



**Mapping the Intellectual Structure of Item Response Theory and Rasch Model
Applications in Physics Education Assessment: A Scopus-Based Bibliometric
Analysis**

**Haeruddin^{1*}, Delthawati Isti Ratnaningtyas¹, Muhammad Jarnawi¹,
Syamsuriwal¹, and Rudi Santoso¹**

Physics Education Study Program, Faculty of Teacher Training and Education,
Universitas Tadulako, Palu, Indonesia

*haeruddin@untad.ac.id

Article Info	Abstract
Article history: Received: March 13, 2026 Accepted: May 22, 2026 Published: May 26, 2026 Keywords: Bibliometric analysis Item response theory Physics education assessment Rasch model Science mapping This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. © 2026 Journal of Mathematics, Science, and Computer Education	Assessment in physics education requires valid and reliable measurement instruments capable of accurately representing students' cognitive abilities, including conceptual understanding, scientific reasoning, and problem-solving skills. Over the past decades, instrument validation has shifted from Classical Test Theory to modern probabilistic measurement models, particularly Item Response Theory (IRT) and the Rasch model. Despite the growing application of these models in physics education research, the expanding literature remains thematically and methodologically fragmented, limiting a comprehensive understanding of the field's intellectual development. Therefore, this study aims to map the intellectual structure and research trends of IRT and Rasch model applications in physics education assessment. This study employed a bibliometric research design using 243 Scopus-indexed publications published between 2001 and 2026. Bibliographic metadata were retrieved using Publish or Perish and analyzed using the Bibliometrix package in RStudio and VOSviewer to examine publication trends, influential sources, collaboration networks, and keyword co-occurrence patterns. The results reveal substantial growth in publications, increasing by more than 1,150% over the past two decades. Five major thematic clusters were identified: measurement frameworks, psychometric modeling, validity and fairness, instrument development and technology, and learning progression. Emerging themes such as computational thinking, digital instrument development, and multidimensional assessment indicate a shift toward competency-based and technology-supported assessment practices. Overall, the findings reveal the intellectual structure and research trends of IRT and Rasch model applications in physics education assessment, demonstrating their growing role as central methodological frameworks in contemporary assessment research. These findings also provide directions for future research, particularly in developing multidimensional Rasch-based instruments and technology-enhanced assessment systems in physics education.
To cite this article: Haeruddin, H., Ratnaningtyas, D. I., Jarnawi, M., Syamsuriwal, S., & Santoso, R. (2026). Mapping the intellectual structure of item response theory and rasch model applications in physics education assessment: A scopus-based bibliometric analysis. <i>Journal of Mathematics, Science, and Computer Education</i>, 6(1), 215-237.	

INTRODUCTION

Assessment of learning outcomes in physics education requires measurement instruments that are not only statistically reliable but also capable of objectively and comprehensively representing students' cognitive abilities (Adams, 2005; De Ayala, 2022). This requirement is particularly crucial because physics learning demands simultaneous assessment of conceptual understanding, scientific reasoning, and problem-solving skills (Diya & Putranta, 2025; Docktor & Mestre, 2014). In this context, the paradigm of educational instrument validation has gradually shifted from Classical Test Theory (CTT) toward modern probabilistic measurement approaches (Hambleton et al., 1991; Hauenstein & Embretson, 2018). Among these approaches, Item Response Theory (IRT) and the Rasch model have emerged as dominant frameworks due to their ability to estimate item characteristics and learner abilities independently on a unified measurement scale (Bond et al., 2020; Embretson & Reise, 2013).

Boone and Scantlebury (2006) and Riyadi et al., (2024) emphasized that Rasch analysis plays an important role in science education research by transforming ordinal scores into interval-scale measurements, thereby enabling more meaningful interpretations. Similarly, Morris et al. (2006) and Amelia and Kriswantoro (2017) highlighted that item response curves provide deeper insights into how individual test items function in measuring students' conceptual understanding.

As physics education increasingly emphasizes higher-order thinking and inquiry-based learning, challenges in instrument validation have become more complex. The application of IRT and Rasch models has extended beyond conventional paper-based testing into various assessment contexts, including performance assessments, diagnostic tests, and technology-mediated learning environments. Scantlebury et al. (2001) and Zayrin et al. (2025) argued that well-validated instruments are essential for monitoring instructional effectiveness and supporting curriculum reform in science education. Furthermore, advances in multidimensional measurement models proposed by Rost and Carstensen (2002) and further developed by Liu and Collard (2005) indicate that modern assessment frameworks must accommodate multiple interrelated competencies within science and engineering education (Intasoi et al., 2020; Qualters et al., 2008).

In the era of digital transformation, the relevance of IRT and Rasch model applications in physics education has grown substantially. Recent studies demonstrate their use in validating instruments aligned with international assessment frameworks, including measures of science identity and scientific literacy associated with PISA-oriented competencies (Mu'aziyah et al., 2025; Pursitasari et al., 2025). In addition, the integration of interactive digital learning environments such as PhET simulations requires rigorous validation procedures to accurately measure complex learning constructs, particularly complex problem-solving skills in physics contexts (Juliyanto et al., 2026; Nurfalalah et al., 2025). While the rapid expansion of research in this domain has produced a substantial body of literature, it has also resulted in increasing fragmentation, making comprehensive synthesis more difficult.

Despite the growing number of publications, existing studies tend to focus primarily on the application of Rasch or IRT models to specific instruments or localized educational contexts (Boone, 2017; Engelhard, 2013). Comprehensive investigations that examine the global intellectual structure, collaboration patterns, and thematic evolution of this field remain limited (Donthu et al., 2021; Zupic & Čater, 2015). In particular, bibliometric studies that explicitly explore the intersection between IRT–Rasch methodologies and instrument validation in physics education remain scarce (Aria & Cuccurullo, 2017; Cobo et al., 2011). This gap makes it difficult for researchers to identify influential journals, key contributors,

and emerging research themes shaping the future direction of the field (Moral-Muñoz et al., 2020). Without a systematic mapping of the scientific landscape, new researchers may encounter challenges in identifying promising research directions and potential gaps with high publication potential (Donthu et al., 2021; Zupic & Čater, 2015).

This growing complexity coincides with increasing demands in physics learning for conceptual understanding, multiple representations, and higher-order problem-solving skills (Docktor & Mestre, 2014; Nieminen et al., 2010). Rasch models and IRT approaches are no longer limited to the development of conventional multiple-choice tests but have also been applied in validating various innovative assessment forms, from concept inventories and measures of scientific attitudes to technology-based assessment systems (Boone & Staver, 2020; Planinic et al., 2019). These developments highlight that instrument validation is not merely a technical statistical issue but also an integral component of science education reform aimed at measuring multidimensional competencies (Liu, 2020; National Research Council, 2012).

In recent years, bibliometric studies have increasingly been used to map global research trends in physics education. Several studies have explored specific topics, such as bibliometric analyses of multiple representations in the physics learning (Addawiyah et al., 2024; Krisnaningsih et al., 2021), gender equity in the physics education (Bezen & Derman, 2025; Doucette & Singh, 2024), mathematical thinking in science education (Evendi, 2022; Kotsis, 2025), general trends in contemporary physics education research (Bancong et al., 2021; Hidayana Nassiri et al., 2025), and the implementation of problem-based learning in physics education using Scopus-based datasets (Nurjanah, Iqbal, et al., 2024; Nurjanah, Sultan, et al., 2024). However, no bibliometric study has specifically mapped the intersection between IRT–Rasch methodologies and instrument validation within physics education as a distinct research domain.

Therefore, this study conducts a comprehensive bibliometric analysis to map publication trends, influential sources, research collaboration networks, and thematic evolution related to the application of IRT and Rasch models in physics education instrument validation over the past two decades (2000–2026). Bibliometric analysis was chosen because it has proven effective in systematically and quantitatively revealing the intellectual structure of a scientific field (Donthu et al., 2021; Zupic & Čater, 2015).

Unlike previous studies that examined Rasch or IRT applications in isolated instructional contexts (Boone, 2017; Planinic et al., 2019), this study constructs a systematically curated bibliometric dataset from Scopus that represents the intersection between Rasch measurement, IRT-based validation, and science education related to physics—a research domain that has not yet been comprehensively mapped.

The novelty of this study lies in two main aspects. First, it identifies publication growth and influential contributors within this research domain. Second, it uncovers emerging thematic convergence between psychometric measurement models, concept-based physics assessment, and technology-mediated learning environments (Aria & Cuccurullo, 2017; van Eck & Waltman, 2009). Through science mapping techniques, this study illustrates how research on instrument validation in physics education has evolved from traditional measurement approaches toward digitally mediated multidimensional assessment practices (Cobo et al., 2011; Liu, 2020). Consequently, the findings are expected to provide a more precise strategic reference compared with general bibliometric analyses in education, while offering guidance for future instrument development and assessment innovation in the physics education research (Haeruddin et al., 2020; Moral-Muñoz et al., 2020).

Therefore, this study aims to map the intellectual structure and research trends of IRT and Rasch model applications in physics education assessment. Through bibliometric analysis of Scopus-indexed publications, this study identifies publication trends, influential

sources and contributors, collaboration networks, and thematic evolution within this field. The findings are expected to provide a comprehensive overview of the scientific landscape of IRT and Rasch research in physics education and reveal potential directions for future research.

METHODS

Research Design

This study employed a bibliometric research design to systematically examine the development of research concerning the application of Item Response Theory (IRT) and the Rasch model in instrument validation within physics education settings. Bibliometric analysis enables quantitative exploration of scientific production, intellectual structure, and thematic evolution based on large-scale bibliographic metadata (Donthu et al., 2021). The methodological procedure followed a modified PRISMA-based framework adapted for bibliometric studies to ensure transparency, reproducibility, and systematic dataset construction (Moher et al., 2009; Page et al., 2021).

Subject of Research or Population and Sample

The population of this study comprised all scholarly publications indexed in the Scopus database related to IRT and Rasch model applications in physics and science education instrument validation. Scopus was selected as the primary data source due to its broad multidisciplinary coverage and high indexing standards, making it widely recognized as a reliable source for bibliometric investigations (Mongeon & Paul-Hus, 2015).

The sample consisted of 243 publications retrieved after systematic filtering from an initial pool of 344 records identified through a structured search query. Publications were included if they were published between 2000 and 2026, addressed IRT or Rasch measurement models, and were contextually relevant to assessment or instrument validation in physics, science education, or STEM domains. Publications with incomplete metadata or irrelevant subject contexts were excluded from the final dataset.

Instruments

Bibliographic metadata was retrieved from the Scopus database using Publish or Perish (PoP) software version 8 (Harzing, 2007) via its Scopus search interface. Publish or Perish facilitated structured retrieval and batch extraction of publication metadata. The search was applied to title, abstract, and keywords fields to ensure topical relevance to measurement models and educational assessment contexts.

The search string used was as follows:

```
TITLE-ABS-KEY(  
  "rasch model" OR "rasch measurement" OR "item response theory"  
  OR "partial credit model" OR "rating scale model"  
  OR "many facet rasch" OR "irt model"  
)  
AND TITLE-ABS-KEY(  
  assessment OR validation OR "instrument validation"  
  OR "test development" OR psychometric* OR measurement  
  OR reliability OR validity  
)  
AND TITLE-ABS-KEY(  
  "physics education" OR physics OR "science education"  
  OR STEM OR "concept inventory"  
)
```

Retrieved records were exported in CSV format containing bibliographic metadata including author names, titles, abstracts, author keywords, publication sources, and citation

information. To enhance metadata completeness, the filtered dataset was subsequently imported into the Mendeley reference manager for automatic synchronization and enrichment of missing bibliographic fields, including DOI identifiers, journal information, author metadata, and publication details.

Data Analysis Technique

Bibliometric analysis was performed using two complementary analytical platforms. Performance analysis and bibliographic data processing were conducted using the Bibliometrix package (version 4.0) implemented within the RStudio environment (Aria & Cuccurullo, 2017). Science mapping, including keyword co-occurrence network visualization and thematic clustering, was performed using VOSviewer (version 1.6.20) (van Eck & Waltman, 2009).

The analytical procedure comprised the following sequential steps: (1) conversion of metadata into Bibliometrix-compatible format; (2) descriptive analysis of annual scientific production; (3) identification of influential journals, authors, and documents based on citation metrics; (4) keyword co-occurrence network analysis; and (5) visualization of thematic clusters representing research development trends. Keyword normalization was applied using a thesaurus-based approach to reduce semantic duplication during visualization, following the procedures recommended by van Eck & Waltman (2009).

Research Procedures

The dataset construction followed four main stages adapted from the PRISMA workflow (Page et al., 2021), as illustrated in Figure 1.

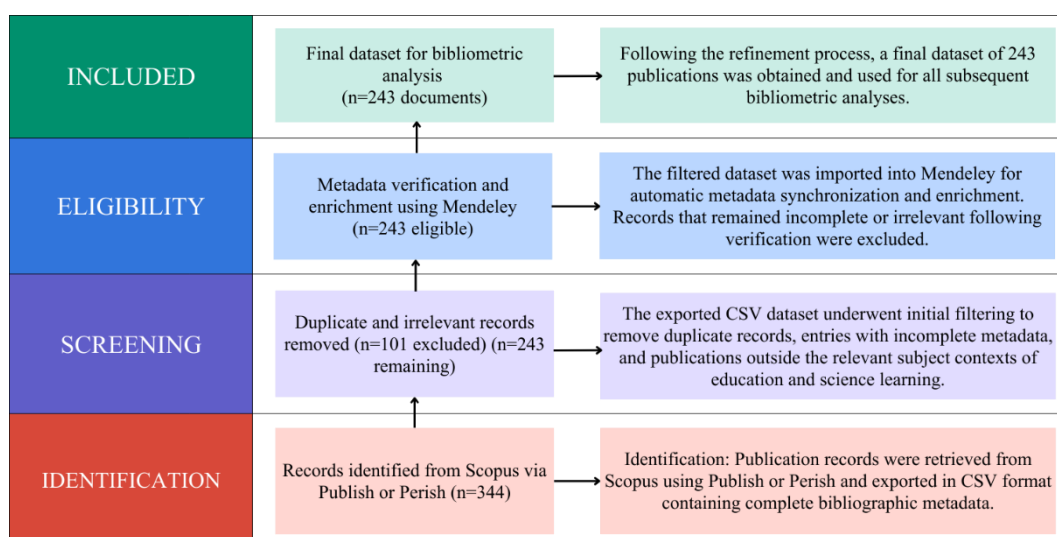


Figure 1 PRISMA-based dataset construction flowchart

The use of Scopus-indexed publications ensured high-quality, peer-reviewed sources. The structured filtering process and metadata verification enhanced dataset accuracy and the reproducibility of the bibliometric analysis (Mongeon & Paul-Hus, 2015).

RESULTS AND DISCUSSION

This section presents the cleaned bibliometric results derived from the final dataset obtained through the systematic data collection and screening process. The analysis focuses on publication trends, influential sources, collaboration patterns, and thematic development within the research domain.

Publication Trends and Scientific Growth

Table 1 presents the main bibliometric characteristics of the dataset. A total of 243 publications from 122 journal sources were identified, contributed by 666 authors with an average of 3.23 authors per document. The dominance of multi-authored papers (80.7%) indicates a strong collaborative culture in this research domain (Aria & Cuccurullo, 2017; Zupic & Čater, 2015).

The most notable finding is the dramatic growth in publication output. Only 18 documents (7.4%) appeared during 2001–2010, compared to 225 documents (92.6%) in 2011–2026, yielding a growth rate of +1,150%. This growth may be partly explained by the emergence of digital instrument development and technology-supported assessment as identified in the keyword clusters of this study, alongside the broader global emphasis on evidence-based science education reform. This trend reflects the increasing recognition of IRT and Rasch models as essential tools for instrument validation in physics education, particularly driven by the growing demand for evidence-based assessment and psychometrically robust measurement frameworks in the science education reform (Boone et al., 2016; Planinic et al., 2019).

Table 1 Main information about dataset

Description	Results
Timespan	2001–2026
Total Documents	243
Total Sources (Journals)	122
Total Authors	666
Total Unique Keywords	705
Average Authors per Document	3.23
Single-authored Documents	47 (19.3%)
Multi-authored Documents	196 (80.7%)
Documents with DOI	233
Documents 2001–2010	18
Documents 2011–2026	225
Growth Rate	+1,150%

Publication Growth Trends (2001–2026) are shown in Figure 2.

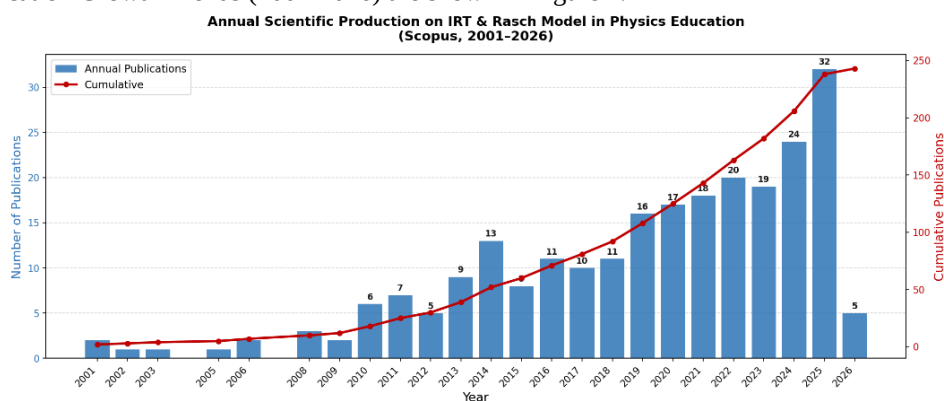


Figure 2 Growth Trend of Publications (2001–2026)

Figure 2 displays the annual scientific production from 2001 to 2026. Publication output remained negligible during 2001–2010 (18 documents total), before accelerating sharply from 2011 onward, peaking at 32 publications in 2025 with a cumulative total of 243 documents. This growth pattern reflects the increasing adoption of IRT and Rasch methodologies in physics education assessment, driven by expanding demands for evidence-based measurement and psychometrically validated instruments (Boone et al., 2016; Donthu et al., 2021; Planinic et al., 2019). The partial count for 2026 (5 documents) reflects data retrieved up to March 2026.

Influential Sources

Table 2 presents the 15 most productive journals publishing research on IRT and Rasch model applications in physics education. Physical Review Physics Education Research leads with 15 articles, which is unsurprising given its status as the most prestigious peer-reviewed journal dedicated exclusively to physics education research and published by the American Physical Society. Its dominance reflects both the journal's high academic standing and its role as the primary outlet for methodologically rigorous studies in the field. This is followed by the Journal of Baltic Science Education (11) and the International Journal of Science and Mathematics Education (10). The predominance of physics and science education journals indicates that this research domain is primarily situated within the science education community rather than within the broader field of general psychometrics (Donthu et al., 2021). Applying Bradford's Law, the top 10 journals collectively account for approximately 81 documents (~33% of the total corpus), constituting the core zone of publication activity. This core zone is significant for the field's development, as it represents the concentrated body of foundational knowledge that researchers must engage with to understand the intellectual basis of IRT and Rasch applications in physics education, functioning as the primary reference point for both established scholars and new entrants to the field. Notably, the presence of Jurnal Pendidikan IPA Indonesia (6 articles) and International Journal of Instruction (6 articles) among the top sources signals a growing contribution from Southeast Asian scholars, particularly from Indonesia, reflecting the expanding global reach of IRT–Rasch research in the physics education (Zupic & Čater, 2015).

Table 2 Top 15 most productive journals

Rank	Journal	Articles
1	Physical Review Physics Education Research	15
2	Journal of Baltic Science Education	11
3	International Journal of Science and Mathematics Education	10
4	International Journal of Science Education	9
5	Journal of Research in Science Teaching	8
6	Science Education	7
7	Jurnal Pendidikan IPA Indonesia	6
8	Journal of Science Education and Technology	6
9	International Journal of Instruction	6
10	Research in Science Education	4
11	School Science and Mathematics	4
12	Journal of Applied Measurement	4
13	Eurasian Journal of Mathematics, Science and Technology Education	4
14	European Journal of Educational Research	4

Rank	Journal	Articles
15	Education and Information Technologies	3

Bradford's Law: 10 core journals (Zone 1) contribute ~81 documents (~1/3 total publications).

Key Contributors

Figure 3 presents the 15 most productive authors in research on IRT and Rasch model applications in physics education. Istiyono, Edi, ranks first with 10 publications, consistently focusing on the development of IRT-based instruments for measuring higher-order thinking skills and problem-solving abilities in the context of senior high school physics education in Indonesia (Novika et al., 2025). Dwandaru, Wipsar Sunu Brams (6 publications), frequently collaborates with Istiyono, reflecting a productive research partnership at Universitas Negeri Yogyakarta. This collaboration pattern is also consistent with the prominence of Jurnal Pendidikan IPA Indonesia as a publication outlet within this research domain.

Fulmer, Gavin W. (6 publications) represents a significant contribution from the United States, with research primarily focusing on the application of Rasch measurement for science assessment and curriculum alignment (Ding & Zhang, 2023). Boone, William (5 publications) is widely recognized as a leading figure in Rasch methodology for science education and has authored several foundational references in the field (Boone & Staver, 2020).

Other authors with three to four publications exhibit a relatively distributed authorship pattern, suggesting that the field is not dominated by a single research group. This pattern is typical of an emerging and multidisciplinary research domain in which contributions originate from diverse institutional and geographical contexts (Kadirhanogullari & Kose, 2024).

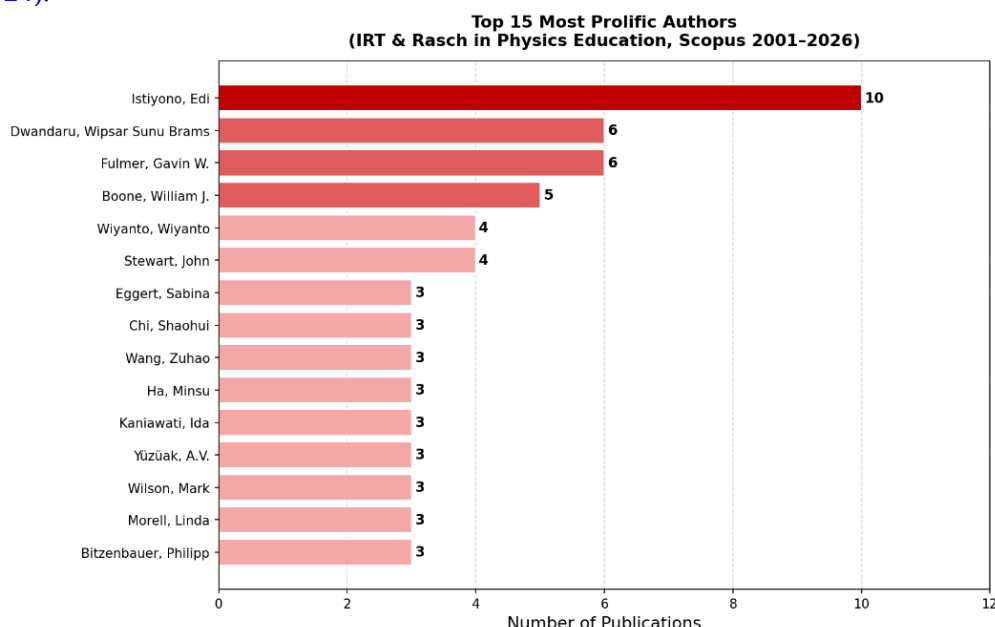


Figure 3 Top 15 most prolific authors

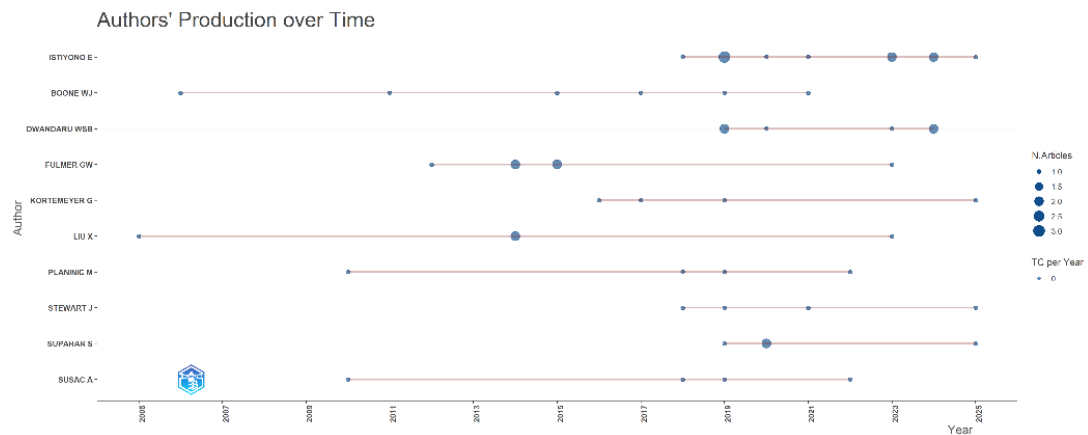


Figure 4 Authors' production over time

Figure 4 illustrates the publication trajectory of the ten most productive authors over time, with bubble size indicating the number of articles published each year. This visualization reveals distinct patterns of research engagement among individual authors. Boone, W. J. and Susac, A. exhibit the longest research span, with contributions traceable back to around 2005–2006, positioning them as early pioneers in the application of IRT–Rasch methodologies in physics and science education. Liu, X. and Planinic, M. also demonstrate early involvement beginning around 2011–2013, reflecting sustained long-term contributions to the field.

In contrast, Istiyono, E. and Dwandaru, W. S. B. display a surge in productivity beginning around 2019, with significantly larger bubble sizes indicating multiple publications per year, consistent with their leading positions shown in Figure 3. This increase coincides with the publication of influential methodological works advocating Rasch analysis in the physics education research (Planinic et al., 2019), which may have catalyzed broader scholarly engagement in the field. The subsequent COVID-19 pandemic further accelerated this trend by intensifying demand for technology-enhanced and remotely administered assessment instruments, consistent with the emergence of digital instrument development as a thematic cluster identified in this study. Fulmer, G. W. shows a productive mid-career period around 2013–2015 before reappearing with new publications in 2023. Meanwhile, Kortemeyer, G., Stewart, J., and Supahar, S. represent relatively recent entrants, with publications concentrated between 2019 and 2025, reflecting the broader acceleration of research activity in this domain in recent years.

Overall, the figure suggests that research activity in this field has evolved from a small group of early pioneers to a broader and more geographically diverse scholarly community, with Indonesian authors becoming increasingly prominent in the most recent period.

Intellectual Structure and Thematic Evolution

Figure 5 presents the 20 most frequently occurring author keywords within the dataset. The Rasch model (50 occurrences) and item response theory (44 occurrences) dominate the keyword distribution, confirming these two frameworks as the core methodological pillars of the research corpus (Boone & Staver, 2020; Planinic et al., 2019). Assessment (28) and science education (26) reflect the primary application contexts, while concept inventory (20) highlights the frequent use of conceptual measurement instruments as objects of validation within this research domain (Morris et al., 2012).

The emergence of keywords such as critical thinking, computational thinking, scientific reasoning, and learning progression indicates a thematic shift toward the measurement of higher-order competencies. This pattern suggests that the field is gradually evolving beyond

conventional psychometric validation toward broader twenty-first-century assessment agendas that emphasize complex cognitive skills and competency-based learning outcomes (Istiyono et al., 2019; Liu, 2020).

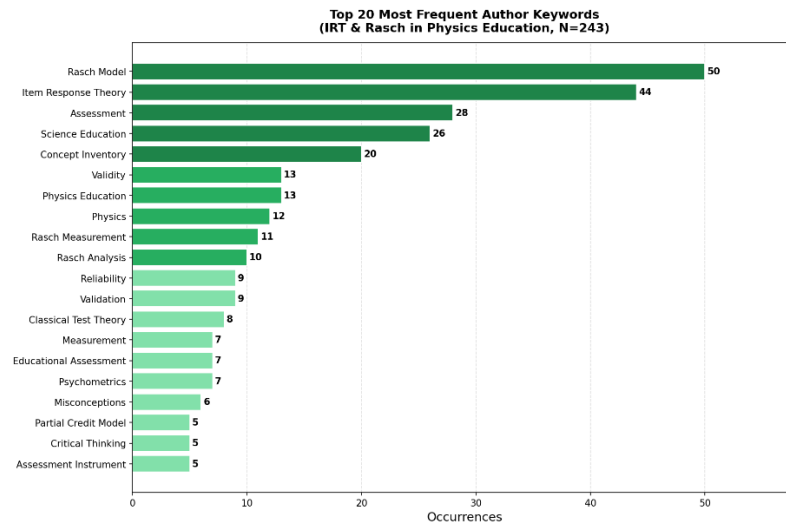


Figure 5 Top 20 most frequent author keywords

Figure 6 presents the keyword co-occurrence network generated using VOSviewer, which identified 29 keywords grouped into five thematic clusters based on a minimum occurrence threshold of five. These clusters represent the major thematic structures shaping the research landscape and provide insight into the key methodological and conceptual directions within the field.

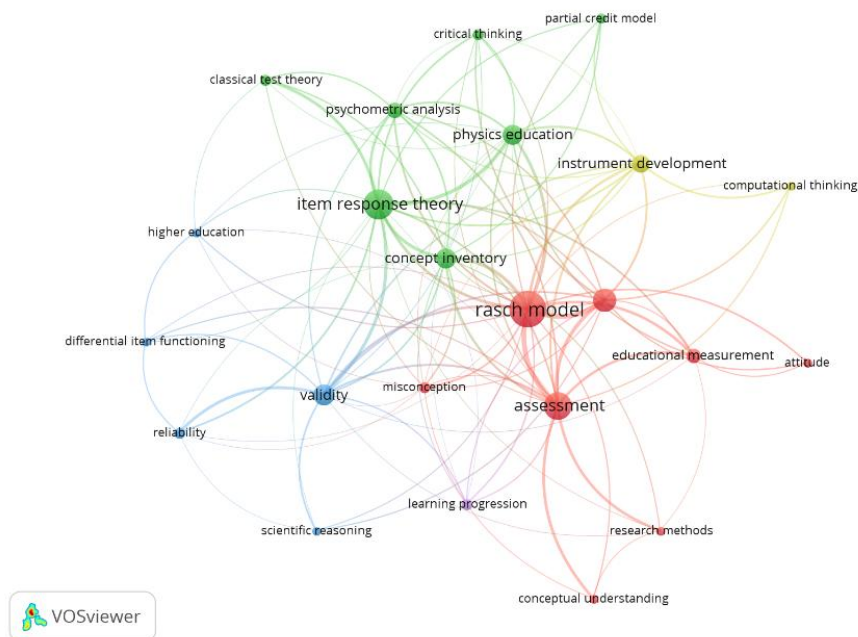


Figure 6 Keyword co-occurrence network

Cluster 1 (Red): Core Measurement and Assessment represents the largest cluster, centered on the Rasch model and assessment, which are strongly connected to science education, educational measurement, misconception, conceptual understanding, research methods, and attitude. This pattern confirms Rasch-based assessment as a dominant research focus within the corpus (Boone & Staver, 2020; Planinic et al., 2019).

Cluster 2 (Green): Psychometric Frameworks and IRT are centered on item response theory and concept inventory, linked with critical thinking, item response theory, partial credit model, physics education, and psychometric analysis. This cluster reflects the ongoing methodological discourse between classical and modern measurement approaches in the educational assessment (Liu, 2020; Morris et al., 2012).

Cluster 3 (Blue): Validity and Fairness groups together validity, reliability, differential item functioning, higher education, and scientific reasoning. The presence of these keywords highlights the sustained scholarly attention to psychometric quality and measurement fairness in science education instruments (Traxler et al., 2018).

Cluster 4 (Yellow): Instrument and Technology Development includes instrument development and computational thinking. This cluster reflects emerging interest in assessing 21st-century competencies within the IRT–Rasch measurement framework (Istiyono et al., 2019; Ramadhan et al., 2019).

Cluster 5 (Purple): Learning progression is represented by learning progression as a distinct thematic node. This cluster reflects a specialized research trajectory focused on mapping students' conceptual development in science through Rasch-based longitudinal assessments (Morell et al., 2017; Plummer & Maynard, 2014).

The strong interconnections among clusters, particularly between the Rasch *model*, *item response theory*, and *assessment*, indicate a high level of thematic coherence across the research corpus. At the same time, peripheral nodes suggest promising directions for future research within this evolving field (Aria & Cuccurullo, 2017; Cobo et al., 2011).

Figure 7 presents the overlay visualization of keyword co-occurrence generated using VOSviewer, where node colors represent the average publication year of documents associated with each keyword, ranging from dark blue (≤ 2018) to yellow-green (≥ 2021). This visualization enables the identification of the temporal evolution of research themes within the corpus.

Established themes (dark blue, avg. year ≤ 2018) include educational measurement, higher education, psychometric analysis, learning progression, validity, concept inventory, partial credit model, item response theory, scientific reasoning, physics education, and classical test theory. The early prominence of these themes confirms their role as foundational methodological pillars underpinning IRT–Rasch research in the physics education (Hambleton et al., 1991; Liu, 2020; Neumann et al., 2011).

Consolidation themes (blue-green, avg. year 2019–2020) include the Rasch model, misconception, instrument development, science education, assessment, reliability, and attitude. The convergence of the Rasch model and assessment as the most central nodes in this period confirms that Rasch measurement reached mainstream adoption as the dominant validation framework in physics education during the 2019–2020 (Boone & Staver, 2020; Planinic et al., 2019).

More recent and emerging themes (green-yellow, ≥ 2020 –2021) include computational thinking, critical thinking, conceptual understanding, and differential item functioning. Notably, computational thinking carries the most recent average publication year across the entire network (2023.33), positioning it as the most emergent research frontier—reflecting the field's responsiveness to 21st-century competency frameworks and technology-enhanced assessment (Istiyono et al., 2019; Zakwandi et al., 2024).

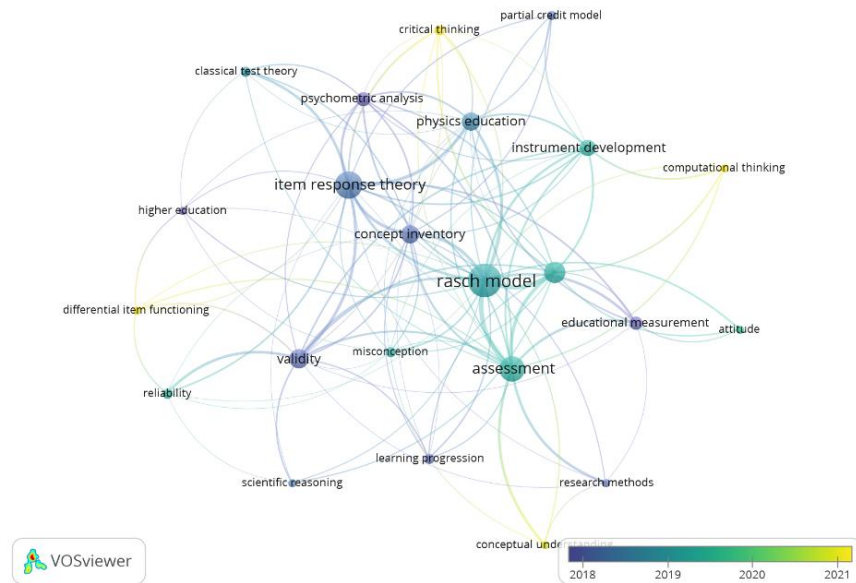


Figure 7 Keyword co-occurrence overlay visualization

Overall, the overlay visualization reveals a clear temporal trajectory from foundational psychometric validation toward competency-oriented and technology-integrated assessment, with computational thinking, critical thinking, and differential item functioning emerging as the most promising directions for future research (Cobo et al., 2011; Donthu et al., 2021).

Figure 8 presents the thematic map generated using the Bibliometrix package, positioning research themes into four quadrants based on centrality (degree of relevance) and density (degree of development), following the strategic diagram approach proposed by Cobo et al. (2011). Based on this mapping, research themes can be categorized into four main groups according to their level of development and their role within the research domain.

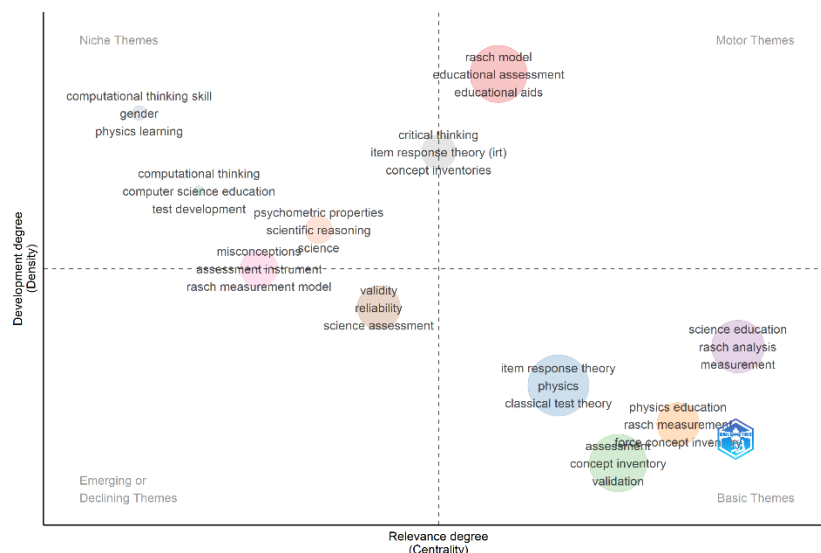


Figure 8 Thematic map generated through bibliometrix

Motor Themes (upper-right quadrant) represent themes that are both well-developed and highly central to the field. Rasch model, educational assessment, and educational aids occupy this quadrant, indicating that Rasch-based educational assessment constitutes one of the most mature and influential research streams in the corpus. These themes exhibit strong internal cohesion and maintain extensive connections with broader research areas (Boone & Staver, 2020; Donthu et al., 2021).

Basic themes (lower-right quadrant) are highly central but less internally developed, functioning as fundamental cross-cutting topics. This quadrant includes science education, Rasch analysis, measurement, physics education, Rasch measurement, force concept inventory, assessment, concept inventory, validation, item response theory, physics, and classical test theory. Collectively, these topics represent the essential methodological and contextual backbone upon which this research domain is built (Liu, 2020; Planinic et al., 2019).

Niche themes (upper-left quadrant) are well-developed internally but remain relatively peripheral to the main research stream. Keywords such as "computational thinking skill," "gender," "physics learning," "computational thinking," "computer science education," "test development," "psychometric properties," "scientific reasoning," and "misconceptions" fall into this category. These themes indicate specialized research trajectories with strong internal coherence but limited integration into the broader IRT–Rasch discourse, representing potential areas for interdisciplinary expansion (Istiyono et al., 2019; Traxler et al., 2018).

Emerging or Declining Themes (lower-left quadrant) include validity, reliability, science assessment, and the Rasch measurement model. Although these topics hold fundamental importance, their position in this quadrant suggests that they either represent emerging subthemes that have not yet gained strong momentum or topics that are gradually declining as they become absorbed into more consolidated research streams (Aria & Cuccurullo, 2017; Cobo et al., 2011).

Overall, the thematic map reveals a research field characterized by a solid and mature core centered on Rasch measurement and educational assessment, while simultaneously displaying dynamic peripheral themes related to competency-based and technology-enhanced assessment. This configuration highlights promising opportunities for future research at the intersection of psychometrics and twenty-first-century science education (Liu, 2020; Ramadhan et al., 2019).

Position of Physics Education within the Thematic Map

Based on the thematic map (Figure 8), physics education is in the Basic Themes quadrant, indicating high centrality but relatively low density. This positioning suggests that physics education is frequently associated with other themes in the corpus and plays a highly connected role within the research network, yet it has not developed as a deeply consolidated and internally cohesive research cluster. In this context, physics education functions primarily as a thematic bridge linking various research streams within the IRT–Rasch domain rather than as an independent and dominant research focus (Cobo et al., 2011).

Meanwhile, in the keyword co-occurrence network (Figure 6), physics education appears within Cluster 2 (Green) together with item response theory, concept inventory, and psychometric analysis. This configuration indicates that studies related to physics education in this corpus are more frequently associated with the development and validation of concept-based instruments using IRT methodologies rather than with broader pedagogical or curricular issues in the physics education (Liu, 2020; Planinic et al., 2019).

Most Influential Documents and Citations

Table 3 presents the ten most highly cited documents in the dataset, providing insight into the foundational works that have most strongly shaped the intellectual landscape of IRT and Rasch model research in physics education.

Marsman et al. (2018) ranks first with 181 citations, introducing network psychometrics as a novel framework that connects the Ising network model with item response theory. This work reflects the growing interest in integrating advanced psychometric modeling approaches within the field. Boone & Scantlebury (2006) ranks second with 118 citations, representing one of the earliest and most influential studies explicitly advocating the use of Rasch analysis in science education research. This contribution has served as an important methodological foundation for many subsequent studies in the corpus.

Quellmalz et al. (2012), with 114 citations, highlights the critical role of simulation-based science assessments validated through IRT, demonstrating how technology-enhanced assessments can expand equitable access to standardized science measurement. Planinic et al. (2019), with 110 citations and the highest citation-per-year rate among the top five documents (15.7), represents one of the most impactful recent contributions. This study explicitly argues for the importance of Rasch measurement in physics education research and has rapidly become a key methodological reference in the field. Neumann et al. (2011), with 105 citations, further reinforces the importance of Rasch-based evaluation of instrument quality in science education contexts.

Notably, several highly cited works—including Traxler et al. (2018), Wang and Bao (2010), and Morris et al. (2012)—specifically apply IRT to the Force Concept Inventory (FCI). These studies highlight the central role of the FCI as one of the most extensively analyzed instruments in physics education psychometrics (Morris et al., 2012; Wang & Bao, 2010). The recurring presence of journals such as Physical Review Physics Education Research and Journal of Research in Science Teaching among the sources of these highly cited documents is also consistent with the findings presented in Table 2, confirming these outlets as some of the most influential publication venues in this research domain.

Citation-per-year metrics further reveal that more recent studies accumulate citations at a faster rate. Planinic et al. (2019) received an average of 15.7 citations per year, followed by Koerber and Osterhaus (2019) with 13.0 citations per year, and Traxler et al. (2018) with 11.2 citations per year. This pattern indicates increasing scholarly engagement with IRT–Rasch methodologies in physics and science education research over the past decade (Donthu et al., 2021).

Table 3 Top 10 most cited documents

Rank	Authors	Years	Journal	Cites	Cites/Years
1	Marsman et al.	2018	Multivariate Behavioral Research	181	22.6
2	Boone & Scantlebury	2006	Science Education	118	5.9
3	Quellmalz et al.	2012	Journal of Research in Science Teaching	114	8.1
4	Planinic et al.	2019	Physical Review Physics Education Research	110	15.7
5	Neumann et al.	2011	International Journal of Science Education	105	7.0
6	Eggert & Bögeholz	2010	Science Education	96	6.0
7	Wang & Bao	2010	American Journal of Physics	92	5.8
8	Koerber & Osterhaus	2019	Journal of Cognition and Development	91	13.0

Rank	Authors	Years	Journal	Cites	Cites/Years
9	Traxler et al.	2018	Physical Review Education Research	90	11.2
10	Plummer & Maynard	2014	Journal of Research in Science Teaching	83	6.9

Figure 9 illustrates the citation distribution across the 243 publications included in the dataset. The distribution exhibits a strongly right-skewed pattern, a common characteristic of bibliometric corpora in emerging and specialized research fields (Donthu et al., 2021; Thelwall, 2016). The majority of documents are concentrated in the lower citation range, with the largest proportion receiving between 0 and 5 citations, while only a small number of publications accumulate substantially higher citation counts.

This skewed pattern reflects the well-documented Matthew effect in academic publishing, whereby a small number of highly visible works attract a disproportionately large share of citations, while the majority of publications receive comparatively limited scholarly attention (Merton, 1968). Specifically, only 5 documents (2.1%) exceed 100 citations, and 23 documents (9.5%) surpass 50 citations. Nevertheless, these highly cited works collectively raise the overall average to 16.84 citations per document, substantially higher than the median of 8 citations, further confirming the pronounced skewness of the distribution.

The presence of 39 uncited documents (16.0%) is largely attributable to recently published works that have not yet had sufficient time to accumulate citations rather than indicating low scientific impact. This pattern is commonly observed in bibliometric studies that include publications from the most recent years (Aria & Cuccurullo, 2017; Donthu et al., 2021). Conversely, 111 documents (45.7%) have already achieved at least 10 citations, suggesting that nearly half of the corpus has attained meaningful scholarly visibility. This pattern indicates a healthy and expanding research community actively engaging with IRT–Rasch applications in the physics education (Planinic et al., 2019).

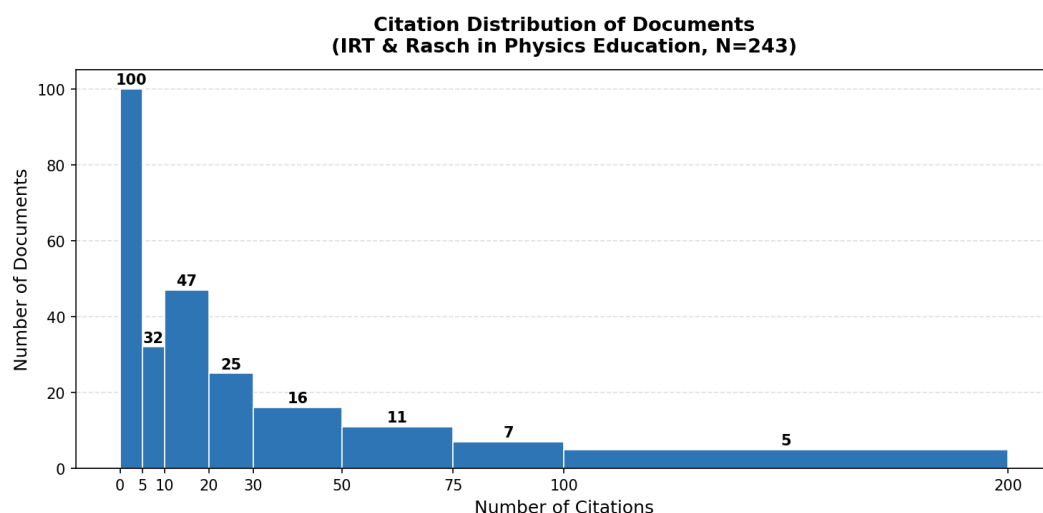


Figure 9 Citation distribution of documents

CONCLUSION

This bibliometric study systematically maps the intellectual structure of IRT and Rasch model applications in physics education assessment based on 243 Scopus-indexed publications published between 2001 and 2026. The analysis reveals a research field that has experienced remarkable growth, with the number of publications increasing by more

than 1,150%, from 18 documents in the first decade (2001–2010) to 225 documents in the subsequent period (2011–2026), reaching a peak of 32 publications in 2025. This trajectory indicates that IRT and Rasch measurement approaches have evolved from specialized psychometric tools into mainstream methodological frameworks in the physics and science education research (Boone & Staver, 2020; Planinic et al., 2019).

Performance analysis identifies *Physical Review Physics Education Research* as the most productive journal (15 articles), while Istiyono, Edi emerges as the most productive author (10 publications), reflecting the growing and significant contribution of Indonesian scholars in this domain. Citation analysis further reveals a right-skewed distribution, characteristic of an emerging research field, with Planinic et al. (2019) achieving the highest citation rate per year (15.7), highlighting its rapid adoption as a key methodological reference.

The keyword co-occurrence network reveals five major thematic clusters: (1) Core Measurement and Assessment, (2) Psychometric Frameworks and IRT, (3) Validity and Fairness, (4) Instrument and Technology Development, and (5) Learning Progression. Together, these clusters represent the multidimensional intellectual structure of the field (Aria & Cuccurullo, 2017; Cobo et al., 2011). Thematic and overlay analyses consistently indicate that while the Rasch model and educational assessment function as well-established motor themes, emerging topics such as computational thinking, instrument development, learning progression, and attitude represent the most recent research directions. These patterns signal a paradigmatic shift from conventional psychometric validation toward competency-based, technology-enhanced, and multidimensional assessment practices in the physics education (Istiyono et al., 2019; Liu, 2020; Zakwandi et al., 2024).

These findings highlight several promising directions for future research. Future studies should prioritize: (1) the development of multidimensional Rasch-based instruments capable of simultaneously measuring conceptual understanding, scientific reasoning, and higher-order thinking in physics; (2) the integration of computerized adaptive testing calibrated through Rasch measurement for real-time personalized physics assessment; (3) Rasch-based validation of instruments designed to measure computational thinking and digital literacy in physics learning contexts; and (4) the construction of Rasch-validated longitudinal learning progressions for core physics topics such as mechanics, electricity, and quantum phenomena. These directions represent high-impact research opportunities at the intersection of psychometrics, physics education, and educational technology.

This study has several limitations that should be acknowledged. First, the dataset was exclusively retrieved from the Scopus database, which may have excluded relevant publications indexed in other databases such as Web of Science, ERIC, or Google Scholar, potentially introducing selection bias. Second, the absence of author affiliation data in the exported metadata prevented the analysis of collaboration patterns at the country and institutional levels, which are typically included in comprehensive bibliometric studies. Third, the keyword co-occurrence analysis was constrained by the limited availability of author keywords in some records, which may have affected the completeness of thematic clustering. Finally, as a bibliometric study, this research focuses on mapping the structure and development of the existing literature rather than evaluating the methodological quality or empirical validity of individual studies.

The findings of this study carry several important implications for multiple stakeholders. For researchers, the identified thematic gaps, particularly in multidimensional assessment, technology-based instrument development, and learning progression, provide a strategic roadmap for future high-impact investigations. For instrument developers and assessment practitioners, the dominance of the Rasch model and Item Response Theory (IRT) as validated measurement frameworks underscores the importance of adopting these approaches rather than relying solely on classical test theory in developing robust physics

education assessment instruments. For journal editors and funding agencies, the rapid growth of this research field, particularly from emerging research communities such as Indonesia, indicates the need to expand platforms and resources that support psychometric research in physics education. Finally, for policymakers and curriculum designers, the convergence of IRT Rasch methodologies with competency-based assessment frameworks offers a scientific foundation for developing assessment systems that align with international standards while remaining responsive to local educational contexts.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS STATEMENT

Conceptualization, HR; Methodology, HR, DIR, and MJ; Software, MJ; Validation, HR, SY, and RS; Formal Analysis, MJ and DIR; Investigation, DIR and SY; Resources, RS; Writing (Original Draft Preparation), DIR and MJ; Writing (Review and Editing), HR and SY; Visualization, MJ; Supervision, HR; Project Administration, HR. All authors have read and agreed to the published version of the manuscript.

AI DISCLOSURE STATEMENT

During the preparation of this work, the authors used ChatGPT 4.0 (Artificial Intelligence) solely for language refinement and grammar checking. Following the use of this tool, the authors thoroughly reviewed, edited, and validated the content to ensure accuracy and academic integrity. The authors take full responsibility for the final content of the publication and declare that the core research, data analysis, interpretations, and manuscript structuring were conducted without the aid of artificial intelligence.

ACKNOWLEDGMENTS

This research received no external funding.

REFERENCES

- Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31(2–3), 162–172. <https://doi.org/10.1016/j.stueduc.2005.05.008>
- Addawiyah, I. R., Astra, I. M., Budi, E., & Wibowo, F. C. (2024). Multiple representations in physics learning: A bibliometric analysis. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 9(2), 277–285. <https://doi.org/10.26737/jipf.v9i2.4907>
- Amelia, R. N., & Kriswantoro, K. (2017). Implementation of item response theory for analysis of test items quality and students' ability in chemistry. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 2(1), 1–12. <https://doi.org/10.20961/JKPK.V2I1.8512>
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Bancong, H., Nurazmi, N., Fiskawarni, T. H., & Park, J. (2021). Trending research topics in the field of physics education from 2017 to 2019 in highly reputable international journals. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 10(1), 29–36. <https://doi.org/10.24042/JIPFALBIRUNI.V10I1.8341>
- Bezen, S., & Derman, İ. (2025). Unveiling global trends in gender equality research within physics education. *Social Sciences & Humanities Open*, 11, 101430. <https://doi.org/10.1016/j.ssaho.2025.101430>
- Bond, T. G., Yan, Z., & Heene, M. (2020). Applying the rasch model: Fundamental measurement in the human sciences. *Applying the Rasch Model: Fundamental*

- Measurement in the Human Sciences*, 1–348. <https://doi.org/10.4324/9780429030499>
- Boone, W. J. (2017). Rasch analysis for instrument development: Why, when, and how? *CBE - Life Sciences Education*, 15(4). <https://doi.org/10.1187/cbe.16-04-0148>
- Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, 90(2), 253–269. <https://doi.org/10.1002/sce.20106>
- Boone, W. J., & Staver, J. R. (2020). Advances in rasch analyses in the human sciences. *Advances in Rasch Analyses in the Human Sciences*. <https://doi.org/10.1007/978-3-030-43420-5>
- Boone, W. J., Towensend, J. S., & Staver, J. R. (2016). Utilizing multifaceted rasch measurement through facets to evaluate science education data sets composed of judges, respondents, and rating scale items: An exemplar utilizing the elementary science teaching analysis matrix instrument. *Science Education*, 100(2), 221–238. <https://doi.org/10.1002/sce.21210>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. <https://doi.org/10.1002/asi.21525>
- De Ayala, R. J. (2022). *The theory and practice of item response theory: Second edition*. 1–670.
- Ding, L., & Zhang, P. (2023). Assessment practices in interactive physics class: explicating theories and applications. *The International Handbook of Physics Education Research: Teaching Physics*. https://doi.org/10.1063/9780735425712_018
- Diya, D. R., & Putranta, H. (2025). Identifying the conceptual understanding of undergraduate physics students in solving basic physics problems: A descriptive quantitative study. *Journal of Mathematics Science and Computer Education*, 5(2), 98–114. <https://doi.org/10.20527/jmscedu.v5i2.14585>
- Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics - Physics Education Research*, 10(2), 1–58. <https://doi.org/10.1103/PhysRevSTPER.10.020119>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Doucette, D., & Singh, C. (2024). Gender equity in physics labs. *Physical Review Physics Education Research*, 20(1), 010102. <https://doi.org/10.1103/PhysRevPhysEducRes.20.010102>
- Eggert, S., & Bögeholz, S. (2010). Students’ use of decision - making strategies with regard to socioscientific issues: An application of the Rasch partial credit model. *Science Education*, 94(2), 230 – 258. <https://doi.org/10.1002/sce.20358>
- Embretson, S. E., & Reise, S. P. (2013). Item response theory for psychologists. *Item Response Theory for Psychologists*, 1–371. <https://doi.org/10.4324/9781410605269>
- Engelhard, G. (2013). Invariant measurement: Using rasch models in the social, behavioral, and health sciences. *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*, 1–288. <https://doi.org/10.4324/9780203073636>
- Evendi, E. (2022). Mathematical thinking styles and its implications in science learning: A bibliometric analysis. *Jurnal Penelitian Pendidikan IPA*, 8(3), 1503–1511. <https://doi.org/10.29303/jppipa.v8i3.1720>
- Haeruddin, Prasetyo, Z. K., Supahar, Sesa, E., & Lembah, G. (2020). Psychometric and structural evaluation of the physics metacognition inventory instrument. *European Journal of Educational Research*, 9(1), 215–225. <https://doi.org/10.12973/EU->

JER.9.1.215

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications.
- Harzing, A. W. (2007). *Publish or Perish (Version 8) [Computer software]*. <https://harzing.com/resources/publish-or-perish/windows>
- Hauenstein, C. E., & Embretson, S. E. (2018). Classical test theory. *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, 278–283. <https://doi.org/10.4135/9781506326139>
- Hidayana Nassiri, S., Abd, M., & Bunyamin, H. (2025). Trends and research frontiers in physics education: A bibliometric analysis from year 2021-2025. *Journal of Modern Education (IJMOE)*, 7(27), 625–642. <https://doi.org/10.35631/ijmoe.727039>
- Intasoi, S., Junpeng, P., Tang, K. N., Ketchatturat, J., Zhang, Y., & Wilson, M. (2020). Developing an assessment framework of multidimensional scientific competencies. *International Journal of Evaluation and Research in Education (IJERE)*, 9(4), 963–970. <https://doi.org/10.11591/IJERE.V9I4.20542>
- 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>
- Juliyanto, E., Wiyanto, W., Nugroho, S. E., & Mindyarto, B. N. (2026). PhET simulations performance to measure complex problem-solving skills in physics subject matter: Evidence of validity and reliability. *International Journal of Information and Education Technology*, 16(2), 422–435. <https://doi.org/10.18178/ijiet.2026.16.2.2515>
- Koerber, S., & Osterhaus, C. (2019). Individual differences in early scientific thinking: assessment, cognitive influences, and their relevance for science learning. *Journal of Cognition and Development*, 20(4), 510–533. <https://doi.org/10.1080/15248372.2019.1620232>
- Konu Kadirhanogullari, M., & Ozay Kose, E. (2024). A bibliometric analysis of articles on bibliometric studies in science education. *International Journal of Research in Education and Science*, 10(2), 315–339. <https://doi.org/10.46328/ijres.3370>
- Kotsis, K. T. (2025). Inquiry-based learning in science: Mathematical reasoning’s support of critical thinking. *Journal of Research in Mathematics, Science, and Technology Education*, 2(1), 60–72. <https://doi.org/10.70232/JRMSTE.V2I1.35>
- Krisnaningsih, E., Putri, M. A. N., Irba, T., Suprpto, N., Deta, U. A., & Hariyono, E. (2021). Bibliometric analysis of multi representation based on problem-solving skills using vosviewer. *Berkala Ilmiah Pendidikan Fisika*, 9(3), 274–284. <https://doi.org/10.20527/BIPF.V9I3.11329>
- Liu, X. (2020). Using and developing measurement instruments in science education: a rasch modeling approach. *Using and Developing Measurement Instruments in Science Education*. <https://doi.org/10.1108/978-1-64113-936-6>
- Liu, X., & Collard, S. (2005). Using the rasch model to validate stages of understanding the energy concept. *Journal of Applied Measurement*, 6(2), 224–241.
- Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., Maas, H. L. J. van der, & Maris, G. (2018). An introduction to network psychometrics: relating ising network models to item response theory models. *Multivariate Behavioral Research*, 53(1), 15–35. <https://doi.org/10.1080/00273171.2017.1379379>
- Merton, R. K. (1968). The matthew effect in science. *Science*, 159(3810), 56–62. <https://doi.org/10.1126/science.159.3810.56>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D’Amico, R., Deeks, J.

- J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P. C., ... Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLOS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Mongeon, P., & Paul-Hus, A. (2015). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* 2015 106:1, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de La Informacion*, 29(1), 1699–2407. <https://doi.org/10.3145/epi.2020.ene.03>
- Morell, L., Collier, T., Black, P., & Wilson, M. (2017). A construct-modeling approach to develop a learning progression of how students understand the structure of matter. *Journal of Research in Science Teaching*, 54(8), 1024–1048. <https://doi.org/10.1002/tea.21397>
- Morris, G. A., Branum-Martin, L., Harshman, N., Baker, S. D., Mazur, E., Dutta, S., Mzoughi, T., & McCauley, V. (2006). Testing the test: Item response curves and test quality. *American Journal of Physics*, 74(5), 449–453. <https://doi.org/10.1119/1.2174053>
- Morris, G. A., Harshman, N., Branum-Martin, L., Mazur, E., Mzoughi, T., & Baker, S. D. (2012). An item response curves analysis of the Force Concept Inventory. *American Journal of Physics*, 80(9), 825–831. <https://doi.org/10.1119/1.4731618>
- Mu'aziyah, S. E. S., Zainal, S., Agni, R., Supriyadi, S., & Isnawati, I. (2025). STEM-focused science identity questionnaire based on the pisa 2025 framework: Rasch model validation. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, 13(3), 737–748. <https://doi.org/10.23947/2334-8496-2025-13-3-737-748>
- National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. *The National Academies Press*, 1–385. <https://doi.org/10.17226/13165>
- Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch - based analyses of a nature of science test. *International Journal of Science Education*, 33(10), 1373–1405. <https://doi.org/10.1080/09500693.2010.511297>
- Nieminen, P., Savinainen, A., & Viiri, J. (2010). Force concept inventory-based multiple-choice test for investigating students' representational consistency. *Physical Review Special Topics - Physics Education Research*, 6(2), 020109. <https://doi.org/10.1103/PhysRevSTPER.6.020109>
- Novika, R., Istiyono, E., & Notiavina, A. A. (2025). Developing an assessment instrument to measure the minimum competency of high school physics using item response theory. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 29(1), 98–113. <https://doi.org/10.21831/pep.v29i1.70700>
- Nurfalah, F., Utama, J. A., & Muslim, M. (2025). Evaluating complex problem-solving test instrument quality in measuring computational thinking skills in global warming: rasch model analysis. *Jurnal Pemberdayaan Masyarakat*, 4(2), 326–337. <https://doi.org/10.46843/JPM.V4I2.425>
- Nurjanah, S., Iqbal, M., Sultan, J., Sholihin, S., Ulyasari, N., & Rashid, S. (2024). Research trends of augmented reality in physics education based on scopus database in thirteen years (2010-2023): A bibliometric approach. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 9(3), 438–452. <https://doi.org/10.26737/JIPF.V9I3.5383>
- Nurjanah, S., Sultan, J., Aisyah, S., Puspita, D., & Ulyasari, N. (2024). Bibliometric

- analysis of problem based learning in physics education: a scopus based study (1996-2023). *Berkala Ilmiah Pendidikan Fisika*, 12(2), 310–326. <https://doi.org/10.20527/bipf.v12i2.18775>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 020111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Plummer, J. D., & Maynard, L. (2014). Building a learning progression for celestial motion: An exploration of students' reasoning about the seasons. *Journal of Research in Science Teaching*, 51(7), 902–929. <https://doi.org/10.1002/tea.21151>
- Pursitasari, I. D., Permana, I., Jaenudin, D., Suriansyah, M. I., Sumiati, E. S., & Aqilah, S. (2025). Digital literacy and scientific literacy of science teachers and student learning styles in science learning. *Jurnal Penelitian Dan Pembelajaran IPA*, 11(1), 26–38. <https://doi.org/10.30870/JPPi.V11i1.30532>
- Qualters, D. M., Sheahan, T. C., Mason, E. J., Navick, D. S., & Dixon, M. (2008). Improving learning in first - year engineering courses through interdisciplinary collaborative assessment. *Journal of Engineering Education*, 97(1), 37–45. <https://doi.org/10.1002/j.2168-9830.2008.tb00952.x>
- Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, 49(3), 363–393. <https://doi.org/10.1002/tea.21005>
- Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The development of an instrument to measure the higher order thinking skill in physics. *European Journal of Educational Research*, 8(3), 743–751. <https://doi.org/10.12973/eu-jer.8.3.743>
- Riyadi, A., Susongko, P., & Pancasakti Tegal, U. (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>
- Rost, J., & Carstensen, C. H. (2002). Multidimensional rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, 26(1), 42–56. <https://doi.org/10.1177/0146621602026001003>
- Scantlebury, K., Boone, W., Kahle, J. B., & Fraser, B. J. (2001). Design, validation, and use of an evaluation instrument for monitoring systemic reform. *Journal of Research in Science Teaching*, 38(6), 646–662. <https://doi.org/10.1002/tea.1024>
- Thelwall, M. (2016). Citation count distributions for large monodisciplinary journals. *Journal of Informetrics*, 10(3), 863–874. <https://doi.org/10.1016/j.joi.2016.07.006>
- Traxler, A., Henderson, R., Stewart, J., Stewart, G., Papak, A., & Lindell, R. (2018). Gender fairness within the force concept inventory. *Physical Review Physics Education Research*, 14(1), 010103. <https://doi.org/10.1103/PhysRevPhysEducRes.14.010103>
- van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2009 84:2, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics*, 78(10), 1064–1070. <https://doi.org/10.1119/1.3443565>
- Zakwandi, R., Istiyono, E., & Dwandaru, W. S. B. (2024). A two-tier computerized adaptive test to measure student computational thinking skills. *Education and*

- Information Technologies*, 29(7), 8579–8608. <https://doi.org/10.1007/s10639-023-12093-w>
- Zayrin, A. A., Nupus, H., Maizia, K. K., Marsela, S., Hidayatullah, R., & Harmonedi, H. (2025). Analisis instrumen penelitian pendidikan (uji validitas dan reliabilitas instrumen penelitian). *QOSIM: Jurnal Pendidikan Sosial & Humaniora*, 3(2), 780–789. <https://doi.org/10.61104/JQ.V3I2.1070>
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>

AUTHOR PROFILES



Dr. Haeruddin, M.Si
Expertise: Assessment of Physics Education
Affiliation: Physics Education Study Program, Faculty of Teacher Training and Education, Tadulako University
Email: haeruddin@untad.ac.id
Google Scholar: <https://scholar.google.com/citations?user=bV--kVwAAAAJ>
Sinta ID: 6182177
ORCID: <https://orcid.org/0000-0003-4654-7637>
SCOPUS ID: 57212874020



Delthawati Isti Ratnaningtyas, M.Pd. is a lecturer at the Faculty of Teacher Training and Education of Universitas Tadulako. Her expertise is in educational assessment and learning media development.
Email: delthawati@untad.ac.id
Google Scholar: <https://scholar.google.com/citations?hl=en&user=mN8bbu4AAAAJ>
Sinta ID: 6762675
ORCID: <https://orcid.org/0009-0002-1394-9306>



Muhammad Jarnawi, S.Pd., M.Pd.
Expertise: written assessment
Affiliation: Physics Education Study Program, Faculty of Teacher Training and Education, Tadulako University
Email: i.am.jarnawi@gmail.com
Scoolar: <https://scholar.google.co.id/citations?user=DJUrodKAAAAJ&hl=id>
Sinta: 6061420
Orchid: <https://orcid.org/0000-0002-0324-599X>
Scopus ID: 57414816500



Syamsuriwal, S.Pd., M.Pd.
Expertise: written assessment
Affiliation: Physics Education Study Program, Faculty of Teacher Training and Education, Tadulako University
Email: riwalpfisika@untad.ac.id
Google Scholar: <https://scholar.google.com/citations?hl=en&user=Y-3jvUQAAAAJ>
Sinta ID: 6722648
Orchid: <https://orcid.org/0009-0005-1005-7996>



Rudi Santoso, S.Pd., M.Pd is a lecturer at the Faculty of Teacher Training and Education of Universitas Tadulako. Her expertise is in learning media development.

Email: rudisantoso@untad.ac.id

Google Scholar:

<https://scholar.google.com/citations?hl=id&user=SHz5piEAAAAJ>

Sinta ID: 6750566

Orcid ID: <https://orcid.org/0009-0000-4880-3898>

Scopus ID: 58036456200