to what extent are they able to integrate technology or interdisciplinary approaches?

Of the 36 articles selected, 12 adopted recent innovations and interdisciplinary approaches in STEM learning assessment methods or models, as shown in Table 8. Most of these articles were from the university level, with 8 articles covering technological innovations relevant to all levels of education. The following is a description of the innovations found in these articles, grouped by education level:

For the university level, there are five articles that explain technological innovation and interdisciplinary approaches, namely: Research by Alomari et al. (2020) proposes a UI/UX evaluation framework to assess digital learning design that focuses on user experience, including four methods: pre/post-test, heuristic evaluation, cognitive search, and UX survey. Innovation: The framework extends evaluation practices by addressing usability aspects, which enhances students' understanding of software development concepts, particularly software testing. Bertolini et al. (2023) used a Bayesian approach to combine LMS data and concept diagnostic assessments. Innovation: This method offers a more personalized and interpretable data-driven assessment to evaluate student performance. Susac et al. (2023) used eye-tracking technology to analyze students' visual attention while solving PISA graphical problems. Innovation: This technology provides deep insight into students' analytical skills in interpreting graphs and visual data. Hinckley et al. (2022) developed a collaboration-based module to teach soil analysis. Innovation: This interdisciplinary approach helps students apply practical knowledge in STEM analysis in the field of environmental science.

At the high school level, research conducted by Lee et al. (2024) applied the Streaming Science model using mobile technology to increase student engagement in STEM. Innovation: The use of mobile technology expands accessibility and increases student interest in STEM learning. Stein et al. (2023) introduced a virtual robotics platform that teaches computational thinking skills through programming challenges. Innovation: Virtual robotics offers challenge-based learning to understand STEM concepts in an applied manner. Zhao et al. (2022) used serious games such as XFEL Crystal Blaster to improve students' understanding of advanced science. Innovation: This educational game combines interactive learning with real-world visualization. Oversby et al. (2019) developed a question generation-based assessment method through a blended learning platform. Innovation: This method provides deeper insight into students' thinking and their understanding of STEM content.

At the junior high school level, research conducted by Lee et al. (2024), explored the use of large language models (LLMs) such as GPT-3.5 and GPT-4 for automated assessment. Innovation: LLMs provide an efficient and consistent solution in scoring students' written answers. For the junior high school level, technological innovation and interdisciplinary approaches were conducted by Chubko et al. (2019) applying SOLO taxonomy to measure students' disciplinary literacy development in a STEM-based astronomy course. Innovation: This taxonomy maps students' skill development through clear structural stages. Then Edith Cowan University (2021) reported on a makerspace project that supports project-based learning in STEM. Innovation: These makerspaces provide interdisciplinary practical experiences to explore STEM concepts. Similarly, interdisciplinary STEM approaches, such as STEM-PjBL, have also been utilized to assess students' higher-order competencies through contextual and problem-based learning activities (Lestari et al., 2024).

The innovations found in these studies show that assessment in STEM is increasingly dynamic and tailored to students' needs. At university level, approaches are more likely to be data-driven and deep analytical, such as eye-tracking and Bayesian approaches. At the high school level, technological innovations focus more on accessibility and challenge-based learning, such as robotics and educational games. At the junior high level, exploration of new technologies such as LLM offers assessment efficiencies. Interdisciplinary approaches and the use of advanced technologies provide new ways to evaluate student understanding, supporting more in-depth, personalized and applied learning. This shows great potential in improving the quality of STEM learning at various levels of education.

Conclusion

Based on the systematic analysis of assessment practices, technological innovations, and interdisciplinary approaches in STEM learning, this study shows that assessment practices across educational levels, from primary education to higher education, are still dominated by summative assessment. Summative assessment is commonly used because it provides an overall description of students' achievement and learning outcomes. However, formative assessment also plays an important role in supporting students' learning development through continuous feedback. The findings indicate that STEM assessment is gradually shifting toward more adaptive and technology-supported approaches that emphasize students' skills, engagement, and the application of STEM concepts. This study contributes to the development of assessment in STEM education by mapping current assessment trends, technological innovations, and interdisciplinary practices across different educational levels. The findings reveal that the integration of technology, including data-driven assessment, eye-tracking technology, Bayesian approaches, mobile-based assessment, virtual robotics, serious games, and large language models (LLMs), offers potential for developing more effective and personalized assessment practices. Furthermore, interdisciplinary approaches, such as makerspace projects and STEM-based courses, support the development of authentic assessments that allow students to integrate knowledge from multiple disciplines in solving real-world problems. The findings of this study provide important implications for STEM educators and researchers. For educational practitioners, the integration of technology and interdisciplinary approaches can support the development of meaningful assessment practices that assess not only conceptual understanding but also higher-order thinking skills, creativity, collaboration, and problem-solving abilities. For researchers, this study provides a basis for further investigation into the effectiveness, validity, and implementation of emerging STEM assessment approaches in various educational contexts. Future research should further investigate the long-term effectiveness of technology-based and interdisciplinary assessments, particularly regarding their effectiveness in supporting students' STEM competencies across different educational levels. In addition, further studies are needed to examine the reliability, validity, and ethical considerations of artificial intelligence-based assessment tools, especially the use of LLMs in assessing complex STEM skills. Research should also explore strategies to overcome implementation challenges, including limited technological infrastructure, teachers' digital competence, and differences in educational contexts, to support the wider implementation of innovative STEM assessment practices.

Acknowledgments
The authors would like to express their gratitude to Yogyakarta State University for the support and facilities that played a crucial role in the smooth progress and completion of this research.

Author Contributions
M.A.: Conceptualization, methodology, literature search, data analysis, interpretation of results, and writing of the original draft. Y.W. and H.K.: Reviewing the research focus and objectives, providing suggestions, and supervising the manuscript development. P.R.A. and B.W.: Article screening, data curation, and selection of relevant studies for analysis. All authors have read and approved the final version of the manuscript.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest

References

Adinda, A. H., Siahaan, H. E., Raihani, I. F., Aprida, N., Fitri, N., & Suryanda, A. (2021). Summative Assessment and Formative Assessment of Online Learning. *Report of Biological Education*, *2*(1), 1–10. https://doi.org/10.37150/rebion.v2i1.1024

Alhayat, A., & J., R. C. (2024). Designing Learning Assessment with STEAM-Based Projects in Implementing of the Kurikulum Merdeka. *Jurnal Penelitian Pendidikan IPA*, *10*(11), 9141–9157. https://doi.org/10.29303/jppipa.v10i11.7266

Alomari, H. W., Ramasamy, V., Kiper, J. D., & Potvin, G. (2020). A User Interface (UI) and User eXperience (UX) evaluation framework for cyberlearning environments in computer science and software engineering education. *Heliyon*, *6*(5), e03917. https://doi.org/10.1016/j.heliyon.2020.e03917

Alshakhi, A. (2024). Speaking Skill Assessment Instrument Validity: An Investigation into Instructors' Perceptions. *Journal of Ecohumanism*, *3*(7). https://doi.org/10.62754/joe.v3i7.4543

Asghar, A., Huang, Y.-S., Elliott, K., & Skelling, Y. (2019). Exploring Secondary Students' Alternative Conceptions about Engineering Design Technology. *Education Sciences*, *9*(1), 45. https://doi.org/10.3390/educsci9010045

Asnawi, A., Sahudra, T. M., Ramadhani, D., Kenedi, A. K., Aosi, G., Wardhana, M. R., & Khalil, N. A. (2023). Development of Digital Diagnostic Test Instruments for Differentiated Learning Process in Elementary Schools. *Jurnal Penelitian Pendidikan IPA*, *9*(SpecialIssue), 460–466. https://doi.org/10.29303/jppipa.v9iSpecialIssue.6125

Bertolini, R., Finch, S. J., & Nehm, R. H. (2023). An application of Bayesian inference to examine student retention and attrition in the STEM classroom. *Frontiers in Education*, *8*, 1073829. https://doi.org/10.3389/feduc.2023.1073829

Broder, E. D., Guilbert, K. E., Tinghitella, R. M., Murphy, S. M., Ghalambor, C. K., & Angeloni, L. M. (2019). Authentic Science with Dissemination Increases Self-Efficacy of Middle School Students. *Integrative and Comparative Biology*, *59*(6), 1497–1508. https://doi.org/10.1093/icb/icz140

Christopoulos, A., Pellas, N., & Laakso, M.-J. (2020). A Learning Analytics Theoretical Framework for STEM Education Virtual Reality Applications. *Education Sciences*, *10*(11), 317. https://doi.org/10.3390/educsci10110317

Chubko, N., Morris, J. E., McKinnon, D. H., Slater, E. V., & Lummis, G. W. (2019). SOLO taxonomy as EFL students' disciplinary literacy evaluation tool in technology-enhanced integrated astronomy course. *Language Testing in Asia*, *9*(1), 19. https://doi.org/10.1186/s40468-019-0095-6

DeCoito, I., & Estaiteyeh, M. (2022). Online teaching during the COVID-19 pandemic: Exploring science/STEM teachers' curriculum and assessment practices in Canada. *Disciplinary and Interdisciplinary Science Education Research*, *4*(1), 8. https://doi.org/10.1186/s43031-022-00048-z

Donham, C., Hong, H., Signorini, A., Menke, E., & Kranzfelder, P. (2024). If you are learner-centered and you know it, raise your hand: Perspectives on and implementation of pedagogical changes by

SCIENCE instructors during the COVID -19 pandemic. *Biochemistry and Molecular Biology Education*, *52*(4), 386–402. https://doi.org/10.1002/bmb.21826

Gunawan, I. Z. B., Gita, R. S. D., & Kustiyowati. (2025). The Urgency of STEM Learning Approach in Improving Students' Creative Thinking and Collaboration Skills. *Jurnal Penelitian Pendidikan IPA*, *11*(9), 157–167. https://doi.org/10.29303/jppipa.v11i9.11890

Hinckley, E.-L. S., & Fendorf, S. (2022). Field science in the age of online learning: Dynamic instruction of techniques to assess soil physical properties. *Frontiers in Education*, *7*, 959776. https://doi.org/10.3389/feduc.2022.959776

Ješková, Z., Lukáč, S., Šnajder, Ľ., Guniš, J., Klein, D., & Kireš, M. (2022). Active Learning in STEM Education with Regard to the Development of Inquiry Skills. *Education Sciences*, *12*(10), 686. https://doi.org/10.3390/educsci12100686

Khalid, I. L., Abdullah, M. N. S., & Mohd Fadzil, H. (2024). A Systematic Review: Digital Learning in STEM Education. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *51*(1), 98–115. https://doi.org/10.37934/araset.51.1.98115

Kumar, P., Sahani, J., Rawat, N., Debele, S., Tiwari, A., Mendes Emygdio, A. P., Abhijith, K. V., Kukadia, V., Holmes, K., & Pfautsch, S. (2023). Using empirical science education in schools to improve climate change literacy. *Renewable and Sustainable Energy Reviews*, *178*, 113232. https://doi.org/10.1016/j.rser.2023.113232

Lee, G.-G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence*, *6*, 100213. https://doi.org/10.1016/j.caeai.2024.100213

Lehmeidi Dong, M., Feng, Z., McLamb, F., Griffin, L., Vazquez, A., Hirata, K. K., Bozinovic, L., Vasquez, M. F., & Bozinovic, G. (2023). Life Science Research Immersion Program Improves STEM-Specific Skills and Science Attitudes among Precollege Students. *Journal of Microbiology & Biology Education*, *24*(1), e00078-22. https://doi.org/10.1128/jmbe.00078-22

Lestari, H. D., Rahmawati, Y., & Usman, H. (2024). STEM-PjBL Learning Model to Enhance Critical Thinking Skills of Students on Magnets, Electricity, and Technology. *Jurnal Penelitian Pendidikan IPA*, *10*(8), 6027–6037. https://doi.org/10.29303/jppipa.v10i8.8153

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 Assesing High School Science Learning. *Jurnal Penelitian Pendidikan IPA*, *10*(3), 1362–1374. https://doi.org/10.29303/jppipa.v10i3.6416

Oversby, J., Sanders, J., Abdul Talib, C., Thoe, N. K., & Esa, N. (2019). Question Generating Supported by Blended Learning Platform: Issues of Social Justice for Environmental Education. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(5). https://doi.org/10.29333/ejmste/105848

Qudratuddarsi, H., Hidayat, R., Shah, R. L. Z. B. R. M., Nasir, N., Imami, M. K. W., & Nor, R. B. M. (2022). Rasch Validation of Instrument Measuring Gen-Z Science, Technology, Engineering, and Mathematics (STEM) Application in Teaching during the Pandemic. *International Journal of Learning, Teaching and Educational Research*, *21*(6), 104–121. https://doi.org/10.26803/ijlter.21.6.7

Rahman, S. M. M. (2021). Assessing and Benchmarking Learning Outcomes of Robotics-Enabled STEM Education. *Education Sciences*, *11*(2), 84. https://doi.org/10.3390/educsci11020084

Ramadhani, D., Kenedi, A. K., Rafli, M. F., Harahap, H., & Mardin, A. (2025). STEM-Based HOTS Digital Assessment Model in Measuring Higher-Order Thinking Skills of Prospective Teacher Students. *Jurnal Penelitian Pendidikan IPA*, *11*(10), 111–120. https://doi.org/10.29303/jppipa.v11i10.12812

Rasmi, D. P., Wibisana, S. S., & Hendri, M. (2025). The Integrating STEM in Hooke's Law and Elasticity Worksheets: Enhancing Student Collaboration and Learning Outcomes. *Jurnal Penelitian Pendidikan IPA*, *11*(2), 573–581. https://doi.org/10.29303/jppipa.v11i2.10166

Rehmat, A. P., & Hartley, K. (2020). Building Engineering Awareness: Problem Based Learning Approach for STEM Integration. *Interdisciplinary Journal of Problem-Based Learning*, *14*(1). https://doi.org/10.14434/ijpbl.v14i1.28636

Saputra, W. T., Rustaman, N. Y., & Rusyati, L. (2025). Exploring the Relationship between Logical Reasoning Skills, Scientific Literacy, and Academic Achievement in Science Courses Among Secondary Students: A Mixed Method Study. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 303–315. https://doi.org/10.29303/jppipa.v11i1.9547

Sato, B. K., Hill, C. F. C., & Lo, S. M. (2019). Testing the test: Are exams measuring understanding? *Biochemistry and Molecular Biology Education*, *47*(3), 296–302. https://doi.org/10.1002/bmb.21231

Shiddiqi, M. H. A., Arthamena, V. D., Ayyubi, M., Manarisip, A. J., & Aznam, N. (2024). Systematic Literature Review: Analysis of Misconception Problems and Diagnostic Instruments for Learning Chemistry. *Jurnal Penelitian Pendidikan IPA*, *10*(4), 168–179. https://doi.org/10.29303/jppipa.v10i4.5189

Sirianansopa, K. (2024). Evaluating students' learning achievements using the formative assessment technique: A retrospective study. *BMC Medical Education*, *24*(1), 1373. https://doi.org/10.1186/s12909-024-06347-5

Solihah, P. A. S., Kaniawati, I., Samsudin, A., & Riandi, R. (2024). Need Analysis of STEM-ESD Based Development on Teaching Materials and Instrument Test. *Jurnal Penelitian Pendidikan IPA*, *10*(3), 1124–1133. https://doi.org/10.29303/jppipa.v10i3.6516

Stein, G., Jean, D., Brady, C., & Lédeczi, Á. (2023). Browser-based simulation for novice-friendly classroom robotics. *Frontiers in Computer Science*, *4*, 1031572. https://doi.org/10.3389/fcomp.2022.1031572

Susac, A., Planinic, M., Bubic, A., Jelicic, K., & Palmovic, M. (2023). Linking information from multiple representations: An eye-tracking study. *Frontiers in Education*, *8*, 1141896. https://doi.org/10.3389/feduc.2023.1141896

Watson, J., Wright, S., Fitzallen, N., & Kelly, B. (2023). Consolidating understanding of variation as part of STEM: Experimenting with plant growth. *Mathematics Education Research Journal*, *35*(4), 961–999. https://doi.org/10.1007/s13394-022-00421-1

Zhao, D., Muntean, C. H., Chis, A. E., Rozinaj, G., & Muntean, G.-M. (2022). Game-Based Learning: Enhancing Student Experience, Knowledge Gain, and Usability in Higher Education Programming Courses. *IEEE Transactions on Education*, *65*(4), 502–513. https://doi.org/10.1109/TE.2021.3136914