JSEP (Journal of Sciences Education and Practice)
Published by Universitas Pakuan
https://jsep-pasca.unpak.ac.id/index.php/jsep
Vol. 10, No. 2, 2026 | Pages 395-410
ISSN: E 2549-7170 / P 2548-950X | https://doi.org/10.33751/jsep.v10i2.30
This article licensed under Creative Common Attribution 4.0 International

The Use of the Rasch Model in Measuring Critical Thinking Skills in Science Education: A Systematic Literature Review

Rahayu Kholifatul Azizah, Novi Ratna Dewi*, Parmin

Science Education Study Program, Universitas Negeri Semarang, Semarang, Indonesia

✉* noviratnadewi@mail.unnes.ac.id

Article Info

Keywords:
Rasch model; critical thinking skills; science education; learning assessment; systematic literature review

Received:
07/08/2026
Revised:
13/08/2026
Accepted:
13/09/2026

Abstract

Background: Assessing critical thinking skills in science education requires valid, reliable, and objective assessment instruments capable of accurately representing students' abilities. Although the Rasch Model has been increasingly applied in educational research, comprehensive synthesis studies examining instrument quality and research trends remain limited. **Methods:** This study employed a Systematic Literature Review (SLR) approach following PRISMA 2020 guidelines. Articles were collected from Google Scholar and Scopus databases covering publications from 2016–2026. A total of 19 studies met the inclusion criteria and were analysed through structured data extraction, quality assessment, narrative synthesis, and bibliometric mapping using VOS viewer. **Results:** The findings reveal that the Rasch Model is predominantly used to evaluate the quality of critical thinking assessment instruments, particularly in physics and chemistry education. Frequently reported indicators include construct validity, reliability, item fit, and person ability. Bibliometric analysis indicates that research trends mainly focus on instrument validation, while integration with contextual learning approaches such as STEM and problem-based learning remains relatively limited. **Conclusion:** The results highlight the need to expand the application of the Rasch Model toward the development of authentic assessment instruments in science learning to support more comprehensive measurement of critical thinking.

INTRODUCTION

Developments in the 21st century require people to possess a variety of skills relevant to the needs of the times, one of which is critical thinking, which is essential for survival and success in an increasingly competitive global environment (Syafitri et al., 2021). In science education, critical thinking skills are an essential ability that not only helps students understand concepts but also involves the process of analyzing information,

evaluating various possibilities, formulating solutions, and drawing conclusions based on the situations or problems students face (Nuryaningsih & Dewi, 2024). The concept of critical thinking itself is understood as students' higher-order cognitive skills that enable them to process information, analyze problems, and make rational, evidence-based decisions or conclusions (Wardana et al., 2025).

The 2022 PISA results show that Indonesia's science score was 383, down from 396 (2018) and 403 (2015), with only 34% of students reaching Level 2 or higher, indicating a low ability to evaluate evidence and draw conclusions as part of critical thinking (OECD, 2023). This is in line with research (Yunita & Dewi, 2020), who state that assessment instruments in schools are still dominated by questions that only measure recall and comprehension (LOTS) skills and do not specifically measure higher-order analytical thinking skills. Therefore, the process of developing critical thinking skills must be supported by an assessment system capable of accurately identifying these abilities through diagnostic tools that not only collect responses but also analyze the results and provide follow-up learning strategies (Novriyanti & Arthur, 2024).

The theory (Maryani et al., 2021) also states that students' critical thinking skills in science learning remain low, with an average score of 53.07, particularly in their ability to analyze arguments, process observational data, make decisions, and express opinions based on evidence. This finding is supported by Putri et al. (2024) through a meta-analysis of 12 studies showing that junior high school students' critical thinking skills in science learning remain in the low-to-moderate range, thus necessitating an assessment tool capable of identifying students' abilities more accurately. However, measurement remains a challenge because instruments tend to focus on simple cognitive aspects and do not yet comprehensively represent the dimensions of critical thinking, resulting in suboptimal measurement accuracy (Prihatni et al., 2016).

Conventional measurement approaches, such as Classical Test Theory (CTT), have limitations because they rely heavily on sample characteristics and item difficulty levels, which can lead to measurement results that are less stable and objective. The limitation of Classical Test Theory (CTT) lies in its focus on total scores as the primary indicator of student ability, thereby failing to provide in-depth information regarding the characteristics of individual test items. This approach does not account for differences in item difficulty levels or variations in individual ability to respond to questions, so the resulting analysis tends to be general and fails to specifically describe the strengths and weaknesses of each student (Andryan et al., 2024)

As an alternative, the Rasch Model, which is part of Item Response Theory (IRT), can be used as an approach to measurement analysis and offers a more comprehensive approach to test instrument analysis. The Rasch Model has various advantages that make it a powerful analytical tool in educational and psychometric research, particularly when used with Winsteps software. One of its main advantages is the ability to analyze person fit that is, the fit of an individual's responses to test items based on the Rasch Model. In Winsteps, this fit is measured through statistics such as infit and outfit mean squares, which serve to assess the consistency of respondents' answers with the model's assumptions. This analysis helps identify respondents who provide inconsistent or insincere answers, thereby providing a more accurate picture of the quality of the data obtained (Adi, 2025). Various studies have developed assessment instruments for critical thinking skills by utilizing the Rasch Model as a psychometric analysis approach capable of providing more comprehensive information regarding the quality of the instruments

(Widodo & Chotimah, 2023; Riyadi et al., 2024). However, although research on the application of the Rasch Model in science education is growing, systematic reviews summarizing perspectives on the use of the Rasch Model in measuring critical thinking skills remain relatively limited. Most studies focus on instrument development or individual empirical analyses, but few have integrated these findings through a systematic literature review approach. This research gap results in a lack of a comprehensive overview of research trends, methodological approaches, implementation advantages, and challenges in using the Rasch Model within the context of science learning.

Although research on the application of the Rasch Model in science education has increased, existing review studies addressing its specific application to critical thinking assessment in science education remain limited. Previous studies have generally focused on individual instrument development and psychometric analysis, while an integrated synthesis of scientific domains, assessment characteristics, psychometric quality, and research trends remains insufficiently documented. Through a systematic literature review approach, this study aims to analyze the distribution of the use of the Rasch Model and the types of assessment instruments in science education, evaluate the quality of test instruments for critical thinking skills or higher-order thinking skills analyzed using the Rasch model, and identify trends in research focus based on bibliometric analysis and density visualization.

METHODS

This research employed a Systematic Literature Review (SLR) to identify, evaluate, and integrate empirical studies examining the application of the Rasch Model in assessing critical thinking skills in science education. The SLR method was selected because it enables literature exploration in a structured, transparent, and reproducible manner, thereby allowing a comprehensive synthesis of scientific evidence. The review procedure followed the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework, which includes the stages of identification, screening, eligibility evaluation, and final inclusion of relevant studies. The literature search was carried out using the Publish or Perish (PoP) software, which is integrated with the Google Scholar and Scopus databases. Google Scholar was selected because it has extensive indexing coverage in the field of science education; furthermore, it includes national journals that are not yet fully indexed by Scopus, making it relevant for identifying empirical studies on the use of the Rasch Model. Scopus was used to ensure academic quality by featuring reputable international articles that have undergone rigorous peer review.

The combination of these two databases aims to minimize database bias through a database triangulation approach and to capture research developments across international contexts and within the context of science education. The search was conducted on March 11, 2026, at 3:35 PM WIB, with a publication range of 2016–2026. A ten-year span was chosen because it effectively represents current trends as well as the evolution of the Rasch Model in a more stable and comprehensive manner, while also selecting the most recent articles relevant to research developments, particularly in the measurement of critical thinking skills in science education (Costa et al., 2021).

The literature search strategy employed a Boolean search to identify studies on the use of the Rasch Model in measuring critical thinking skills in science education, using the search string "critical thinking" AND "Rasch model" AND ("science education" OR physics

OR chemistry OR biology OR "integrated science"). The search was conducted in the keywords field, limited to peer-reviewed journal articles in English and Indonesian within the field of science education. All search results were exported in RIS format and subsequently managed using Zotero for document organization and duplication removal prior to the article screening stage. The article selection criteria were established prior to the screening process to ensure the objectivity of the review, as presented in the following table:

Table 1. Criteria for Study Inclusion and Exclusion

Selection Criteria	Inclusion Criteria	Exclusion Criteria
Research topic	The study examines the use of the Rasch Model and critical thinking in science.	Does not cover the Rasch Model or critical thinking in science.
Field of study	Research related to the field of science	Research outside the field of science.
Type of publication	Peer-reviewed scientific journal articles.	Proceedings, books, book chapters, editorials, undergraduate theses, master's theses, or non-scientific documents.
Analysis method	The study uses the Rasch model.	Uses other statistical analyses
Publication years	Published between 2016 and 2026.	Published outside that time frame.
Publication language	Articles are written in English or Indonesian.	The article is written in a language other than English or Indonesian.
Document access	Articles are available in full text and can be accessed in their entirety.	The article is not fully accessible.

The study selection process in this research was conducted systematically using the PRISMA guidelines, as illustrated in the following diagram.

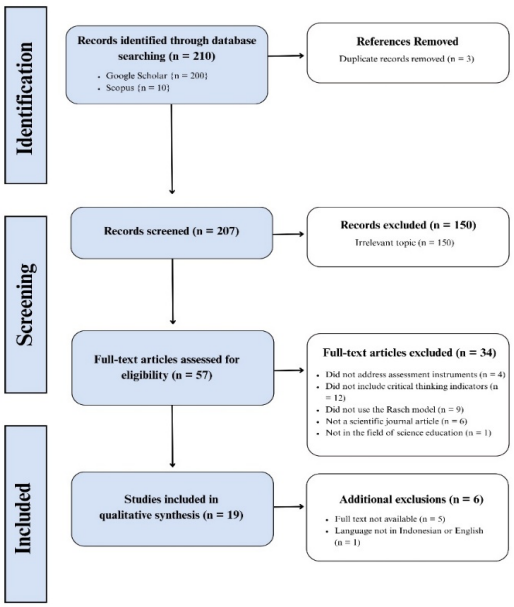


Figure 1. SLR using the prism method

The article selection procedure done according to guidelines in PRISMA 2020, consisting of the identification, screening, eligibility evaluation, and inclusion phases. During the identification stage, 210 articles were identified from Google Scholar (200) and Scopus (10); three (3) duplicate articles were then removed, leaving 207 articles. The screening stage, reviewed based on the title and abstract, then eliminates candidates that meet the criteria 150 articles, leaving 57 articles. At the eligibility stage, the full-text review excluded 23 articles for several reasons: they did not address assessment instruments (4), lacked critical thinking indicators (8), did not apply the Rasch Model (7), were not classified as scientific articles (2), were outside the science field (1), or were not aligned with the research focus (1), resulting in 34 remaining articles. The subsequent inclusion stage removed articles that were not fully accessible (5) and were in languages other than Indonesian or English (1), leaving 19 articles to be analyzed in the systematic literature review. To ensure transparency in the selection process, the total number of articles and the grounds for exclusion at each stage are summarized in the following table.

Table 2. Summary of the Article Selection Process Based on PRISMA 2020

Selection Stage	Number of Articles	Criteria / Activities	Reasons for Exclusion
Identification	210	Articles were retrieved from the Google Scholar and Scopus databases	-
References Removed	207	Removal of duplicate articles	Duplicate article detected
Screening (Title & Abstract)	57	Checking for topic relevance	Irrelevant to the topic; not in the field of science
Full-text articles assessed for eligibility	34	Analysis of methodology, research focus, and use of the Rasch model	Does not discuss instruments; does not include critical thinking indicators; does not use the Rasch model; not a scientific article
Included Studies	19	The articles met all inclusion criteria and were analyzed in the systematic review	Full text not available; language does not match

The metadata of eligible articles were exported in RIS format, compiled into a single database, and subjected to a data-cleaning process (removal of duplicates, standardization of authors and keywords, and verification of methodological consistency). Next, key data, including publication details, scientific field, instrument type, use of the Rasch Model, software, sample size, instrument quality, and research findings, are extracted for cross-study synthesis to identify research patterns and trends.

Quality assessment was conducted to ensure that the included articles met adequate methodological standards and were relevant to the research objectives. The screening, eligibility assessment, and quality assessment processes were conducted independently by a single reviewer using a structured rubric that included six criteria: topic relevance, clarity of research design, adequacy of sample size for Rasch analysis, reporting of Rasch analysis, instrument quality, and clarity of results reporting. Each criterion was

scored 1 if fully met, 0.5 if partially met, and 0 if not met, resulting in a total score ranging from 0 to 6, with quality categories of high (4–6), moderate (1.5–3.5), and low (0–1). Since the entire process was conducted by a single reviewer, inter-rater agreement was not calculated, and there was no procedure for resolving disagreements between raters; however, to ensure consistency, all articles were evaluated systematically using the same rubric and indicators.

Quality assessment is conducted to ensure that articles demonstrate adequate methodological quality and are relevant to the research objectives. The evaluation is carried out using a structured rubric that covers topic relevance, clarity of research design, and clarity of results and discussion. Each aspect is scored as 1 (the study fully met the criteria), 0.5 (the study partially met the criteria), and 0 (the study did not meet the criteria), with a total score ranging from 0-6 (Kumar et al., 2025).

Table 3. Summary of the Article Selection Process Based on PRISMA 2020

Criteria	Indicators
Topic relevance	Rasch and critical thinking in science
Research design	Methods clearly explained
Sample size	Sufficient for Rasch analysis
Rasch analysis	Reports item fit/reliability
Instrument quality	Evidence of validity and reliability provided
Reporting of results	Results clearly presented

Score categories: 4-6= High | 1.5-3.5= Moderate | 0-1= Low

Data synthesis was conducted using a narrative synthesis method, which involves grouping findings according to the research questions (RQ1–RQ3). The analysis focused on identifying common patterns, research trends, contradictions in the findings, and research gaps related to the use of the Rasch Model in measuring critical thinking skills in science education. A bibliometric analysis was conducted to identify research trends and thematic focuses related to the use of the Rasch Model in measuring critical thinking abilities in science education. The analysis was conducted on 19 articles that met the inclusion criteria using VOS viewer software version 1.6.20. Article metadata was exported from the database in RIS format, followed by a data cleaning process that included keyword normalization, merging terms with similar meanings, and removal of irrelevant keywords. Bibliometric visualization was used to identify dominant keywords, relationships between research themes, and research areas that have been rarely studied. The results of this analysis were then integrated with the findings of the systematic literature review to strengthen the interpretation of research trends and research gaps concerning the use of the Rasch Model in science education.

RESULT AND DISCUSSION

This Results and Discussion section presents a synthesis of the selected articles and a comprehensive analysis to identify patterns, trends, and implications of the use of the Rasch Model in science education, which are summarized in Table 4 below.

Table 4. Results of Article Review

Author, year	Science domain	Instrument type	Rasch model / software	Main findings	Implication
(Yulianto et al., 2021)	Science	HOTS Questions	Rasch Model	The HOTS instrument is valid, reliable, and capable of grouping students' abilities without bias.	Development of contextual HOTS assessment tools
(Yokhebed et al., 2025)	Biology	CT Self-Assessment	Rasch Model	The instrument has good validity and reliability and is capable of distinguishing	Development of assessment tools in other contexts
(Widiyawati et al., 2021)	Science	HOTS Test & Digital Literacy	Rasch Model	Students' HOTS skills are low to moderate, with differences based on gender.	Teachers need training in HOTS assessment
(Susongko et al., 2021)	Chemistry	Science Literacy Test	Rasch Model (DIF)	Most items are gender-neutral, even though students' abilities are still at a low-to-moderate level.	Assessment tools must be sensitive to student characteristics
(Mulyanti et al., 2022)	Chemistry	Multiple-Choice Practical Test	Rasch (Winsteps)	The instrument is valid, highly reliable, and suitable for use in laboratory settings.	Broader development of laboratory assessment tools
(Misbah et al., 2018)	Chemistry	Critical Thinking Test	Rasch (Winsteps)	The instrument meets the criteria for validity, reliability, and unidimensionalitas.	Development of innovative assessment tools and learning methods
(Laliyo, 2023)	Chemistry	Scientific Explanation Test	Partial Credit Model	There is variation in students' abilities and in the difficulty level of the items when it comes to formulating scientific explanations.	Learning strategies to improve CER
(Istiyono et al., 2019)	Physics	Computer-Based Computing Test	PCM (IRT)	CAT-based instruments demonstrate high validity and reliability in measuring critical thinking.	Implementation of CAT in other subjects
(Harta et al., 2020)	Chemistry	CAT (Competency Test)	Rasch Model	Students' analytical skills are low despite the high quality of the questions.	HOTS learning strategies are needed
(Hamdu et al., 2020)	Science	Multiple-choice test	Rasch (Winsteps)	The instrument is capable of measuring item difficulty levels and students' critical thinking skills.	Broader development of assessment tools
(Fauziah & Sukmawati, 2023)	Biology	STEM	Rasch (stacking)	RADEC-based instruction has been shown to improve students' higher-order thinking skills (HOTS).	Apply RADEC in other subjects

Author, year	Science domain	Instrument type	Rasch model / software	Main findings	Implication
(Ekawati et al., 2024)	Physics	HOTS (RADEC) test	Rasch (Quest)	A valid diagnostic tool for identifying students' early critical thinking skills.	Broader implementation
(Sujatmika, 2025)	Biology	Polytomous diagnostic test	Rasch Model	The diagnostic instrument is valid for identifying students' initial critical reasoning abilities.	Focus on lower-level aspects
(Sadhu & Laksono, 2018)	Chemistry	CT Test & Chemical Literacy	Rasch Model	The instrument is valid, reliable, and capable of measuring critical thinking skills and chemical literacy in an integrated manner.	The instrument can be used directly
(Sabekti & Khoirunnisa, 2018)	Chemistry	CT Test	PCM	Most items fit with sufficient reliability to measure critical thinking.	A larger sample size is needed
(Risdianto et al., 2021)	Physics	Student Response Questionnaire	Rasch Model	Rasch Model The questionnaire instrument is valid and shows positive responses to ethoscience learning.	Ethoscience learning is recommended
(Rachmadtullah, 2020)	Physics	CTIT Test	Rasch Model	The CTIT instrument is valid and reliable in measuring students' critical thinking.	Testing on a larger sample
(Putri & Istiyono, 2017)	Physics	STEM Performance Assessment	Rasch Model	The STEM-based performance assessment instrument is valid and reliable.	Digital development & other materials
(Purnami et al., 2021)	Science	Eco-Critical Thinking Test	Rasch Model	The instrument is capable of mapping abilities and item suitability well.	Development of environment-based instruments

The following is an analysis of the distribution of Rasch model usage by science subject area and the types of assessment instruments used in science education; the results are presented in Table 5 below.

Table 5. Distribution of Rasch Model Usage by Science Field

Field of Science	Number of Studies	Dominant Instrument Type	Assessment Method	Rasch Model Application
Physics	6	Multiple-choice/written tests, critical thinking tests, performance assessments	HOTS, critical thinking test, diagnostic test, STEM performance	Development of critical thinking tests, diagnostic assessments, computer-adaptive testing (CAT), performance analysis
Chemistry	6	Written tests, practical tests, integrated tests	Chemistry literacy, HOTS, experiments, integrated assessment	Instrument validation, item analysis, differential item functioning (DIF), science literacy
Biology	3	Self-assessment/surveys, critical thinking tests	Self-assessment, learning evaluation	Measurement of critical thinking, learning evaluation
Integrated Science	4	Contextual instruments, written tests	HOTS, STEM, eco-critical thinking	STEM assessment, development of contextual critical thinking

The analysis summary table indicates that the Rasch Model is used most in physics and chemistry (6 studies each), followed by biology (3 studies) and integrated science (4 studies). This pattern reflects the tendency to apply the Rasch Model in quantitative, structured science domains, which generally use written test instruments such as multiple-choice questions, critical-thinking tests, and HOTS-based diagnostic tests. Within the Item Response Theory (IRT) framework, the Rasch Model can accurately estimate key parameters, including item difficulty, student ability, and item fit. These findings align with (Misbah et al., 2018), who showed that the Rasch Model is effective for assessing reliability, Unidimensionalitas, and item quality, making it widely utilized in the development and validation of instruments in physics and chemistry, while biology and integrated science are more oriented toward mapping student abilities. Furthermore, the trend toward using written test instruments aligns with the revised Anderson and Krathwohl taxonomy, which classifies analysis, evaluation, and creation as indicators of higher-order thinking that are relatively easy to measure through tests (Umami et al., 2021). Theoretically, this dominance underscores the superiority of the Rasch Model over Classical Test Theory (CTT) in producing independent and accurate parameter estimates (Mulyanti et al., 2022).

However, there is a methodological contradiction between the psychometric strengths of the Rasch Model and the demands of 21st-century assessment. Although it excels in quantitative analysis, research remains dominated by objective test instruments and thus does not fully represent the complexity of critical thinking skills that develop through authentic scientific practice. In fact, according to the concept of authentic assessment outlined in the Ministry of Education and Culture Regulation No. 104 of 2014 (Rosidah et al., 2021), critical thinking abilities should be evaluated through activities such as investigation, data interpretation, and evidence-based argumentation. This situation indicates a research gap, particularly in the development of authentic assessment instruments based on problem-solving, socioscientific issues, and STEM approaches analyzed using the Rasch Model, especially in the fields of biology and integrated science. Therefore, from the theoretical and methodological implications of these findings is the

need to expand the use of the Rasch Model toward the development of authentic and contextual assessment instruments that require the analysis of real-world problems, evaluation of information, and the systematic formulation of solutions. This integration is expected to strengthen the construct validity of critical thinking skills while making a more meaningful contribution to science education practices focused on developing 21st-century skills. The following is a summary of the psychometric qualities of the critical thinking and HOTS instruments based on a Rasch model analysis, as presented in the table below.

Table 6. Quality of Rasch Model Instruments

Indicator	Rasch Criteria	Meets Criterion, n (%)	Does Not Meet Criterion, n (%)	Not Reported, n (%)	Total Studies
Item Fit (Infit–Outfit MNSQ)	0.5–1.5	16 (84%)	3 (16%)	0 (0%)	19
Person Reliability	≥0.80	14 (74%)	5 (26%)	0 (0%)	19
Item Reliability	≥0.80	17 (89%)	2 (11%)	0 (0%)	19
Separation Index	≥2.0	15 (79%)	4 (21%)	0 (0%)	19
Unidimensionality	Variance ≥20%	13 (68%)	6 (32%)	0 (0%)	19
DIF (Fairness Analysis)	DIF analysis reported	8 (42%)	–	11 (58%)	19
Cronbach's Alpha	≥0.70	18 (95%)	1 (5%)	0 (0%)	19

A total of 84% of the studies reported item fit that met the criteria, indicating that the majority of the instrument's items functioned consistently in measuring the construct of critical thinking skills. This indicates that data modelling using the Rasch Model can produce objective ability estimates because student responses align with the measurement model's expectations. This finding suggests that the collected data fit well with the assumptions of the Rasch measurement model, indicating that students' responses correspond to model expectations. Consistent with the study conducted by Pratiwi and Sugiharto (2025), infit and outfit MNSQ values within the acceptable range demonstrate that the test items function effectively in differentiating students' ability levels. Furthermore, item reliability in the "very good" category indicates the stability of the instrument's parameter estimates, meaning the instrument can produce consistent measurements across different populations, as emphasized by (Novriyanti & Arthur, 2024), who noted that high reliability reflects the instrument's ability to accurately distinguish students' ability levels.

A high separation index value also reinforces the argument that the instrument is capable of hierarchically grouping ability levels, in accordance with modern measurement principles that emphasize ability differentiation as the basis for 21st-century learning evaluation. The dominance of quality indicator reporting in physics and chemistry research compared to biology and integrated science can be explained by the fact that quantitative, problem-solving-based questions more readily satisfy the unidimensionality assumption of the Rasch Model, as explained by the hierarchical and measurable cognitive structure theory of science. However, methodological weaknesses remain: 32% of the studies have not demonstrated strong evidence of unidimensionality, and more than half have not reported Differential Item Functioning (DIF) analysis. This indicates that some studies still focus on technical reliability without thoroughly examining construct validity and measurement equity. Therefore, future research should integrate Rasch analysis into the

development of authentic, problem-solving-based assessments so that the instruments are not only statistically reliable but also conceptually valid in measuring students' critical thinking skills and 21st-century competencies. Below is a VOS viewer visualization of a bibliometric analysis using an overlay visualization of keyword co-occurrence.

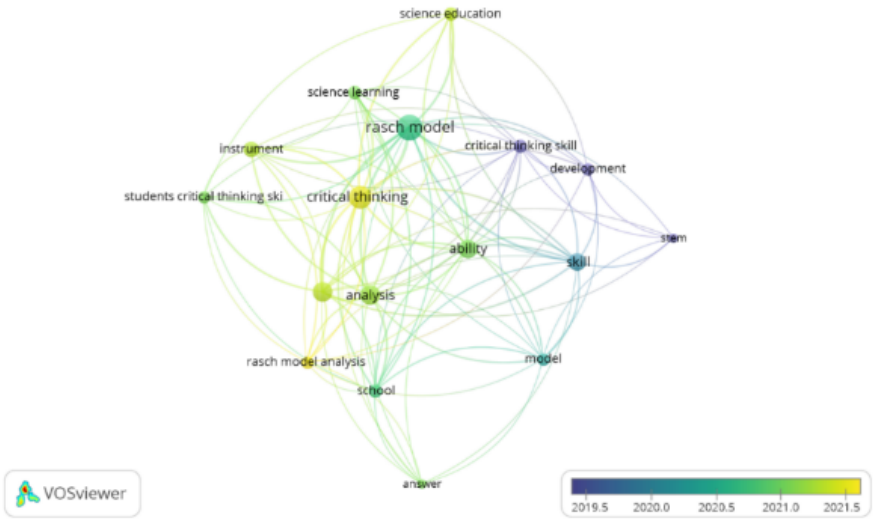


Figure 2. Bibliometric analysis using VOS viewer

A bibliometric analysis of 19 selected articles was conducted using VOSviewer with a co-occurrence approach based on author keywords. After keyword cleaning and term normalization, keywords that met the minimum occurrence threshold of ≥2 were visualized in a network of relationships. Based on the overlay visualization, the research network reveals connections among several main themes, namely the Rasch model, critical thinking, ability, analysis, instrument, skill, science education, development, and STEM. Structurally, the Rasch model and critical thinking occupy relatively central positions in the network and are connected to the themes of instrument development and evaluation. These connections indicate that, in the analyzed corpus of studies, the use of the Rasch model is primarily related to the measurement and analysis of critical thinking skills as well as the evaluation of instrument quality. The relationship between the keywords "instrument," "analysis," and "Rasch model" also reveals a close connection between instrument development and the psychometric evaluation process, while the connections to "science education," "skill," and "development" indicate the application of the Rasch model in the broader context of science education. To enhance the transparency of the analysis, these results should be supplemented with a table reporting the occurrences, number of links, total link strength, and cluster composition for each keyword.

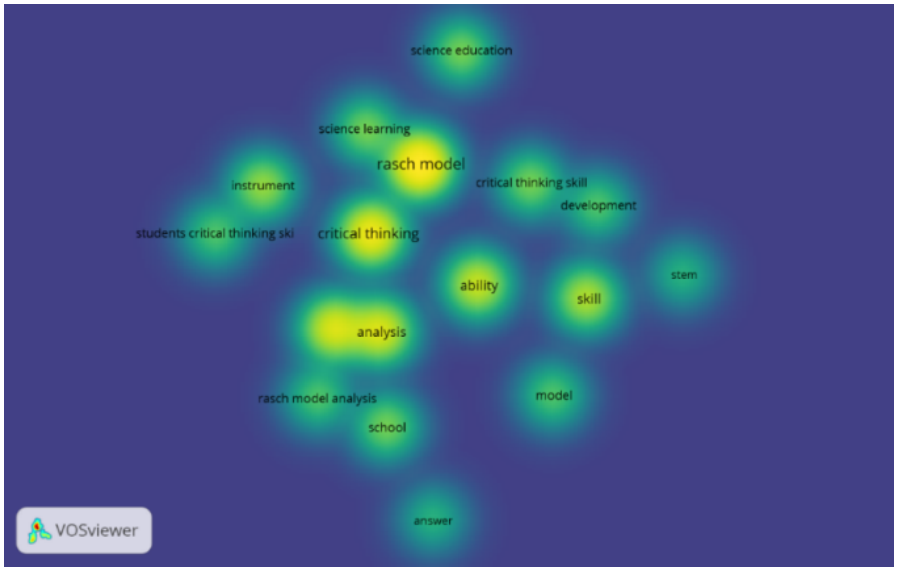


Figure 3. Density visualization menggunakan VOSviewer

Based on the temporal dimension in the overlay visualization, the network colors indicate differences in the average year of publication within the range of approximately 2019.5–2021.5. Keywords with a more yellow hue, such as "critical thinking," "Rasch model," "analysis," and "ability," indicate a relatively more recent emergence compared to some other keywords with darker colors. This finding is supported by the density visualization, which shows the highest density for "Rasch model," "critical thinking," and "analysis." Conversely, keywords such as "STEM," "development," "instrument," and "students' critical thinking skills" exhibit lower density. This pattern suggests that, across the 19 studies analyzed, research focus remains concentrated on the application of the Rasch Model for the analysis and evaluation of critical thinking skills, while its integration with learning contexts such as STEM and the development of more context-based instruments remains relatively limited. Therefore, STEM, development, and context-based instrument development can be considered research opportunities that warrant further investigation. However, since the analysis is based on only 19 articles, these results should be interpreted as thematic trends within the corpus of reviewed studies, rather than as a general representation or global trend in Rasch model research in science education.

A synthesis of 19 studies indicates a consistent trend toward the use of the Rasch model in science education research. The dichotomous Rasch model and the rating scale model are the most frequently used approaches, primarily applied in the validation and development of written assessment instruments oriented toward Higher Order Thinking Skills (HOTS). The most frequently used software is Winsteps due to the ease of analysis and interpretation of results. The sample size of the studies generally ranged from 100 to 300 respondents, which was deemed sufficient for the stability of parameter estimates. The most frequently reported instrument quality indicators included item fit statistics, person and item reliability, the separation index, and unidimensionality tests. However, a common weakness identified is the limited use of Differential Item Functioning (DIF) analysis and the dominance of written-test-based assessments, resulting in critical thinking measurement that remains focused on individual cognitive performance. Therefore, the next direction for development is to create authentic, problem-solving-based instruments analyzed using the Rasch Model to support the enhancement of 21st-century skills in a more contextual and holistic manner.

CONCLUSION

Based on the results of the Systematic Literature Review, the use of the Rasch Model in science education is still dominated by written tests to measure critical thinking skills and Higher-Order Thinking Skills (HOTS), with its primary application being the psychometric analysis of assessment instruments. Density visualization analysis shows that research focus is concentrated on the themes of the Rasch model, critical thinking, and analysis, while integration with learning contexts, STEM, and authentic assessment remains limited. Theoretically, future research needs to expand the application of the Rasch Model to more authentic assessment contexts, while practically, it is necessary to develop test items that are contextual, based on 21st-century learning, and maintain strong psychometric quality.

DECLARATION OF GENERATIVE AI

During the preparation of this manuscript, the author used ChatGPT (OpenAI) to assist with language editing, grammar and spelling checks, manuscript structure organization, and refinement of the presentation of research results. All research decisions, the article selection process, data extraction, analysis, interpretation of results, and formulation of conclusions were carried out and verified by the author. After using this tool, the author reviewed and re-edited all generated output as needed and assumes full responsibility for the accuracy, originality, interpretation, and entire content of the published manuscript.

CONFLICTING INTEREST STATEMENT

The authors declare that they have no known financial interests or personal relationships with any individuals or organizations that could influence or potentially introduce bias into the research reported in this article.

FUNDING INFORMATION STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORSHIP CONTRIBUTION STATEMENT

The first author contributed to the overall conceptualization, methodology, and conduct of the research, as well as to data collection and analysis, interpretation of results, and the drafting and revision of the manuscript. The second author contributed by providing input, supervising the work, and reviewing the manuscript and offering suggestions. The third author contributed by providing academic input.

REFERENCES

Adi, N. R. M. (2025). Analisis Validitas dan Reliabilitas Menggunakan Model Rasch untuk Mengukur Kualitas Instrumen Literasi Media (Validity and Reliability Analysis Using the Rasch Model to Measure Quality of Media Literacy Instruments). *Briliant: Jurnal Riset dan Konseptual, 10*(2), 323–336. https://doi.org/10.28926/briliant.v10i2.2048

Andryan Fitraynsyah, M., Hilmiyati, F., & Sultan Maulana Hasanuddin, U. (2024). Peran Tingkat Kesukaran dan Daya Pembeda dalam Analisis Butir Tes: Kajian Literatur untuk Pendidikan Menengah. *Jrep :Jurnal Riset Dan Evaluasi Pendidikan*, *1*(4), 252–262. https://doi.org/10.23887/jere.v8i2.69259

Costa, S. L. R., & Broietti, F. C. D. (2021). Scientific Practices in Science Education: A Systematic Review of Research Characteristics and Trends. *Ensino & Pesquisa*, *19*(3), 203-219. https://doi.org/10.33871/23594381.2021.19.3.203-219

Ekawati, E. Y., Azzahra, S. A., & Fithriyah, H. (2024). Construction of Critical Reasoning Skills Assessment Instruments as Diagnostic Assessments in Physics Learning with Polytomus Scoring. *Journal of Education Research and Evaluation, 8*(2), 338–349. https://doi.org/10.23887/jere.v8i2.69259

Fauziah, N., & Sukmawati, W. (2023). Stacking Analysis of Higher Thinking Skills of Class V Elementary School Students on the Material of Movement Organs Using the RADEC Model. *Jurnal Penelitian Pendidikan IPA*, (Quer 9(7), 9(7), 5263–527. http://dx.doi.org/10.26877/mpp.v17i1.15388

Hamdu, G., Fuadi, F. N., Yulianto, A., & Akhirani, Y. S. (2020). Items Quality Analysis Using rasch Model to Measure Elementary School Students' Critical Thinking Skill on Stem Learning. *Jurnal Pendidikan Indonesia,* 9(1), 61–74. https://doi.org/10.23887/jpi-undiksha.v9i1.20884

Harta, J., Rasuh, N. T., & Seriang, A. (2020). Using HOTS-Based Chemistry National Exam Questions to Map the Analytical Abilities of Senior High School Students. *Journal of Science Learning*, *3*(3), 143–148. https://doi.org/10.17509/jsl.v3i3.22387

Istiyono, E., Dwandaru, W. S. B., Lede, Y. A., Rahayu, F., & Nadapdap., A. (2019). Developing IRT-Based Physics Critical Thinking Skill Test: A CAT to Answer 21st Century Challenge. *International Journal of Instruction,* 12(4), 267–280. https://doi.org/10.29333/iji.2019.12417a

Kumar, J., Sharma, V., & Mishra, S. (2025). Assessing the Quality of Review Corpus: A Key to Rigorous Systematic Literature Review. *South Asian Journal of Business and Management Cases*, *14*(2), 79-86. https://doi.org/10.1177/22779779251362291

Laliyo, L. A. R. (2023). Evaluating Students' Ability in Constructing Scientific Explanations on Chemical Phenomena. *Eurasia Journal of Mathematics Science and Technology Education*, *19*(9). https://doi.org/10.29333/ejmste/13524

Maryani, Y., Prasetyorini, P., & Permana, I. (2021). Critical Thinking Skills of Junior High School Students in Science Learning. *Scientiae Educatia: Jurnal Pendidikan Sains*, *10*(1), 20-28. 10.24235/sc.educatia.v10i1.8034

Misbah, M., Mahtari, S., Wati, M., & Harto, M. (2018). Analysis of Students' Critical Thinking Skills in Dynamic Electrical Material. *Kasuari: Physics Education Journal, 1*(2), 103–110. https://doi.org/10.37891/kpej.v1i2.19.

Mulyanti, S., Sukmawati, W., & Tarkin, N. E. H. (2022). Development of Items in Acid-Base Identification Experiments Using Natural Materials: Validity Test with Rasch Model Analysis. *Phenomenon: Jurnal Pendidikan MIPA, 12*(1), 17–30. https://doi.org/10.21580/phen.2022.12.1.10703

Novriyanti, E., & Arthur, R. (2024). Analisis Kualitas Butir Soal Ujian Tengah Semester Biologi Umum Menggunakan Model Rasch. *JagoMIPA: Jurnal Pendidikan Matematika Dan IPA*, *4*(4), 718–733. https://proceedings.unnes.ac.id/snpptk/article/view/3291

Nuryaningsih, S., & Dewi, R. N. (2024). Model Pembelajaran Problem Based Learning untuk Meningkatkan Critical Thinking Siswa Kelas 9D SMP Negeri 41 Kota Semarang pada Mata Pelajaran IPA. *In Prosiding Seminar Nasional Pendidikan Dan Penelitian Tindakan Kelas*, (1), 1528–1535. https://proceedings.unnes.ac.id/snpptk/article/view/3291

OECD. (2023b). PISA 2022 Results (Volume I). The State of Learning and Equity in Education, PISA

OECD Publishing, Paris, https://doi.org/10.1787/53f23881-en.

Pratiwi, E., & Sugiharto, B. (2025). Analisis Validitas Butir dan Tepat Target Ukur Instrumen Biologi Kelas XII: Aplikasi Pemodelan Rasch. *Jurnal Pendidikan Biologi* , *16* (2), 155-167. https://doi.org/10.17977/um052v16i2p155-167

Prihatni, Y., Kumaidi, & Mundilarto. (2016). Pengembangan Instrumen Diagnostik Kognitif pada Mata Pelajaran IPA di SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*, *20*(1), 111-125. https://doi.org/10.21831/pep.v20i1.7524

Purnami, W., Ashadi, A., Suranto, S., Suyanti, R. D., & Mocerino, M. (2021). Investigation of Person Ability and Item Fit Instruments of Eco Critical Thinking Skills in Basic Science Concept Materials for Elementary Pre-service Teachers. *Jurnal Pendidikan IPA Indonesia, 10*(1), 127–137. https://doi.org/10.15294/jpii.v10i1.25239

Putri, F. S., & Istiyono, E. (2017). The Development of Performance Assessment of STEM-Based Critical Thinking Skill in the High School Physics Lessons. *International Journal of Environmental and Science Education, 12*(5), 1269–1281. https://eric.ed.gov/?id=EJ1278188.

Putri, A. N. L., Sutarto, S., & Wahyuni, D. (2024). Meta Analisis Kemampuan Berpikir Kritis Siswa SMP dalam Pembelajaran IPA. *Jurnal Penelitian Pembelajaran Fisika, 15(1), 43-48.* https://doi.org/10.26877/jp2f.v15i1.15580

Rachmadtullah, R. (2020). Critical Thinking Instrument Test (CTIT): Developing and Analyzing Sundanese Students' Critical Thinking Skills on Chysics Concepts Using Rasch Analysis. *International Journal of Psychosocial Rehabilitation, 24*(8). https://doi.org/10.37200/IJPR/V24I8/PR281423

Risdianto, E., Dinissjah, M. J., Nirwana, N., & Putri, D. H. (2021). Analysis of Student Responses Toward Ethnoscience Based Direct Instruction Learning Model in Learning Physics Applying Rasch Model Approach. *Journal of Physics: Conference Series, 1731*(1), 012081. https://doi.org/10.1088/1742-6596/1731/1/012081

Riyadi, A., Susongko, P., & Munadi, M. (2024). Model Asesmen Literasi Sains pada Peserta Didik Sekolah Dasar dengan Aplikasi Model Rasch. *Journal of Education Research*, *5*(3), 3044–3054. https://doi.org/10.37985/jer.v5i3.1411

Rosidah, C. T., Pramulia, P., & Susiloningsih, W. (2021). Analisis Kesiapan Guru Mengimplementasikan Asesmen. *Jurnal Pendidikan Dasar*, *12*(1), 87–103. https://doi.org/10.21009/jpd.v12i01.21159

Sabekti, A. W., & Khoirunnisa, F. (2018). Penggunaan Rasch Model untuk Mengembangkan Instrumen Pengukuran Kemampuan Berfikir Kritis Siswa pada Topik Ikatan Kimia. *Jurnal Zarah, 6*(2), 68–75. https://doi.org/10.31629/zarah.v6i2.724

Sadhu, S., & Laksono, E. W. (2018). Development and Validation of an Integrated Assessment for Measuring Critical Thinking and Chemical Literacy in Chemical Equilibrium. *International Journal of Instruction, 11*(3), 557–572. https://doi.org/10.12973/iji.2018.11338a

Sujatmika, S. (2025). Applying the Rasch Model to Measure Students' Critical Thinking Skills on the Science Topic of the Human Circulatory System. *Eurasia Journal of Mathematics Science and Technology Education*, *21*(4). https://doi.org/10.29333/ejmste/16221

Susongko, P., Arfiani, Y., & Kusuma, M. (2021). Determination of Gender Differential Item Functioning in Tegal Students' Scientific Literacy Skills with Integrated Science (SLiSIS) Test Using Rasch Model. *Jurnal Pendidikan IPA Indonesia, 10*(2), 270–281. https://doi.org/10.15294/jpii.v10i2.26775

Syafitri, E., Armanto, D., & Rahmadani, E. (2021). Aksiologi Kemampuan Berpikir Kritis. *Journal of Science and Social Research,* 4(3), 320–325. http://jurnal.goretanpena.com/index.php/JSSR

Umami, R., Rusdi, M., & Kamid, K. (2021). Pengembangan Instrumen Tes untuk Mengukur Higher Order Thinking Skills (HOTS) Berorientasi Programme for International Student Asessment (PISA). *JP3M, 7*(1), 57–68. https://doi.org/10.37058/jp3m.v7i1.2069

Wardana, II, Saptono, S., & Dewi, NR (2025). Suplemen Elektronik Etnosains untuk Berpikir Kritis dan Hasil Kognitif. *Jurnal Pendidikan Sains Unnes* , *14* (2), 236-244. https://doi.org/10.15294/usej.v14i2.28898

Widiyawati, Y., Nurwahidah, I., Sari, D. S., Budiyanto, C. W. (2021). The 21 st Century Science Learning: HOTS and Digital Literacy Among Junior High School Students in Semarang, Indonesia. *Journal of Physics: Conference Series, 1842*(1), 012081. https://doi.org/10.1088/1742-6596/1842/1/012081

Widodo, W., & Chotimah, C. (2023). Adaptasi dan Analisis Psikometri Skala Kompetensi Multikultural Calon Guru Menggunakan Pemodelan Rasch. *Jurnal Pendidikan Dan Kebudayaan*, *8*(2), 153–172. https://doi.org/10.24832/jpnk.v8i2.4228

Yokhebed, Y., Karmadi, R. M. D., & Nastiti, L. R. (2025). Validity and Reliability Analysis of a Socioscientific Issues-Based Critical Thinking Self-Assessment Instrument Using the Rasch Model. *Jurnal Pendidikan Biologi Indonesia, 11*(1), 73–82. https://doi.org/10.22219/jpbi.v11i1.38902

Yulianto, A., Yuliyanto, A., Hamdu, G., Nur, L., Hamzah, N. (2021). Disclosure of Student Ability in Working on Higher-Order Thinking Skills Questions through Rasch Modeling. *Al Ibtida, 8*(1), 130–143. http://dx.doi.org/10.24235/al.ibtida.snj.v8i1.7865

Yunita, Y., & Dewi, N. R. (2020). The Development of Two-tier Multiple Choice Assessment Instrument to Measure Higher Order Thinking Skills (HOTS) of The Students on Excretion System Material. *Unnes Science Education Journal*, *9*(1), 30–39. https://journal.unnes.ac.id/sju/usej/article/view/36472.