Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6, 2003-2022
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i6.8310

Unlocking the predictive power of the Rasch model: A systematic literature review on educational instrument calibration and assessment accuracy

I Made Arjana[1]*, Desak Putu Parmiti[2], I Made Candiasa[3], Ni Ketut Widiartini[4]
[1,2,3,4]Ganesha University of Education, Udayana Street, No.11, Bali 81116, Indonesia; imdarjana@gmail.com (I.M.A.)

Abstract: This systematic literature review explores the application of the Rasch Model in educational measurement, highlighting its role in psychometric validation, DIF detection, and multidimensional assessments. The study examines key software tools and diverse research applications, with a focus on STEM education and large-scale testing programs. It also reviews the advantages of different Rasch-based software, including Winsteps, RUMM2030, and ConQuest, in facilitating accurate measurement. The findings reveal methodological challenges, including limited cross-cultural validations, inconsistent model applications, and insufficient sample representation, which impact the reliability and generalizability of Rasch-based assessments. The review identifies gaps in scaling methodologies, response category designs, and adaptation processes across different educational contexts. Future research should prioritize AI-driven Rasch analysis, comparative model evaluations, and interdisciplinary integrations to refine educational assessments. Additionally, expanding real-time psychometric evaluations and cross-cultural validations will enhance the applicability of the Rasch Model in diverse educational settings. Strengthening methodological rigor and ensuring greater transparency in validation procedures are crucial to advancing the field. Addressing these issues will promote more equitable, reliable, and innovative measurement frameworks, ultimately improving the accuracy and fairness of educational assessments.

Keywords: *AI-driven analysis, Assessment reliability, Differential item functioning, Educational measurement, Large-scale testing, Psychometric validation, Rasch Model.*

1. Introduction

Educational assessment plays a fundamental role in shaping academic policies, instructional strategies, and student learning outcomes. The effectiveness of assessment tools depends on their ability to measure constructs reliably and validly. Over the last decade, the Rasch Model has gained prominence in educational measurement as a powerful tool for analyzing and validating assessment instruments. The model, which is rooted in item response theory, allows researchers to evaluate the properties of test items systematically, ensuring the accuracy and fairness of educational assessments [1]. By focusing on individual item performance and person parameters, Rasch analysis offers a significant improvement over Classical Test Theory (CTT), which relies heavily on sample-dependent statistics [2].

The Rasch Model has been widely applied to assess educational instruments, ranging from subject-specific self-concept inventories to competency assessments in STEM education [3]. One of the key advantages of the Rasch Model is its ability to ensure unidimensionality, meaning that a test measures a single latent trait, which is crucial for valid educational assessments [1]. Moreover, Rasch modeling enables the detection of differential item functioning (DIF), which identifies potential biases in test items across different demographic groups, thus promoting fairness in assessment practices [4, 5]. This feature is particularly valuable in large-scale educational assessments, where fairness and validity are essential.

History: Received: 17 April 2025; Revised: 20 May 2025; Accepted: 23 May 2025; Published: 23 June 2025
* Correspondence: imdarjana@gmail.com

Another critical advantage of the Rasch Model is its capacity for individualized item and person parameter estimation, enhancing measurement precision [6, 7]. Unlike CTT, which assumes that measurement error is constant across all ability levels, the Rasch Model recognizes that error varies depending on the difficulty level of an item and the ability of a respondent. This results in a more refined analysis of student performance and instrument effectiveness [8]. Additionally, the generalizability of Rasch-based findings makes the model particularly useful in educational research, as it minimizes the influence of specific sample characteristics, reducing biases and potential distortions in measurement outcomes [2, 9].

The practical applications of the Rasch Model in educational research have expanded significantly in recent years. For example, the model has been employed to evaluate the psychometric properties of the Chemistry Self-Concept Inventory, confirming its suitability for assessing students' perceptions of their abilities in science education, particularly during the COVID-19 pandemic [1]. Similarly, it has been applied to the development of tools for measuring critical thinking skills in physics education [10] and competency assessments in STEM education [3]. These studies underscore the versatility of the Rasch Model in assessing a wide range of educational constructs, thereby reinforcing its role in advancing psychometric research.

Furthermore, the increasing integration of digital assessments and technology-enhanced learning environments has further expanded the applicability of the Rasch Model. Recent studies highlight the model's relevance in assessing student performance within blended learning frameworks, where traditional assessment approaches may fall short in capturing complex learning interactions [11]. The ability of the Rasch Model to provide detailed feedback on both item performance and participant ability makes it an essential tool in modern education, particularly in the context of digital learning and adaptive assessments.

Despite its advantages, research on the Rasch Model continues to reveal gaps that warrant further exploration. One of the primary limitations is the scarcity of longitudinal studies that examine the stability of Rasch-analyzed instruments over time [2]. While many studies validate instruments at a single time point, there is a need for research that investigates how these tools perform across different testing periods and educational settings. Additionally, there is limited research on the application of the Rasch Model in non-traditional learning environments, such as informal education settings and culturally diverse classrooms [12]. Given the increasing globalization of education, more studies are needed to explore how the Rasch Model can be effectively utilized in varied cultural contexts.

This systematic literature review (SLR) aims to provide a comprehensive synthesis of existing research on the Rasch Model's application in educational measurement. Specifically, the review seeks to identify key trends in Rasch-based assessments, examine methodological approaches, and highlight areas for future research. The scope of this review includes studies published over the past decade, with a focus on Rasch applications in educational assessments across different disciplines and learning environments. The review also considers the role of software tools in facilitating Rasch analysis, given the increasing reliance on computational methods in psychometric research.

The review is structured as follows: The next section outlines the systematic review methodology, including search strategies, inclusion criteria, and quality assessment procedures. This is followed by a discussion of the theoretical foundations of the Rasch Model, comparing its principles with alternative measurement theories. The subsequent section synthesizes findings from recent studies, categorizing research trends, methodological approaches, and practical applications. The discussion section then interprets these findings, highlighting theoretical and practical implications while identifying research gaps. Finally, the review concludes with recommendations for future research directions, emphasizing the need for further exploration of the Rasch Model's applicability in diverse educational contexts.

By synthesizing current research and identifying gaps, this review contributes to the ongoing discourse on educational assessment and psychometric modeling. The insights gained from this analysis will inform future studies on how to enhance the validity, reliability, and fairness of educational assessments using the Rasch Model. Moreover, it will provide educators, policymakers, and researchers with a deeper understanding of the model's strengths and limitations, ultimately guiding its application in evidence-based educational practices.

2. Methods

2.1. Systematic Review Framework

This study follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure methodological transparency and reproducibility [13]. The PRISMA framework provides a structured approach to systematic reviews, outlining a 27-item checklist and a flow diagram that guides researchers through the stages of identification, screening, eligibility assessment, and inclusion [14]. The implementation of PRISMA in educational research has demonstrated its adaptability beyond healthcare, enabling structured syntheses in fields such as educational technology and pedagogical evaluation [15, 16].

This systematic literature review (SLR) focuses on applications of the Rasch Model in educational measurement. The review follows a predefined protocol to ensure a rigorous and unbiased selection of studies. The research strategy includes a comprehensive search of relevant databases, application of strict inclusion and exclusion criteria, and a robust quality assessment process. The methodology is detailed below and visually represented in Figure 1 (PRISMA Statement).

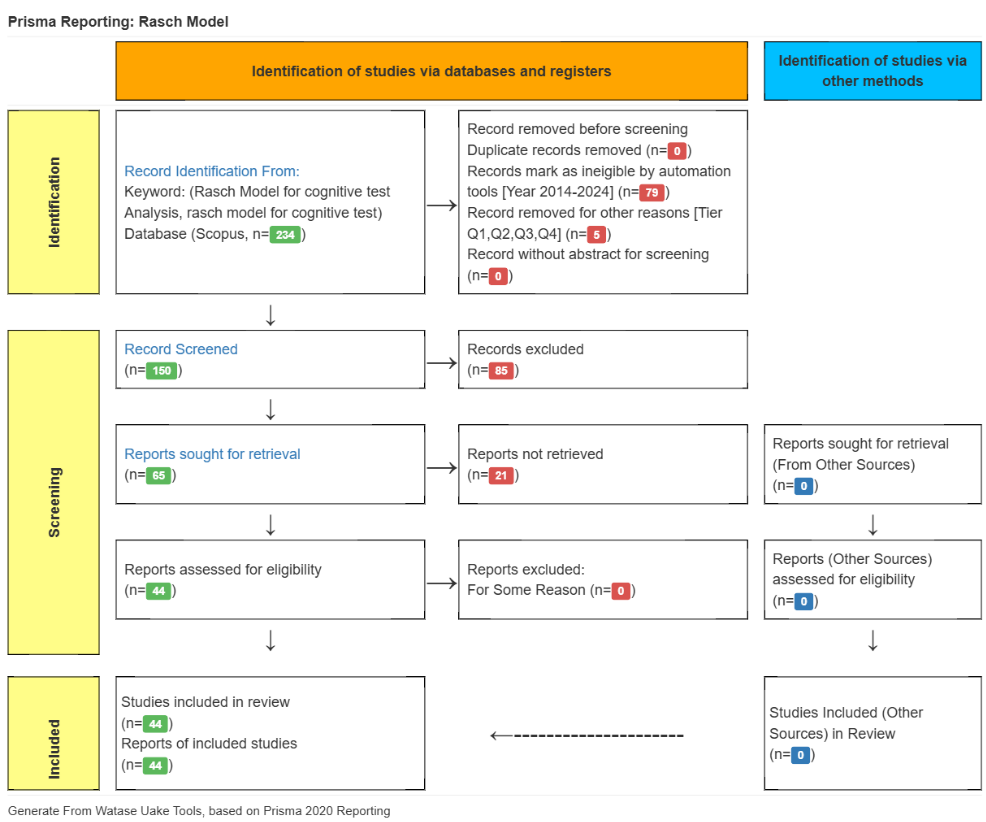

Figure 1.
PRISMA Statement.

2.2. Data Sources and Search Strategy

A systematic search was conducted in major academic databases, ensuring comprehensive coverage of relevant studies. The selected databases include Scopus, Web of Science, and ERIC, as these platforms offer extensive literature on educational assessments, psychometric methodologies, and

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2003-2022, 2025
DOI: 10.55214/25768484.v9i6.8310

applications of the Rasch Model [17]. To maximize retrieval of pertinent studies, Boolean operators were applied using the following search strings:

- *("Rasch Model" OR "Rasch Analysis") AND ("educational assessment" OR "measurement")*
- *("psychometric evaluation" AND "Rasch Model") AND ("instrument calibration" OR "validity")*

Search queries were refined by applying filters to limit results to peer-reviewed journal articles published between 2014 and 2024. Additionally, reference lists of relevant papers were screened to identify additional sources, ensuring a comprehensive literature search. Studies in multiple languages were considered, provided they included English abstracts and were published in reputable journals.

2.3. Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were established to maintain the relevance and quality of the selected studies. These criteria were formulated based on best practices in psychometric research [18, 19].

Inclusion Criteria: Studies published between 2014 and 2024, empirical research applying the Rasch Model in educational measurement, articles discussing psychometric validation of educational assessments, studies published in peer-reviewed journals, and research with sufficient methodological detail to allow for quality appraisal.

Exclusion Criteria: Studies without clear psychometric analysis or validation focus, papers that only provide theoretical discussions without empirical data, research that does not apply Rasch modeling in education-related contexts, non-peer-reviewed publications such as conference papers, book chapters, and unpublished theses, and articles without an abstract in English, limiting accessibility.

2.4. Data Extraction

Extracted data included study characteristics such as author, year, journal, and country of study, methodological details including sample size, instrument type, and Rasch model variant used, key findings relating to psychometric properties assessed and major conclusions, and software utilized such as Winsteps, RUMM2030, and ConQuest. This standardized approach enhances the reliability and consistency of the review process. The findings from the extracted data are synthesized and discussed in subsequent sections, providing insights into research trends and methodological advancements in Rasch-based educational measurement. By implementing rigorous search, screening, and quality assessment procedures, this study ensures the selection of high-quality literature that contributes to a comprehensive understanding of Rasch Model applications in educational research. The next section discusses the theoretical underpinnings of the Rasch Model, comparing it with alternative psychometric frameworks and highlighting its strengths in instrument calibration and assessment accuracy.

3. Theoretical Framework and Background

3.1. Rasch Model in Modern Test Theory

The Rasch Model is a probabilistic measurement model that has gained prominence in modern test theory due to its advantages over Classical Test Theory (CTT). While CTT relies on aggregate statistics such as total test scores and assumes that measurement error is constant across all ability levels, the Rasch Model provides a more precise estimation of item and person parameters by employing a logistic function [2]. This item response theory (IRT)-based approach enables a more individualized assessment, reducing dependency on sample-specific characteristics and improving measurement precision.

A key advantage of the Rasch Model is its ability to ensure measurement invariance, meaning that item parameters remain consistent across different groups of test-takers. This feature makes it a valuable tool for evaluating differential item functioning (DIF), allowing researchers to detect biases in assessment items [20]. Unlike CTT, which does not rigorously account for individual variations in response behavior, the Rasch Model provides a framework for fairer assessments by ensuring that test items measure the intended construct equally across diverse populations.

ISSN: 2576-8484
Vol. 9, No. 6: 2003-2022, 2025
DOI: 10.55214/25768484.v9i6.8310

The Rasch Model has been widely applied in large-scale educational assessments, such as the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) [6]. These large-scale assessments benefit from Rasch's robust linking methods, which allow for the calibration of item difficulty across different test administrations and diverse populations. Additionally, the model supports adaptive testing environments, where item difficulty can be dynamically adjusted based on a respondent's ability, further enhancing test validity [21]. This adaptability underscores the growing importance of the Rasch Model in contemporary educational measurement and evaluation.

3.2. Variants of the Rasch Model

Different variations of the Rasch Model have been developed to accommodate diverse assessment needs. These include dichotomous, polytomous, and rating scale models, each designed to handle different response formats and enhance measurement precision.

3.2.1. Dichotomous Rasch Model

The dichotomous Rasch model is the simplest form, applied in assessments where responses are binary (e.g., correct/incorrect, true/false). This model is widely used in standardized testing and multiple-choice assessments, where it helps evaluate item difficulty and person ability with minimal estimation bias [6]. It ensures that test items contribute equally to the measurement of the underlying construct, making it a preferred choice for objective educational assessments.

3.2.2. Polytomous Rasch Model

The polytomous Rasch model extends the dichotomous approach by accommodating multiple-graded responses, making it useful for assessments that involve partial credit scoring or categorical response scales. This model is particularly valuable in STEM education, where assessments of problem-solving skills and conceptual understanding often require nuanced grading [1]. Its ability to capture varying levels of proficiency allows for a more detailed analysis of student performance beyond binary correctness.

3.2.3. Rating Scale Model

The rating scale model is a specialized form of the polytomous Rasch model, designed to handle ordinal data, such as responses on Likert scales. This model is commonly applied in educational psychology and humanities research, where attitudes, perceptions, and self-efficacy beliefs need to be measured reliably [20]. In studies on teaching efficacy and student motivation, the rating scale model has been instrumental in ensuring that survey instruments provide valid and comparable data across different respondent groups [22].

3.2.4. Applications of Rasch Model Variants in Educational Assessment

The flexibility of Rasch Model variants has led to their widespread use across educational disciplines. In STEM education, polytomous Rasch models have been utilized to assess critical thinking skills in physics and e-learning readiness among students [5]. Similarly, in humanities and social sciences, rating scale models have been applied to evaluate student attitudes toward learning environments and teacher effectiveness [22].

The selection of an appropriate Rasch Model variant depends on the nature of the assessment and the data type being collected. For assessments involving binary responses, the dichotomous model is preferred, whereas evaluations of subjective constructs or multi-level competencies benefit from polytomous and rating scale models [23]. Researchers also consider statistical and psychometric requirements, ensuring that the chosen model aligns with the theoretical framework of the construct being measured [24].

Overall, the adaptability and rigor of Rasch Model variants make them indispensable in educational assessment, ensuring measurement accuracy, fairness, and applicability across diverse learning contexts. The next section of this review explores recent research trends and methodological advancements in Rasch-based educational assessment.

4. Review of Themes and Findings

4.1. Trends in Publication and Research Focus

Over the past decade, research on the Rasch Model has experienced significant growth, reflecting an increasing interest in its applications across various fields, particularly in educational measurement. Studies have shown that the Rasch framework has been widely used to validate assessment instruments, improving their psychometric properties in disciplines such as STEM and humanities [25, 26]. Recent research has expanded beyond traditional applications, incorporating innovative uses such as developmental screening tools [27] cognitive assessments [28] and multimedia-based testing [29]. Figure 2 presents a WordCloud based on Scopus keywords, illustrating the dominant themes and concepts in Rasch-based research.

In terms of global contributions, developed countries dominate the research output on the Rasch Model, often focusing on advanced methodologies and large-scale assessments like those associated with PISA and TIMSS [30, 31]. Meanwhile, researchers in developing countries increasingly adopt the Rasch Model but tend to prioritize basic psychometric validation and local adaptation of existing tools [26]. This contrast reflects disparities in research infrastructure, access to advanced statistical software, and differing educational priorities. Figure 3 illustrates the year-wise publication classification, while Figure 4 and Figure 5 provide insights into country-wise contributions to Rasch Model research.

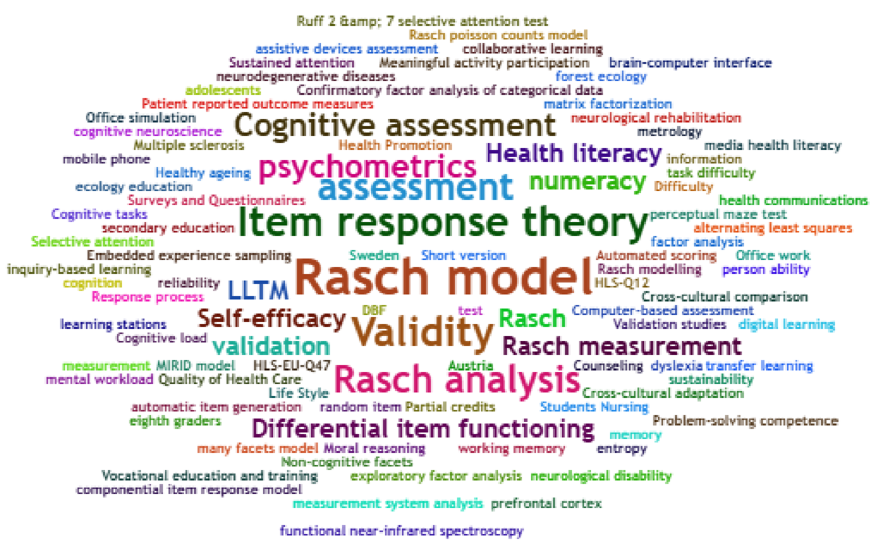

Figure 2.
WordCloud Based on Keywords from Scopus.

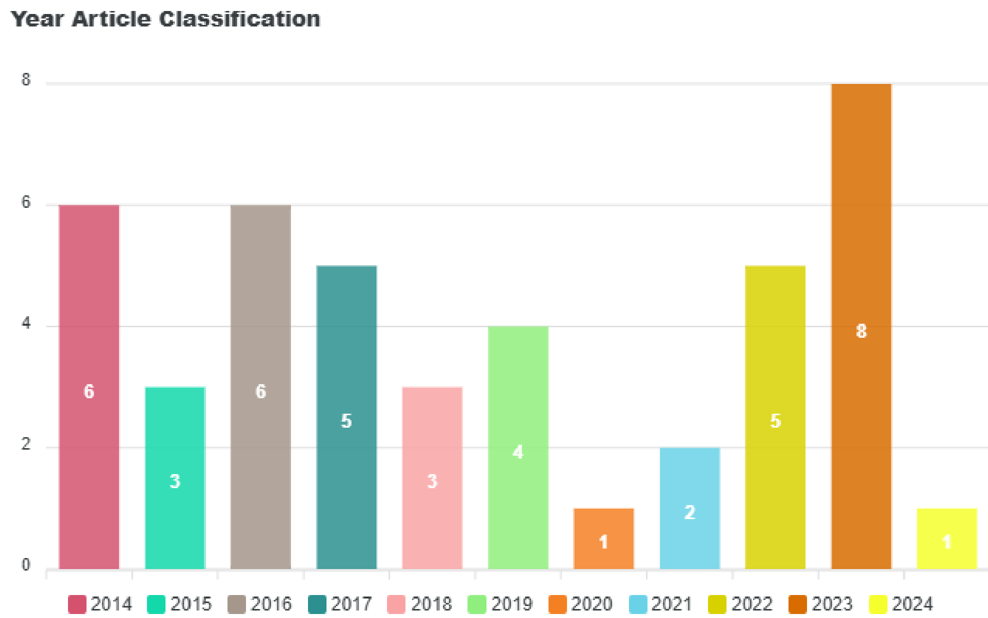

Figure 3.
Publication of Article.

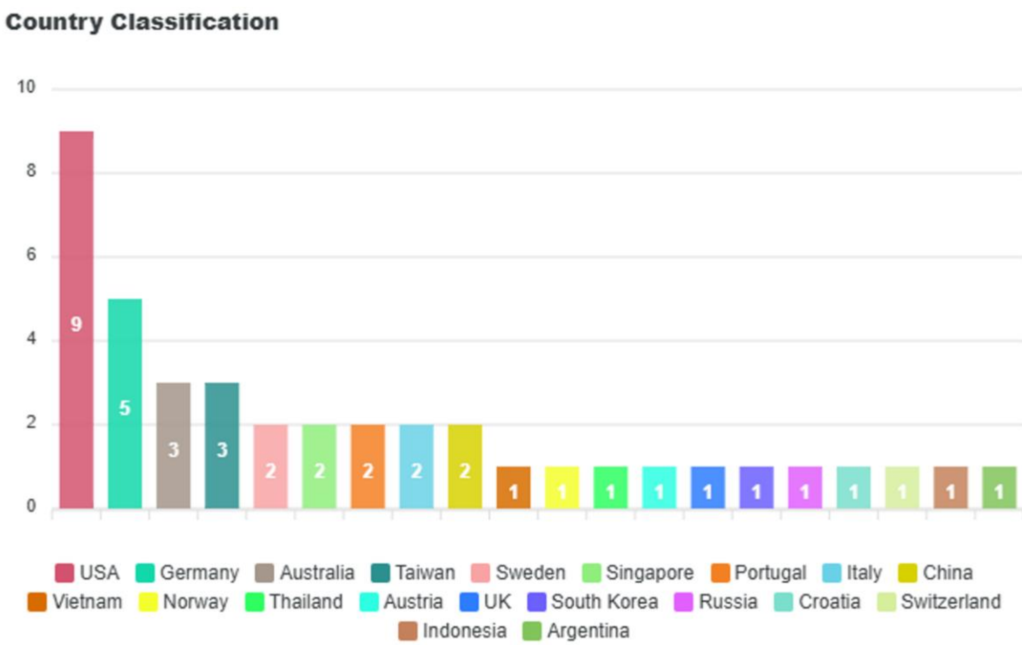

Figure 4.
Country of research.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2003-2022, 2025
DOI: 10.55214/25768484.v9i6.8310

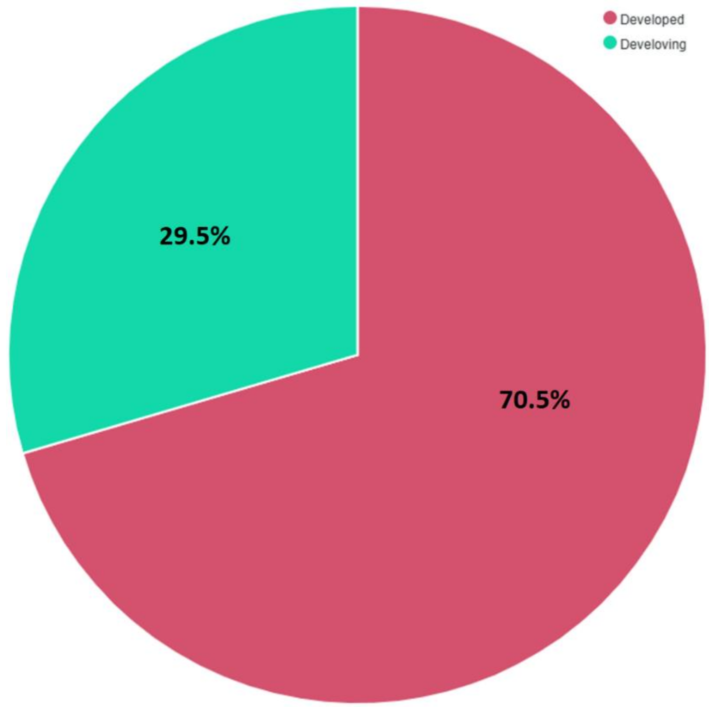

Figure 5.
Country Classification.

4.2. Contexts of Rasch Model Application

The Rasch Model has been extensively applied across diverse educational settings, particularly in K-12 education, higher education, and vocational training. However, it is most frequently utilized in higher education and large-scale K-12 assessments. Within higher education, the Rasch Model plays a crucial role in the psychometric evaluation of entrance exams and course assessments, ensuring their validity and reliability for diverse student populations [32]. In K-12 settings, its application is prominent in formative assessments and standardized testing programs, which aim to improve educational accountability and learning outcomes [33]. Figure 6 visualizes the context classification of Rasch Model applications across different educational settings.

Disciplinary applications of the Rasch Model also vary widely. In STEM fields, it is commonly used to analyze problem-solving skills in physics and mathematics, allowing for deeper insights into student comprehension and learning trajectories [34]. Conversely, in humanities and social sciences, the Rasch Model is frequently employed to evaluate student attitudes, moral reasoning, and metacognitive awareness [35]. These variations highlight the model's adaptability in measuring both cognitive and affective constructs within educational research. Figure 7 presents an overview of theory classifications in Rasch-based studies, demonstrating the broad theoretical foundations utilized.

Emerging interdisciplinary trends suggest an increasing convergence of Rasch analysis with qualitative methodologies to provide richer interpretations of student engagement and learning behaviors [36, 37]. Additionally, technological advancements, including game-based assessments and online adaptive testing, have contributed to the expansion of Rasch-based research, facilitating personalized learning experiences [34, 37].

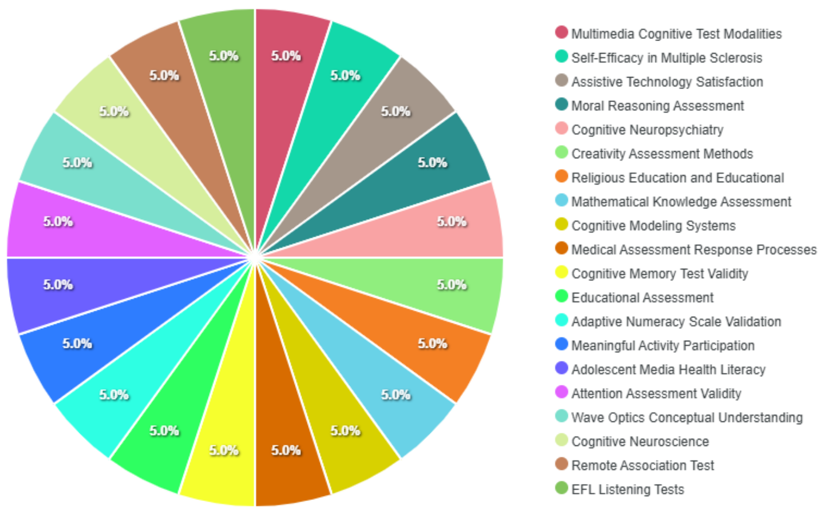

Figure 6.
Context Classification.

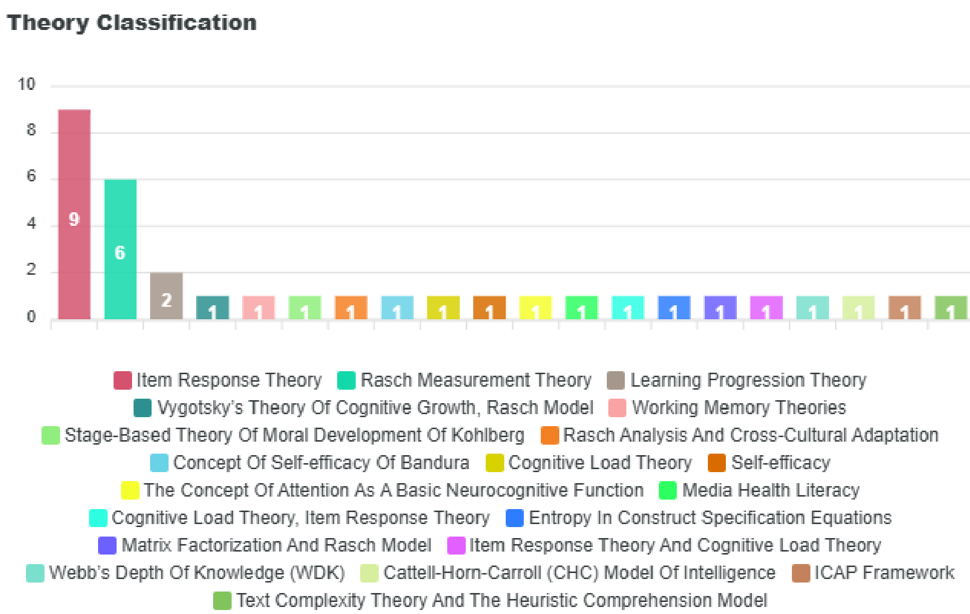

Figure 7.
Theory classification.

4.3. Methodological Approaches

The most prevalent research designs in Rasch-based studies include survey research and mixed-methods approaches, while experimental designs remain relatively rare. Surveys are widely used to collect data on student knowledge, attitudes, and competencies, whereas mixed-methods approaches integrate quantitative Rasch analysis with qualitative insights to provide a comprehensive understanding of educational phenomena [38, 39]. Figure 8 and Figure 9 illustrate the classification of research methods and research design types commonly employed in Rasch studies.

Data collection strategies have evolved significantly, particularly with the rise of digital assessments and online surveys, increasing accessibility and participant responsiveness [27, 40]. The incorporation of technology-enhanced tools has allowed researchers to collect more reliable and extensive datasets, fostering improved psychometric analysis and validation. Figure 10 presents an overview of data collection methods used in Rasch Model research.

Statistical methods commonly employed in Rasch Model-based research include:

- Goodness-of-fit statistics – To evaluate how well items align with the Rasch Model's assumptions.
- Item difficulty calibration – To measure how challenging each test item is for respondents.
- Reliability estimates – To determine the consistency and accuracy of test scores.

Additionally, differential item functioning (DIF) analysis is frequently applied to examine fairness across demographic groups, ensuring that assessment instruments are free from bias [41]. These methodological approaches collectively reinforce the validity and reliability of Rasch-based educational assessments [33, 34].

This review of themes and findings highlights the increasing prominence of the Rasch Model in educational research, its broad applicability across multiple disciplines and settings, and the evolution of methodological approaches enhancing its effectiveness. Future research should explore longitudinal applications of the Rasch Model, particularly in assessing learning progression over time and evaluating technological advancements in educational measurement. Moreover, further investigation is needed into cross-cultural adaptations of Rasch-based assessments to enhance their global applicability and fairness in diverse educational contexts.

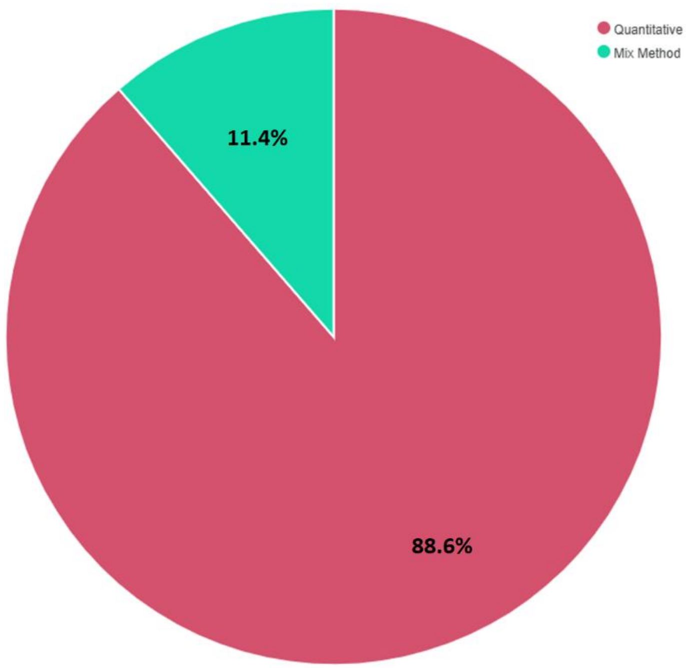

Figure 8.
Method of classification.

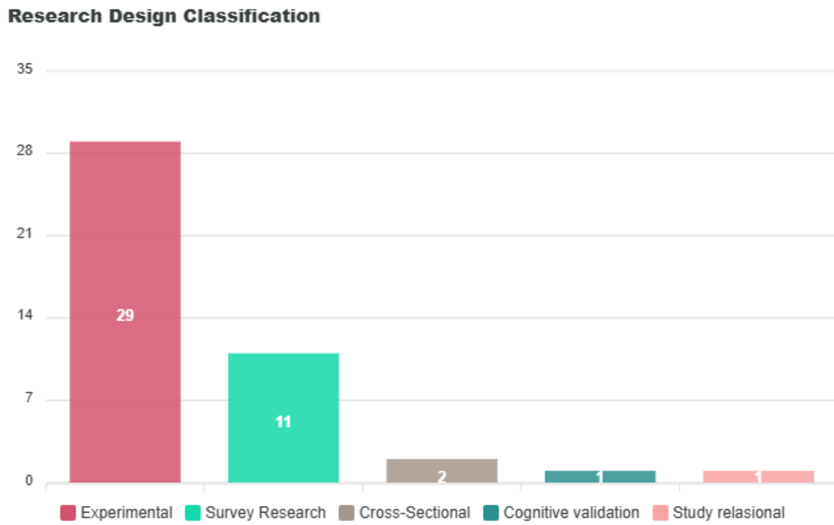

Figure 9.
Research design.

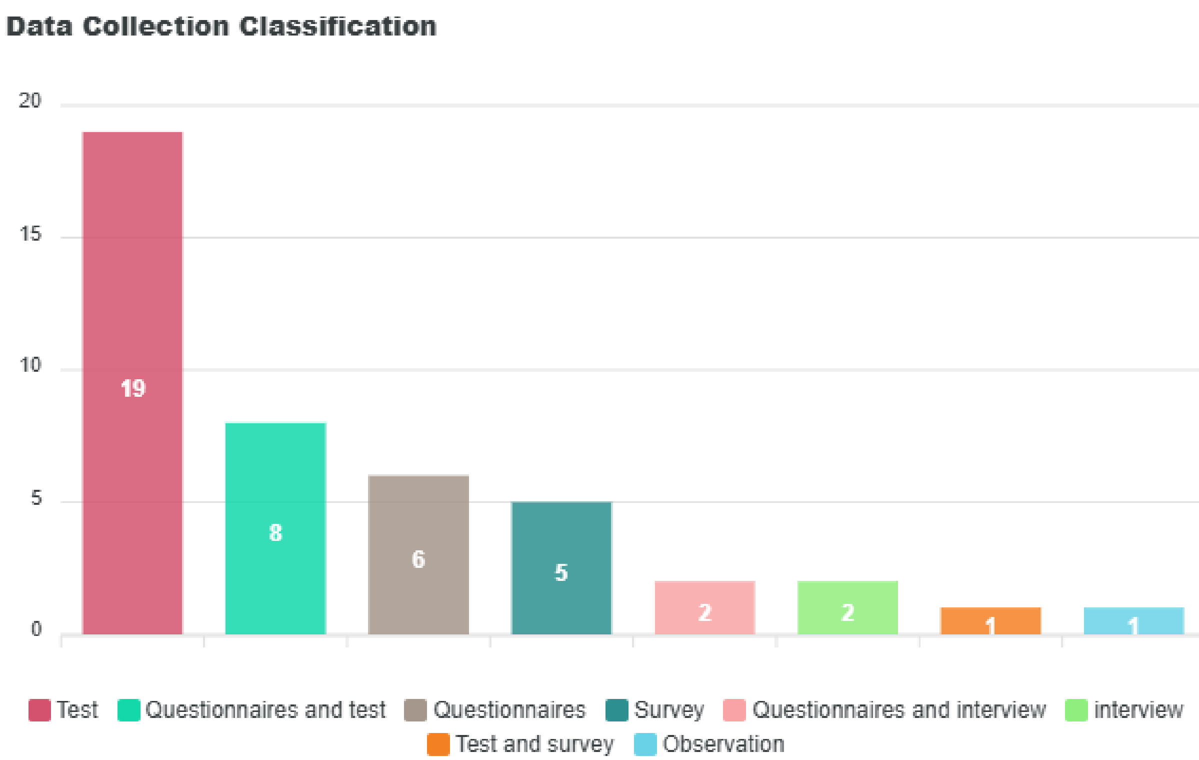

Figure 10.
Data Collection.

4.4. Impact of Rasch-Based Software

The selection of Rasch-based software plays a crucial role in determining the accuracy, validity, and reliability of educational assessment studies. Various software packages provide unique features and analytical capabilities tailored for specific research needs. Winsteps, for example, is widely recognized for its user-friendly interface and extensive reporting functionalities, making it a popular choice in K-12 educational assessments [38]. RUMM2030, on the other hand, is favored for its flexibility in analyzing polytomous data and supporting multiple Rasch model frameworks, making it a preferred tool in advanced psychometric research [25]. ConQuest is particularly suited for handling large-scale assessments like PISA, given its robust modeling capabilities and support for complex, multidimensional datasets [32].

The distribution of Rasch-based software usage is illustrated in Figure 11, showing that Winsteps leads with 31.8%, followed by R Package (20.5%) and RUMM2030 (15.9%). These preferences highlight researchers' inclinations towards software that balances ease of use with analytical sophistication.

The choice of software significantly influences research outcomes by affecting the precision of item calibration, person ability estimation, and model fit statistics. For instance, while Winsteps provides comprehensive fit indices suitable for basic Rasch analysis, ConQuest accommodates complex multidimensional constructs, offering superior analytical flexibility for large-scale evaluations [42]. This distinction is particularly important for researchers working with hierarchical datasets and adaptive learning models.

Emerging trends indicate an increasing integration of machine learning techniques and real-time data processing functionalities into Rasch-based software [41]. There is also a growing emphasis on user-friendly platforms to facilitate access for researchers with limited statistical expertise, promoting the democratization of Rasch analysis tools [39]. The expansion of software applications across fields is shown in Figure 12, where educational research leads with 63.5%, followed by health sciences (25.5%) and psychology (11.4%). Additionally, the classification of Rasch model types in Figure 13 highlights that Unidimensional Rasch Models dominate at 81.8%, with limited but notable applications of Multidimensional and Mixture Rasch Models.

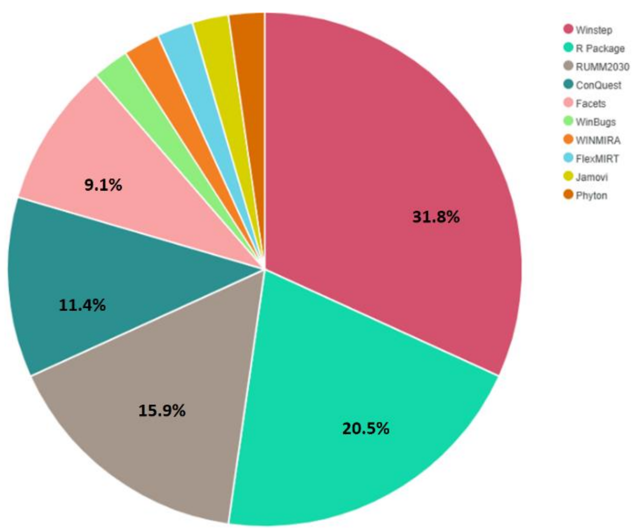

Figure 11.
Software Classification.

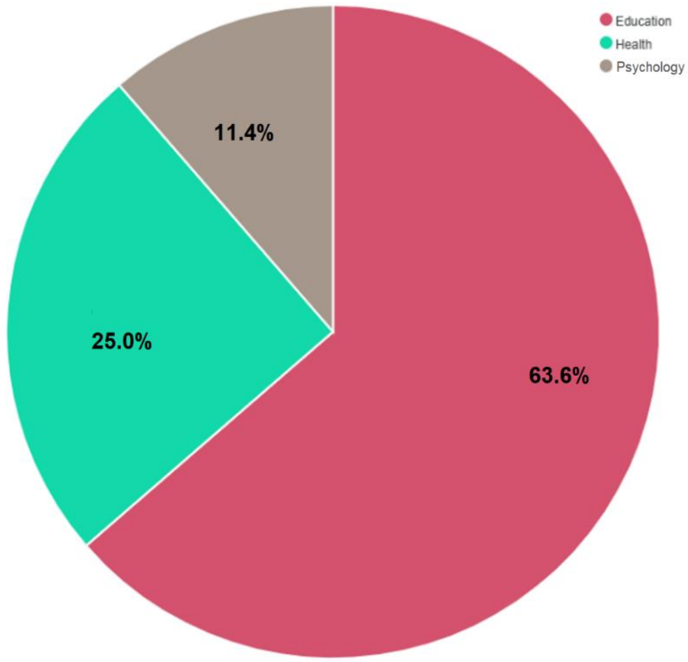

Figure 12.
Research Fields.

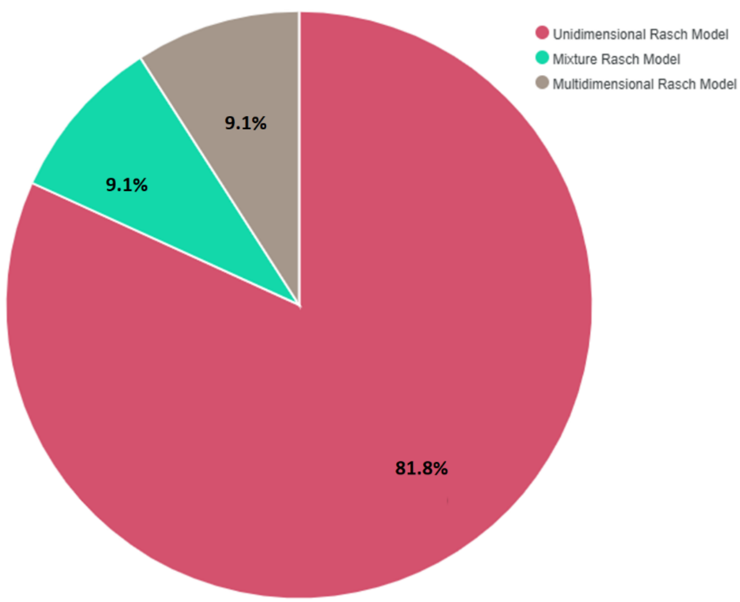

Figure 13.
Rasch model types.

4.5. Citation Impact and Key Contributions

The influence of Rasch Model research is reflected in citation metrics, highlighting the most significant contributions in both theoretical and practical advancements. One of the most cited studies, "Psychometric Properties of the Montreal Cognitive Assessment (MoCA)" by Freitas, et al. [32] set a benchmark for cognitive assessment in neuropsychology through its comprehensive Rasch validation. Similarly, Finbråten, et al. [43] made a significant impact in health literacy research, employing Rasch

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2003-2022, 2025
DOI: 10.55214/25768484.v9i6.8310

analysis to develop a validated short version of the HLS-Q12 survey. These studies exemplify the practical applications of Rasch modeling in educational, health, and psychological assessments.

A detailed breakdown of top-cited authors is presented in Table 1, where Finbråten, et al. [43] lead with 78 citations, followed by Herrmann-Abell and DeBoer [44] with 64 citations for their study on learning progressions in science education. Citation trends also demonstrate the growing prominence of Rasch-based research in high-impact journals. As seen in Table 2, BMC Health Services Research and Journal of Research in Science Teaching rank among the most influential Q1 journals publishing Rasch Model studies.

Theoretical advancements in Rasch measurement have been significantly influenced by Primi [25] who combined many-facet Rasch measurement with cognitive psychology to refine creativity assessments. Similarly, Herrmann-Abell and DeBoer [44] contributed to the understanding of learning progressions, applying Rasch analysis to track cognitive development in science education.

From a practical perspective, top-cited studies have enhanced psychometric validity in multiple disciplines, providing frameworks for developing robust assessment tools. For example, Antonietti, et al. [30] explored technology integration in education, helping educators adopt evidence-based strategies for digital learning assessment. These contributions demonstrate the continued relevance of Rasch Model applications in shaping educational policies, cognitive evaluations, and cross-disciplinary psychometric research.

Overall, the findings underscore the widespread influence of Rasch-based software in assessment research, as well as the substantial impact of highly cited studies in advancing theoretical and applied psychometric methodologies. Future research should explore the integration of AI-driven Rasch analysis tools and further assess the scalability of Rasch-based assessments across diverse cultural contexts, ensuring continued innovation in measurement science and educational evaluation.

Table 1.
Author citation.

No	Author	Year	Title	Journal	Citation
1	Freitas, et al. [28]	2018	Establishing the HLS-Q12 short version of the European Health Literacy Survey Questionnaire latent trait analyses applying Rasch modelling and confirmatory factor analysis	BMC Health Services Research	78
2	Herrmann-Abell and DeBoer [44]	2017	Investigating a learning progression for energy ideas from upper elementary through high school	Journal of Research in Science Teaching Journal of Research in Science Teaching	64
3	Antonietti, et al. [30]	2023	Development and validation of the ICAP Technology Scale to measure how teachers integrate technology into learning activities	Computers & Education	45
4	Freitas, et al. [32]	2014	Psychometric Properties of the Montreal Cognitive Assessment (MoCA) An Analysis Using the Rasch Model	The Clinical Neuropsychologist	39
5	McKendrick, et al. [41]	2019	Theories and Methods for Labeling Cognitive Workload Classification and Transfer Learning	Frontiers in Human Neuroscience	32

6	Peters, et al. [45]	2021	Construction and validation of a game-based intelligence assessment in minecraft	Computers in Human Behavior	23
7	Rausch, et al. [46]	2016	Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain	Empirical Research in Vocational Education and Training	23
8	Ding [47]	2017	Progression Trend of Scientific Reasoning from Elementary School to University a Large-Scale Cross-Grade Survey Among Chinese Students	International Journal of Science and Mathematics Education	21
9	Förster, et al. [48]	2015	Assessing the financial knowledge of university students in Germany	Empirical Research in Vocational Education and Training	15
10	Blum, et al. [49]	2016	Task difficulty prediction of figural analogies	Intelligence	14

Table 2.
Top cited journal.

No	Journal	Tier	Citation	Year
1	BMC Health Services Research	Q1	78	2018
2	Journal of Research in Science Teaching Journal of Research in Science Teaching	Q1	64	2017
3	Computers & Education	Q1	45	2023
4	The Clinical Neuropsychologist	Q1	39	2014
5	Frontiers in Human Neuroscience	Q2	32	2019
6	Computers in Human Behavior	Q1	23	2021
7	Empirical Research in Vocational Education and Training	Q3	23	2016
8	International Journal of Science and Mathematics Education	Q1	21	2017
9	Empirical Research in Vocational Education and Training	Q3	15	2015
10	Intelligence	Q1	14	2016

5. Discussion and Implications

5.1. Interpretation of Findings

The findings of this systematic review confirm the growing application of the Rasch Model in educational assessment and psychometric validation. Studies have increasingly employed Rasch analysis to ensure measurement invariance and differential item functioning (DIF) assessments, which contribute to more equitable and reliable educational measurements [4]. For example, Werner, et al. [1] demonstrated how Rasch analysis facilitated the evaluation of the Chemistry Self-Concept Inventory, ensuring appropriate item fit and psychometric effectiveness. Similarly, studies integrating DIF analysis, such as Rodríguez, et al. [20] have reinforced the importance of rigorous validation processes in educational measurement.

Despite these advancements, gaps persist in methodological rigor and validation consistency across studies. Some research has displayed insufficient validation steps, including the absence of fit statistics reporting and inadequate item calibration [24]. Such limitations underscore the need for greater transparency and consistency in Rasch-based studies to enhance the reliability of educational assessments. Furthermore, discrepancies in sample size and representativeness remain a recurring issue, potentially biasing estimates of item difficulty and person abilities [50].

5.2. Theoretical Contributions

Theoretical advancements in Rasch measurement theory have significantly contributed to refining psychometric frameworks, particularly in addressing multidimensional constructs and measurement invariance. Rodríguez, et al. [20] highlighted the potential of multidimensional Rasch models in assessing complex educational constructs, moving beyond traditional unidimensional approaches. These theoretical developments offer new perspectives on Rasch applications, allowing researchers to better account for latent traits and underlying response structures.

Additionally, this review aligns with recent discussions on cross-cultural validation, which is critical for ensuring assessment fairness and comparability across diverse educational settings. As global assessments increasingly rely on Rasch modeling (e.g., PISA and TIMSS), the model's adaptability in different cultural contexts remains a vital consideration for future research and policy implementation [51, 52].

5.3. Practical Implications for Educational Research

The findings from this review provide actionable recommendations for educators, policymakers, and researchers aiming to implement Rasch-based assessments effectively. Studies emphasize the necessity of adopting rigorous validation processes by incorporating DIF analysis, item fit statistics, and person reliability measures to enhance assessment precision [4]. Institutional collaboration among universities and educational organizations is essential for developing and validating Rasch-based instruments, as such partnerships improve psychometric quality and enhance generalizability [53]. With the increasing prevalence of online assessments and digital learning platforms, embedding Rasch analysis within adaptive testing frameworks will enable educators to track student progress dynamically and tailor instructional interventions accordingly [50]. Additionally, policies should encourage the cross-cultural adaptation of Rasch-based assessments to minimize bias and mitigate cultural discrepancies in test performance, ensuring fair and equitable assessment outcomes for diverse populations [51].

5.4. Study Limitations

Despite the valuable insights gained from this review, several methodological constraints and biases should be acknowledged. A recurring limitation across reviewed studies is the lack of representative samples, which may limit the generalizability of findings [50]. Some studies fail to report model fit indices, making it difficult to assess the adequacy of Rasch model applications [54]. Issues related to response category inconsistencies and subjective judgments in self-reported assessments present additional sources of bias [55]. Furthermore, the choice of Rasch software, such as Winsteps, RUMM2030, or ConQuest, may influence study outcomes since different programs offer varying model specifications and data processing capabilities [32]. Addressing these limitations in future research will be critical to enhancing the credibility and applicability of Rasch-based educational assessments.

5.5. Future Research Directions

Future research should expand on the integration of AI-driven Rasch analysis and cross-cultural validation to improve measurement accuracy and assessment fairness. Leveraging machine learning algorithms to automate DIF detection, item calibration, and adaptive assessment designs will enhance the efficiency and scalability of Rasch applications [32]. Additional research is needed to examine how Rasch Model applications perform across different linguistic and cultural groups, ensuring fair and equitable assessment interpretations [52]. Comparative studies should evaluate the effectiveness of unidimensional, multidimensional, and mixture Rasch models in different educational assessment contexts, helping to determine which modeling approach best fits various learning environments [20]. Expanding Rasch-based research into emerging fields, such as game-based learning assessments and cognitive neuroscience applications, can further optimize assessment methodologies and contribute to innovative psychometric applications [53].

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2003-2022, 2025
DOI: 10.55214/25768484.v9i6.8310

6. Conclusion

This systematic literature review highlights the growing adoption of the Rasch Model in educational measurement and psychometric research. The findings demonstrate its effectiveness in ensuring measurement validity, detecting differential item functioning (DIF), and supporting multidimensional assessments. The review also identifies key software tools, including Winsteps, RUMM2030, and ConQuest, emphasizing their impact on research outcomes. Additionally, the study underscores the increasing application of the Rasch Model in STEM education, social sciences, and large-scale assessments like PISA and TIMSS.

Despite these advancements, several challenges persist. Notably, methodological inconsistencies, limited cross-cultural validation studies, and variability in model application remain critical concerns. Many studies lack robust sample representation and fail to report model fit indices, which raises concerns about generalizability and measurement precision.

Future research should focus on AI-driven Rasch modeling, cross-cultural validation, and comparative analyses of Rasch model variants to enhance educational assessment frameworks. Expanding interdisciplinary applications and integrating real-time psychometric evaluations will further optimize the Rasch Model's utility in diverse educational settings. Addressing these gaps will ensure the continued evolution of equitable, reliable, and innovative assessment methodologies.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] S. M. Werner, Y. Chen, and M. Stieff, "Examining the psychometric properties of the chemistry self-concept inventory using rasch modeling," *Journal of Chemical Education*, vol. 98, no. 11, pp. 3412-3420, 2021. https://doi.org/10.1021/acs.jchemed.1c00436

[2] W. J. Boone, "Rasch analysis for instrument development: why, when, and how?," *CBE—Life Sciences Education*, vol. 15, no. 4, p. rm4, 2016. https://doi.org/10.1187/cbe.16-04-0148

[3] N. F. Ramli, O. Talib, S. A. Hassan, and U. K. A. Manaf, "Development and validation of an instrument to measure stem teachers' instructional preparedness," *Asian Journal of University Education*, vol. 16, no. 3, pp. 193-206, 2020. https://doi.org/10.24191/ajue.v16i3.11084

[4] D. Adams, K. M. Chuah, A. Mohamed, B. Sumintono, V. Moosa, and M. Shareefa, "Bricks to clicks: Students' engagement in e-learning during the COVID-19 pandemic," *Asia Pacific Journal of Educators and Education*, vol. 36, no. 2, pp. 99-117, 2021. https://doi.org/10.21315/apjee2021.36.2.6

[5] H. Qudratuddarsi, R. Hidayat, N. Nasir, M. K. W. Imami, and R. bin Mat Nor, "Rasch validation of instrument measuring Gen-Z science, technology, engineering, and mathematics (STEM) application in teaching during the pandemic," *International Journal of Learning, Teaching and Educational Research*, vol. 21, no. 6, pp. 104-121, 2022. https://doi.org/10.26803/ijlter.21.6.7

[6] M. Browall, A. H. Pakpour, C. Melin-Johansson, C. L. Hagelin, J. Österlind, and I. Henoch, "Development and psychometric evaluation of a new short version of the swedish frommelt attitudes toward care of the dying scale," *Cancer nursing*, vol. 44, no. 4, pp. 305-313, 2021. https://doi.org/10.1097/ncc.0000000000000794

[7] S. N. Phillipson, S. Phillipson, and M. A. Francis, "Validation of the family educational and learning capitals questionnaire: An Australian perspective," *Journal for the Education of the Gifted*, vol. 40, no. 4, pp. 350-371, 2017. https://doi.org/10.1177/0162353217734373

[8] D. Adams, K. M. Chuah, B. Sumintono, and A. Mohamed, "Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: a Rasch analysis," *Asian Education and Development Studies*, vol. 11, no. 2, pp. 324-339, 2022. https://doi.org/10.1108/aeds-05-2020-0100

[9] H. H.-T. Liu and Y.-S. Lee, "Measuring self-regulation in second language learning: A Rasch analysis," *SAGE Open*, vol. 5, no. 3, p. 2158244015601717, 2015. https://doi.org/10.1177/2158244015601717

[10] S. Kassiavera, A. Suparmi, C. Cari, and S. Sukarmin, "Application of rasch model in two-tier test for assessing critical thinking in physics education," *Journal of Baltic Science Education*, vol. 23, no. 6, pp. 1227-1242, 2024. https://doi.org/10.33225/jbse/24.23.1227

[11] P. Yuanyuan, B. S. Alias, A. N. Mansor, and M. K. M. Nasir, "Validation of the Teachers' Digital Competence Instrument Using Rasch Measurement Model," *Journal of Ecohumanism*, vol. 3, no. 7, pp. 4497–4511, 2024. https://doi.org/10.62754/joe.v3i7.4564

[12] A. Rachmatullah, S. Diana, and M. Ha, "Identifying Indonesian Upper-Secondary School Students' Orientations to Learn Science and Gender Effect through the Use of Structural Equation Modeling," *Journal of Baltic Science Education*, vol. 17, no. 4, pp. 633-648, 2018. https://doi.org/10.33225/jbse/18.17.633

[13] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, 2021. https://doi.org/10.1136/bmj.n71

[14] A. S. Pinto, A. Abreu, E. Costa, and J. Paiva, "How machine learning (ml) is transforming higher education: A systematic literature review," *Journal of Information Systems Engineering and Management*, vol. 8, no. 2, p. 21168, 2023. https://doi.org/10.55267/iadt.07.13227

[15] K. Al-Jaf, C. Öz, H. Mahmud, and T. A. Rashid, "Leveraging chatbots for effective educational administration: A systematic review," 2024. https://doi.org/10.20944/preprints202410.0238.v1

[16] J. Q. Pérez, T. Daradoumis, and J. M. M. Puig, "Rediscovering the use of chatbots in education: A systematic literature review," *Computer Applications in Engineering Education*, vol. 28, no. 6, pp. 1549-1565, 2020. https://doi.org/10.1002/cae.22326

[17] C. North, M. Shortt, M. A. Bowman, and B. Akinkuolie, "How instructional design is operationalized in various industries for job-seeking learning designers: Engaging the talent development capability model," *TechTrends*, vol. 65, pp. 713-730, 2021. https://doi.org/10.1007/s11528-021-00636-2

[18] D. G. Sánchez, R. G. García, I. M. López-Medina, and A. I. Cuesta-Vargas, "Suffering measurement instruments in palliative care: Protocol for a systematic psychometric review," *BMJ open*, vol. 9, no. 4, p. e027524, 2019. https://doi.org/10.1136/bmjopen-2018-027524

[19] Z. Khadjesari, S. Vitoratou, N. Sevdalis, and L. Hull, "Implementation outcome assessment instruments used in physical healthcare settings and their measurement properties: a systematic review protocol," *BMJ open*, vol. 7, no. 10, p. e017972, 2017. https://doi.org/10.1136/bmjopen-2017-017972

[20] M. F. M. Rodríguez, M. Martí-Vilar, M. A. N. Peláez, J. M. G. Lozano, J. P. Martínez-Ramón, and A. Caracuel, "Psychometric properties of the affective dimension of the generic macro-competence assessment scale: Analysis using Rasch model," *Sustainability*, vol. 13, no. 12, p. 6904, 2021.

[21] L. J. Richardson and J. J. Bélanger, "COVID-19 knowledge test: an assessment tool for health educators during the COVID-19 pandemic," *Frontiers in Public Health*, vol. 8, p. 580204, 2020. https://doi.org/10.3389/fpubh.2020.580204

[22] L. A. R. Laliyo, R. Utina, R. Husain, M. K. Umar, M. R. Katili, and C. Panigoro, "Evaluating students' ability in constructing scientific explanations on chemical phenomena," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 19, no. 9, p. em2328, 2023. https://doi.org/10.29333/ejmste/13524

[23] T. Chia Chun, M. A. Samsudin, and M. N. L. Y. Abdullah, "Psychometric properties assessment of postgraduate self-determined learning questionnaire (PSLQ): A mixed method approach," *SAGE Open*, vol. 14, no. 2, p. 21582440241242860, 2024. https://doi.org/10.1177/21582440241242860

[24] Y. Su and L. S. Behar-Horenstein, "Assessment of psychometric properties of an oral health care measure of cultural competence among dental students using rasch partial credit model," *Journal of dental education*, vol. 82, no. 10, pp. 1105-1114, 2018. https://doi.org/10.21815/jde.018.107

[25] R. Primi, "Divergent productions of metaphors: Combining many-facet Rasch measurement and cognitive psychology in the assessment of creativity," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 8, no. 4, pp. 461-471, 2014. https://doi.org/10.1037/a0038055

[26] L. Lim and E. Chapman, "Validation of the Moral Reasoning Questionnaire against Rasch measurement theory," *Journal of Pacific Rim Psychology*, vol. 16, p. 18344909221087418, 2022. https://doi.org/10.1177/18344909221087418

[27] A.-W. Hwang, Y.-T. Chou, C.-L. Hsieh, W.-S. Hsieh, H.-F. Liao, and A. M.-K. Wong, "A developmental screening tool for toddlers with multiple domains based on Rasch analysis," *Journal of the Formosan Medical Association*, vol. 114, no. 1, pp. 23-34, 2015. https://doi.org/10.1016/j.jfma.2014.07.001

[28] S. Freitas *et al.*, "Using the Rasch analysis for the psychometric validation of the Irregular Word Reading Test (TeLPI): A Portuguese test for the assessment of premorbid intelligence," *The Clinical Neuropsychologist*, vol. 32, no. sup1, pp. 60-76, 2018. https://doi.org/10.1080/13854046.2018.1468481

[29] V. Van, D. and B. Csapó, "Effects of multimedia on psychometric characteristics of cognitive tests: A comparison between technology-based and paper-based modalities," *Studies in Educational Evaluation*, vol. 77, p. 101254, 2023. https://doi.org/10.1016/j.stueduc.2023.101254

[30] C. Antonietti, M.-L. Schmitz, T. Consoli, A. Cattaneo, P. Gonon, and D. Petko, "Development and validation of the ICAP Technology Scale to measure how teachers integrate technology into learning activities," *Computers & Education*, vol. 192, p. 104648, 2023. https://doi.org/10.1016/j.compedu.2022.104648

[31] D. Andrich, I. Styles, A. Mercer, and I. B. Puddey, "On the validity of repeated assessments in the UMAT, a high-stakes admissions test," *Advances in Health Sciences Education*, vol. 22, pp. 1245-1262, 2017. https://doi.org/10.1007/s10459-017-9761-6

[32] S. Freitas, G. Prieto, M. R. Simões, and I. Santana, "Psychometric properties of the Montreal Cognitive Assessment (MoCA): An analysis using the Rasch model," *The Clinical Neuropsychologist*, vol. 28, no. 1, pp. 65-83, 2014. https://doi.org/10.1080/13854046.2013.870231

[33] B. Seebacher *et al.*, "German translation, cultural adaption and validation of the unidimensional self-efficacy scale for multiple sclerosis: A study protocol," *BMJ open*, vol. 9, no. 8, p. e029565, 2019. https://doi.org/10.1186/s12883-021-02183-y

[34] G. Hossain and M. Yeasin, "Cognitive ability-demand gap analysis with latent response models," *IEEE Access*, vol. 2, pp. 711-724, 2014. https://doi.org/10.1109/ACCESS.2014.2339328

[35] Y.-C. Yeh, D. Park, S.-Y. Yang, and C.-C. Kuo, "Developing and validating the Taiwan version of the meaningful activity participation assessment (T-MAPA) with Rasch analysis," *BMC geriatrics*, vol. 23, no. 1, p. 159, 2023. https://doi.org/10.1186/s12877-023-03839-9

[36] V. Vattanavanit, S. Ngudgratoke, and P. Khaninphasut, "Validation of response processes in medical assessment using an explanatory item response model," *BMC Medical Education*, vol. 22, no. 1, p. 855, 2022. https://doi.org/10.1186/s12909-022-03942-2

[37] M. C. Silverstein, P. Bjälkebring, B. Shoots-Reinhard, and E. Peters, "The numeric understanding measures: Developing and validating adaptive and nonadaptive numeracy scales," *Judgment and Decision Making*, vol. 18, p. e19, 2023. https://doi.org/10.1017/jdm.2023.17

[38] B. S. Loe and J. Rust, "The perceptual maze test revisited: Evaluating the difficulty of automatically generated mazes," *Assessment*, vol. 26, no. 8, pp. 1524-1539, 2019. https://doi.org/10.1177/1073191117746501

[39] P. Nezhnov, E. Kardanova, M. Vasilyeva, and L. Ludlow, "Operationalizing levels of academic mastery based on Vygotsky's theory: The study of mathematical knowledge," *Educational and Psychological Measurement*, vol. 75, no. 2, pp. 235-259, 2015. https://doi.org/10.1177/0013164414534068

[40] Y. Lee and M. Wilson, "Random item MIRID modeling and its application," *Applied Psychological Measurement*, vol. 41, no. 2, pp. 97-114, 2017. https://doi.org/10.1177/0146621616675835

[41] R. McKendrick, B. Feest, A. Harwood, and B. Falcone, "Theories and methods for labeling cognitive workload: Classification and transfer learning," *Frontiers in human neuroscience*, vol. 13, p. 295, 2019. https://doi.org/10.3389/fnhum.2019.00295

[42] J. Fleissner-Martin, F. X. Bogner, and J. Paul, "Digital vs. analog learning—two content-similar interventions and learning outcomes," *Forests*, vol. 14, no. 9, p. 1807, 2023. https://doi.org/10.3390/f14091807

[43] H. S. Finbråten, B. Wilde-Larsson, G. Nordström, K. S. Pettersen, A. Trollvik, and Ø. Guttersrud, "Establishing the HLS-Q12 short version of the European Health Literacy Survey Questionnaire: latent trait analyses applying Rasch modelling and confirmatory factor analysis," *BMC health services research*, vol. 18, pp. 1-17, 2018. https://doi.org/10.1186/s12913-018-3275-7

[44] C. F. Herrmann-Abell and G. E. DeBoer, "Investigating a learning progression for energy ideas from upper elementary through high school," *Journal of Research in Science Teaching*, vol. 55, no. 1, pp. 68-93, 2018. https://doi.org/10.1002/tea.21411

[45] H. Peters, A. Kyngdon, and D. Stillwell, "Construction and validation of a game-based intelligence assessment in minecraft," *Computers in Human Behavior*, vol. 119, p. 106701, 2021. https://doi.org/10.1016/j.chb.2021.106701

[46] A. Rausch, J. Seifried, E. Wuttke, K. Kögler, and S. Brandt, "Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain," *Empirical Research in Vocational Education and Training*, vol. 8, pp. 1-23, 2016. https://doi.org/10.1186/s40461-016-0035-y

[47] L. Ding, "Progression trend of scientific reasoning from elementary school to university: A large-scale cross-grade survey among Chinese students," *International Journal of Science and Mathematics Education*, vol. 16, no. 8, pp. 1479-1498, 2018. https://doi.org/10.1007/s10763-017-9844-0

[48] M. Förster, S. Brückner, and O. Zlatkin-Troitschanskaia, "Assessing the financial knowledge of university students in Germany," *Empirical Research in Vocational Education and Training*, vol. 7, pp. 1-20, 2015. https://doi.org/10.1186/s40461-015-0017-5

[49] D. Blum, H. Holling, M. S. Galibert, and B. Forthmann, "Task difficulty prediction of figural analogies," *Intelligence*, vol. 56, pp. 72-81, 2016. https://doi.org/10.1016/j.intell.2016.03.001

[50] M. Planinic, W. J. Boone, A. Susac, and L. Ivanjek, "Rasch analysis in physics education research: Why measurement matters," *Physical Review Physics Education Research*, vol. 15, no. 2, p. 020111, 2019. https://doi.org/10.1103/physrevphyseducres.15.020111

[51] M. F. Romero *et al.*, "Measuring COVID-19 health literacy: validation of the COVID-19 HL questionnaire in Spain," *Health and Quality of Life Outcomes*, vol. 20, no. 1, p. 138, 2022. https://doi.org/10.1186/s12955-022-02050-5

[52] L. B. Sousa, G. Prieto, M. Vilar, H. Firmino, and M. R. Simoes, "The adults and older adults functional assessment inventory: A rasch model analysis," *Research on aging*, vol. 37, no. 8, pp. 787-814, 2015. https://doi.org/10.1177/0164027514564469

[53] N. Schott, B. Johnen, and T. J. Klotzbier, "Einschätzung des wohlbefindens von bewohnern in pflegeeinrichtungen: Ubersetzung und validierung der deutschen version des laurens well-being inventory for gerontopsychiatry (LWIG-GER)," *German Journal of Exercise and Sport Research*, vol. 51, pp. 474-486, 2021. https://doi.org/10.1007/s12662-021-00776-w

[54] P. Lambriex-Schmitz, M. R. Van der Klink, S. Beausaert, M. Bijker, and M. Segers, "Towards successful innovations in education: Development and validation of a multi-dimensional Innovative Work Behaviour Instrument," *Vocations and Learning*, vol. 13, pp. 313-340, 2020. https://doi.org/10.1007/s12186-020-09242-4

[55] C. B. Hamilton *et al.*, "Shortening and validation of the Patient Engagement In Research Scale (PEIRS) for measuring meaningful patient and family caregiver engagement," *Health Expectations*, vol. 24, no. 3, pp. 863-879, 2021. https://doi.org/10.1111/hex.13227

DOI: 10.55214/25768484.v9i6.8310