Available online at: http://journal.uny.ac.id/index.php/jpe

Jurnal Prima Edukasia, 13 (1), 59-74

Development of Multi-Tier Diagnostic Tests in Primary Schools: A Systematic Literature Review

Hafizhah Rahmiati Auliya[1], Martha Christianti[2], Benny Hidayat[2]*
[1]Department of Primary Education, Universitas Negeri Yogyakarta
Colombo Street No. 1, Karangmalang, Yogyakarta 55281, Indonesia
[2]Department of Early Childhood Education, Universitas Negeri Yogyakarta
Colombo Street No. 1, Karangmalang, Yogyakarta 55281, Indonesia
*Corresponding Author. E-mail: bennyhidayat.2023@student.uny.ac.id

Received: 11 August 2024; Revised: 11 November 2024; Accepted: 6 January 2025

Abstract: The implementation of multi-tiered diagnostic tests is crucial for identifying and addressing student misconceptions in primary education. This study systematically reviews the development and application of these diagnostic tools in primary schools from 2019 to 2024. Following the PRISMA protocol, 26 studies were selected from an initial 178 records obtained from the Scopus database. The analysis highlights trends in test development, key content areas, and potential future applications in primary mathematics education. The findings emphasize the importance of tailored diagnostic tools for early identification of misconceptions and suggest directions for future research to enhance their effectiveness in educational assessments. The findings are expected to contribute to education in Indonesia by providing tools and strategies tailored to the unique needs of primary school students, thereby enhancing their foundational mathematical understanding and long-term academic success.
Keywords: multi-tier diagnostic test, student misconceptions, diagnostic tool, educational assessment

How to Cite: Auliya, H. R., Christianti, M., & Hidayat, B., (2025). Development of multi-tier diagnostic tests in primary schools: A systematic literature review. *Jurnal Prima Edukasia, 13*(1), 59-74. doi: http://dx.doi.org/10.21831/jpe.v13i1.76906

Introduction

In the process of student interaction with learning content, misconceptions-defined as misunderstandings of concepts caused by ontogenic, didactic, and epistemological obstacles-often emerge and hinder learning. These misconceptions are particularly problematic in foundational education, as they can obstruct the understanding of advanced topics and disrupt cumulative learning processes. (Gurel et al., 2015) explains that "misconception" is the term most commonly used to describe understanding that is contrary to the views of experts (Clement et al., 1989; Driver & Easley, 1978; Helm, 1980). In addition to this term, the literature also recognizes other terms such as "alternative conception" (Klammer, 1998; Palmer & Sarju, 2022; Wandersee et al., 1994), "naive belief" (McCloskey et al., 1980), "children's ideas" (Osborne et al., 1993), "conceptual difficulties" (McDermott, 1993), "primitive phenomenology" (diSessa, 1993), and "mental model" (Greca & Moreira, 2002). The multidimensional nature of misconceptions necessitates diagnostic tools capable of addressing these challenges effectively.

Research shows that misconceptions can hinder student learning and affect linkages to other more complex concepts (Gurel et al., 2015). For example, students who misunderstand fundamental arithmetic operations often struggle with algebraic reasoning and problem-solving, creating persistent gaps in their mathematical competence. Therefore, identifying and addressing misconceptions is crucial for measuring students' level of mastery, tracking their progress, and evaluating the effectiveness of educational programs (Sujinah et al., 2024). Assessment for Learning (AfL) approaches have demonstrated significant positive impacts on mathematics learning by providing feedback to address student misconceptions effectively (Yusron & Sudiyatno, 2021). Recent studies have highlighted the shifts in assessment practices post-COVID-19, where innovative tools and methodologies have been

http://dx.doi.org/10.21831/jpe.v13i1.76906

emphasized to address these educational challenges effectively (Azis et al., 2024). In addition, the identification of misconceptions is also effective in evaluating the competence of teachers and prospective teachers (Osadchuk, 2021; Shamigulova et al., 2022). Research highlights that pre-service mathematics learning for primary school teachers' beliefs about assessment significantly influence their teaching practices and their ability to address student misconceptions effectively (Tamba et al., 2021). Despite the availability of diagnostic tools, their application in mathematics learning for primary school is limited, particularly in addressing common misconceptions such as fractions, place value, and arithmetic strategies. This gap underscores the need for research aimed at developing tools tailored to the needs of younger learners.

Diagnostic tests are one of the tools that can be used to identify student misconceptions. These tests can be in the form of interviews, mind maps, open-ended questions, multiple-choice tests, or graded multiple-choice tests (Gurel et al., 2015; Treagust, 1986). Each type of diagnostic test offers unique advantages and limitations. For instance, open-ended questions allow for detailed exploration of students' thought processes, while multiple-choice tests provide efficiency in identifying common misconceptions across larger groups of students. Educators need to tailor the purpose of using the test to the specific needs in their classroom. The exploration of alternative assessments in mathematics, particularly during uncertain times, highlights the importance of adaptable diagnostic tools to support diverse educational needs (Mahmud et al., 2021). Recent advancements, such as multi-tiered diagnostic tests, have combined multiple perspectives—including confidence levels and reasoning processes—to address the complexities of student misconceptions more effectively (Lai & Chen, 2010; Nainggolan et al., 2022). Ensuring the quality of these diagnostic tools is critical, as research has demonstrated the importance of construct validity and reliability in assessment instruments for primary schools (Otaya et al., 2020) and the alignment of examination questions with quality standards through the Rasch Model (Banawi et al., 2023).

The use of diagnostic tests in education has increased in various fields. These tests not only help identify students' strengths and weaknesses but also allow educators to adjust teaching strategies to improve learning outcomes (Shim et al., 2017). For example, in science education, diagnostic tests have been used to identify misconceptions about Newtonian mechanics and genetics (Lai & Chen, 2010; Lim & Poo, 2021). Similarly, in language education, diagnostic tools are applied to assess students' grammar and vocabulary proficiency (Wibowo & Rosalina, 2019). The application of diagnostic tests in education parallels their use in other fields, such as clinical microbiology, where molecular diagnostics have revolutionized disease detection (Ramanan et al., 2018). These interdisciplinary parallels underscore the importance of precision diagnostics in improving outcomes across domains, including education. However, the transfer of best practices from other fields, such as medical diagnostics, into educational contexts remains limited, highlighting the need for further integration of proven methodologies (Austin-Tse et al., 2022; Hueth et al., 2022; Liu & Zanotti, 2011; Tuut et al., 2022).

Currently, the development of multi-tiered diagnostic tests, including two-tier, three-tier, four-tier, and five-tier formats, has become a trend in educational assessment to more effectively identify student misconceptions across different subjects. Two-tier diagnostic tests, which combine multiple-choice questions with reasoning, are widely used in science education to uncover misconceptions in topics such as genetics, physics and biology (Lai & Chen, 2010; Lim & Poo, 2021; Uyulgan et al., 2014). The three-tier test adds a dimension of student confidence in their answers, which improves accuracy in identifying misconceptions (Nainggolan et al., 2022; Putri et al., 2023). This feature is particularly important for detecting deeply rooted misconceptions, as students often exhibit high confidence in incorrect answers. Meanwhile, four- and five-tier tests assess more detailed aspects, such as reasoning, confidence, and reasoning behind students' answers, thus providing a more comprehensive assessment and aiding in the development of more targeted learning strategies (Habiddin & Page, 2019; Ramadhani & Ermawati, 2021). However, despite these advancements, their application in mathematics learning for primary school remains underexplored, particularly for foundational topics such as fractions, arithmetic operations, and place value.

In mathematics education, multi-tiered diagnostic tests have also been used to assess students' understanding and identify misconceptions that may be hindering their learning. These tests have proven effective in diagnosing challenges in areas such as arithmetic operations, fractions, and algebraic reasoning (Roos et al., 2023; Yang & Lin, 2015). However, most existing research focuses on secondary and higher education levels, leaving a significant gap in the application of these diagnostic tools in mathematics learning for primary school. Research shows that these tests can improve students' arithmetic competence through stepwise intervention models, yet their application in foundational mathematics education remains limited. This research aims to address this gap by focusing on mathematics learning for primary school, a critical stage where early intervention can prevent the entrenchment of misconceptions. The findings are expected to contribute to education in Indonesia by providing tools and strategies tailored to the unique needs of primary school students, thereby enhancing their foundational mathematical understanding and long-term academic success.

Through a systematic review of the literature, this study aims to outline trends in the development of multi-tiered mathematics diagnostic tests, describe the content areas that have been developed, and provide recommendations for further development. Specifically, this study addresses gaps in the application of diagnostic tools for mathematics learning for primary school by focusing on how these tools can be adapted to foundational concepts such as arithmetic operations, fractions, and place value. By exploring these trends, this research seeks to bridge the gap between theoretical advancements in diagnostic testing and practical applications in primary education. The research questions posed in this paper include: (1) what are the trends in the development of multi-tiered diagnostic tests in primary schools? (2) on what variables or materials have multi-tiered diagnostic tests been developed? and (3) what kind of multi-tier math diagnostic tests can be developed in the future? The findings of this study are expected to provide actionable insights for educators and researchers, contributing to the development of more effective diagnostic tools and teaching strategies in Indonesia's primary education system.

Methods

A search for relevant literature on multiple-tier diagnostic tests in primary schools was conducted using the Scopus database. To capture a wide range of studies, keywords such as "multi-tier diagnostic test", "education", "five-tier diagnostic test", "four-tier diagnostic test", "three-tier diagnostic test", and "two-tier diagnostic test" were used. Boolean operators (AND, OR, NOT) were used to narrow and focus the search results, ensuring that the search was broad and topic-specific. The search included literature published from 2019 to March 2024 to look at the overall process of diagnostic test development as well as the latest methodologies in this field.

Inclusion criteria for study selection included: (1) studies published in peer-reviewed journals, to ensure the credibility and reliability of the research; (2) studies focusing on tiered diagnostic tests in educational contexts; (3) studies involving the development, implementation, or evaluation of these diagnostic tools; and (4) articles written in English to maintain language consistency and accessibility. Conversely, the exclusion criteria eliminated: (1) studies that did not focus on school environments, as the focus was on educational contexts; (2) articles that had not been peer-reviewed or opinion articles, to maintain high standards of academic rigor; (3) studies that lacked full text, as access to full studies was necessary for a thorough evaluation; and (4) articles published in languages other than English, to avoid language barriers and ensure comprehensive understanding.

The selection process involved initial screening of titles and abstracts by two independent reviewers to assess relevance to the topic of graded diagnostic tests in primary schools. Articles deemed relevant then underwent full-text review to ensure eligibility based on established inclusion and exclusion criteria. Any differences between reviewers during this phase were resolved through discussion, with a third reviewer consulted when necessary to reach consensus.

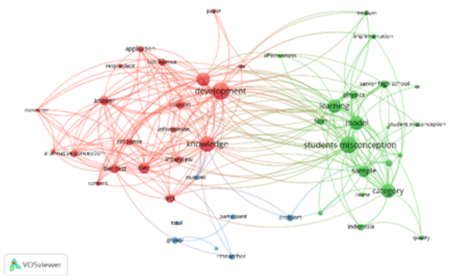

Figure 1. Initial Network Visualization

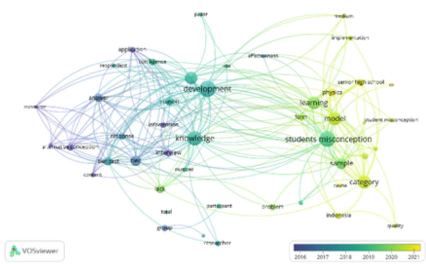

Figure 2. Initial Overlay Visualization

Data extraction was performed using a standardized form (see Table 1) to maintain consistency and completeness across all included studies. The form captured important information such as study characteristics (authors, year of publication, and journal), study design, sample population, specific interventions or diagnostic tools used, outcomes measured, and key findings. This structured approach facilitates the systematic collection of relevant data for analysis.

The extracted data was analyzed using VOSviewer to create thematic visualizations and identify key trends, clusters, and interdependencies within the dataset. Figure 1 shows the initial network visualization, highlighting the thematic organization within the dataset, including various clusters of research-related terms, educational methodologies, and participant demographics. Figure 2 presents the initial overlay visualization, adding a temporal dimension with a color gradient indicating the average publication year of the terms, showing the evolution of research themes over time.

Based on the mapping of 47 keywords scattered in 3 clusters and referring to the occurrence and total link strength on VOSviewer, it was found that the themes most related to multiple-tier diagnostic tests were development (occurrence: 47, total link strength: 302), student misconceptions (occurrence: 46, total link strength: 229), knowledge (occurrence: 42, total link strength: 271), categories (occurrence: 39, total link strength: 204), models (occurrence: 38, total link strength: 227), and topics (occurrence: 37, total link strength: 242). These keywords hint at what the multiple-tier diagnostic test is all about.

The PRISMA flowchart (Figure 3) provides a detailed overview of the systematic review process, illustrating the stages of study identification, screening, eligibility assessment and inclusion. Initially, 178 studies were identified, which were then screened down to 26 studies that met the inclusion criteria. This diagram highlights the rigorous selection methodology used to ensure the inclusion of high-quality and relevant research studies.

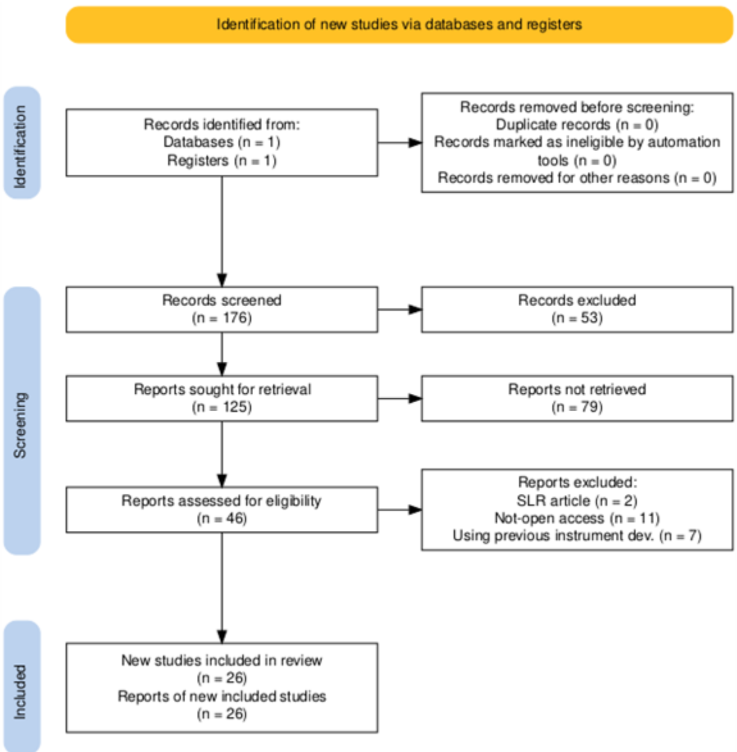

Figure 3. PRISMA Flow Diagram for Systematic Review (Haddaway et al., 2022)

This methodical approach ensures that the selected studies are relevant, of high quality, and provide a comprehensive picture of the development and implementation of tiered diagnostic tests in primary schools. Using standardized procedures for data search, selection and extraction, this systematic review aims to provide robust and insightful conclusions that can guide future research and practice in educational assessment.

Results and Discussion

In the results and discussion section, this study focuses on three main subtopics related to multiple-tier diagnostic tests. First, the trend of multiple-tier diagnostic tests is analyzed to highlight their growing adoption as a powerful tool for understanding and addressing students' misconceptions in depth. This aligns with the research objective of identifying the temporal, geographical, and disciplinary distribution of these tests. Second, the study examines the specific variables and materials developed within the context of multiple-tier diagnostic tests, providing a detailed overview of their application in educational research. This subtopic is critical in addressing the research objective of evaluating the design and implementation of these tests in various contexts. Finally, potential future developments in multiple-tier diagnostic tests are discussed, particularly in mathematics, to emphasize their role in enhancing students' conceptual understanding and procedural skills through more accurate diagnostic tools and targeted interventions. This discussion is closely tied to the objective of proposing recommendations for future test designs in mathematic learning for primary school.

Table 1. Finding 26 Articles from the Scopus Database

No.	Article	Country	Year	Study	Level	Grade	Sample
1	(Retone & Prudent, 2023)	Philippines	2023	Biology	3-tier	college	109
2	(Nasir et al., 2023)	Indonesia	2023	Physics	2-tier	college	81
3	(Aksoy & Erten, 2022)	Turkey	2022	Physics	4-tier	college	401
4	(Kiray & Simsek, 2021)	Turkey	2021	Physics	4-tier	college	470
5	(Halim et al., 2021)	Indonesia	2021	Physics	3-tier	college	281
6	(Muhajir, 2021)	Indonesia	2021	Geography	4-tier	college	147
7	(Nurfalah et al., 2020)	Indonesia	2020	Math	2-tier	college	35
8	(Andariana et al., 2020)	Indonesia	2020	Biology	3-tier	college	128
9	(Liampa et al., 2019)	Greece	2019	Science	3-tier	college	219
10	(Yang & Sianturi, 2019)	Taiwan	2019	Math	3-tier	elementary school	125
11	(Anam et al., 2019)	Indonesia	2019	Science	5-tier	elementary school	69
12	(Suparman et al., 2024)	Indonesia	2024	Chemistry	5-tier	senior high school	580
13	(López-Garduza et al., 2024)	Mexico	2024	Chemistry	4-tier	senior high school	8
14	(Astuti et al., 2023)	Indonesia	2023	Physics	4-tier	senior high school	270
15	(Jumadi et al., 2023)	Indonesia	2023	Physics	4-tier	senior high school	241
16	(Istiyono et al., 2023)	Indonesia	2023	Physics	4-tier	senior high school	700
17	(Murni et al., 2022)	Indonesia	2022	Chemistry	2-tier	senior high school	137
18	(Lim & Poo, 2021)	Malaysia	2021	Biology	2-tier	senior high school	500
19	(Hadinugrahaningsih et al., 2020)	Indonesia	2020	Chemistry	2-tier	senior high school	127
20	(Métioui & Trudel, 2020)	Canada	2020	Physics	2-tier	senior high school	25
21	(Wardani et al., 2020)	Taiwan	2020	Chemistry	3-tier	senior high school	68
22	(Prodjosantoso et al., 2019)	Indonesia	2019	Chemistry	3-tier	senior high school	56
23	(Madlazim & Izzah, 2019)	Indonesia	2019	Physics	4-tier	senior high school	32
24	(Ribič et al., 2024)	Slovenia	2024	Chemistry	3-tier	junior high school	503
25	(Yeo et al., 2022)	Taiwan	2022	Biology	3-tier	junior high school	106
26	(Majer et al., 2019)	Slovenia	2019	Chemistry	3-tier	junior high school	1012

The Trend of Multiple-tier Diagnostic Test Development

The trend of Multi-Tier Diagnostic Test development in the field of education demonstrates a growing significance, supported by the analysis of metrics represented in Figures 4, 5, and 6. Figure 4 highlights the distribution of research across various subject areas, with Physics (36%) and Chemistry (32%) emerging as the most studied disciplines. This trend underscores the strong interest in these core scientific subjects, driven by their complex concepts that require detailed diagnostic tools to accurately address misconceptions and learning gaps. Biology (12%), Mathematics (8%), and General Science (8%) have also shown notable research activity, reflects the expansion of the application of Multi-Tier Diagnostic Tests in various fields of educational science. However, Geography (4%), as an underrepresented area, illustrates the need for broader exploration to leverage these tools across diverse educational contexts. The identification of trends in the development of these diagnostic assessments highlights the potential for extending their use into less explored disciplines, such as Mathematics and General Science.

Figure 5 illustrates the geographical distribution of research on Multi-tier Diagnostic Tests, with the largest concentration found in Indonesia (56%). This dominance indicates that researchers and educational institutions in Indonesia place significant emphasis on the development and application of these diagnostic tools. This focus may be attributed to supportive educational policies, adequate research funding, and heightened awareness of the need for diagnostic tools to address student misconceptions in the region. Significant contributions also come from countries such as Slovenia (8%), Turkey (8%), and Taiwan (8%), reflecting their active involvement in global efforts to improve educational practices through diagnostic innovations. Meanwhile, smaller contributions from countries such as Greece, Canada, Malaysia, the Philippines, and Mexico (4% each) demonstrate growing international interest, although the distribution remains uneven. This pattern highlights the expanding international collaboration in educational research while underscoring the potential to extend the use of these diagnostic tools across diverse educational contexts worldwide.

Figure 6 illustrates the distribution of publication years, highlighting the temporal trends in the development of Multi-Tier Diagnostic Tests. The data reveals a peak in research activity in 2019 (24%), followed by consistent contributions in 2020 (20%), 2021 (16%), and 2022 (8%). The surge in 2019 reflects heightened interest and innovation in this field, possibly driven by emerging educational challenges and advancements in diagnostic methodologies. However, the decline observed in 2024 (12%) may signify a shift in research priorities, potentially towards newer or alternative educational technologies and assessment methods. Despite these fluctuations, the steady volume of research over the years underscores the sustained relevance of Multi-Tier Diagnostic Tests in educational research. These tools remain critical in identifying and addressing student misconceptions, supporting educators in improving learning outcomes through evidence-based interventions.

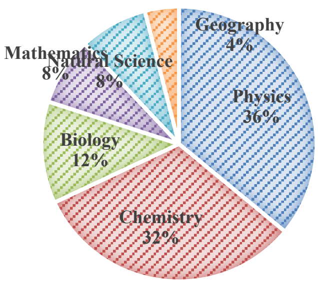

Figure 4. Field Study

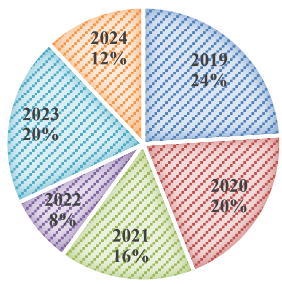

Figure 5. Publication Year

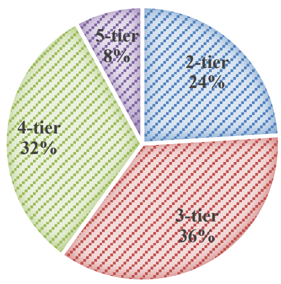

Figure 6. Geo. Distribution

Variables/contents Have Multiple-tier Diagnostic Tests Been Developed On

Figure 7 illustrates the distribution of study designs employed in Multi-Tier Diagnostic Test research, offering insights into the complexity and depth of these diagnostic tools. The 3-tier design emerges as the most utilized, accounting for 36% of studies. This widespread use can be attributed to its balanced approach, which provides detailed diagnostics without being overly complex or difficult to administer. The 4-tier design, comprising 32% of studies, reflects a preference for more comprehensive diagnostic tools aimed at pinpointing specific misconceptions and learning gaps with greater precision. Meanwhile, the 2-tier design, at 24%, demonstrates the continued reliance on simpler diagnostic formats due to their ease of implementation and broad applicability across educational contexts. The 5-tier design, representing just 8% of studies, highlights its limited adoption, likely due to the increased complexity and resource demands associated with these detailed assessments. This distribution underscores the diversity of diagnostic test designs, each tailored to address specific educational needs. The prevalence of 3- and 4-tier designs suggests a balance between diagnostic precision and practicality, while the lower adoption of 5-tier designs signals the challenges of integrating highly detailed tools into everyday educational practices.

Figure 8 provides an analysis of the sample sizes utilized in Multi-Tier Diagnostic Test studies, offering insights into the robustness and generalizability of their findings. The majority of studies (56%) employed sample sizes between 6 and 208, reflecting a preference for smaller, more manageable research groups that allow for detailed, individualized analyses. Such sample sizes are particularly suitable for in-depth studies exploring specific variables or educational contexts. A significant portion of studies (20%) used sample sizes ranging from 209 to 408, indicating an effort to strike a balance between detailed exploration and broader applicability of findings. Moderately larger sample sizes, between 409 and 610, accounted for 16% of the studies, demonstrating a moderate inclination towards generalizing findings to wider populations. The smallest proportions (4% each) were observed in the ranges of 611 to 810 and 811 to 1012, suggesting that very large sample sizes are less common due to the logistical and analytical challenges they entail. This distribution highlights the diverse methodological approaches in sample size selection, with a predominance of small-scale studies designed for specificity and precision. The variation in sample sizes reflects the adaptability of Multi-Tier Diagnostic Test research to different educational needs and contexts.

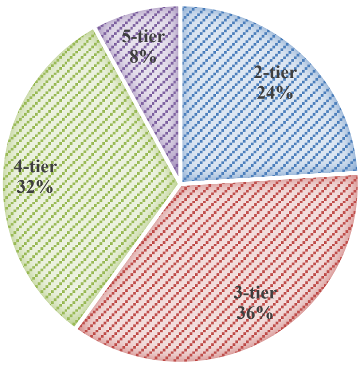

Figure 7. Study Design

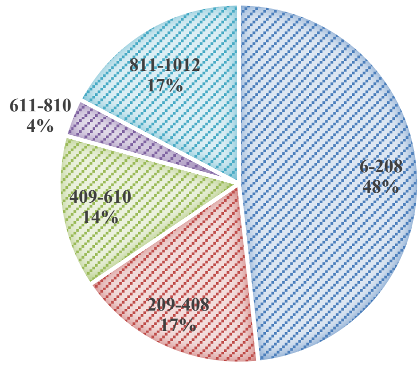

Figure 8. Sample Size

Figure 9 illustrates the distribution of educational levels in Multi-Tier Diagnostic Test research, revealing a predominant focus on secondary education, which constitutes 46.2% (12 studies), followed by higher education (PT) at 34.6% (9 studies). This highlights a strong emphasis on addressing misconceptions in subjects such as physics and chemistry, commonly taught at these levels. In contrast, middle school and elementary school are underrepresented, accounting for only 11.5% (3 studies) and 7.7% (2 studies), respectively. The limited focus on earlier educational levels underscores a critical gap in research, where foundational learning in mathematics and science is vital for long-term academic success. Expanding research efforts to include younger students is essential to develop diagnostic tools that effectively address misconceptions early, preventing compounded learning difficulties in later stages.

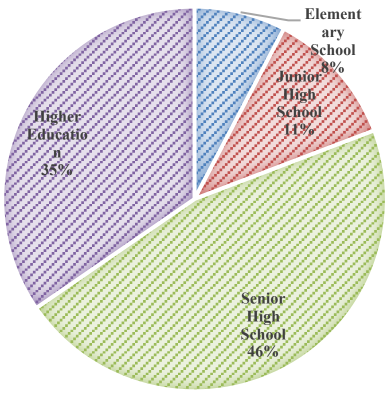

Figure 9. Education Level

A total of 26 studies on the development of Multi-Tier Diagnostic Tests highlight a diverse range of interventions tailored to address student misconceptions across various disciplines and educational levels. These interventions, which predominantly leverage advanced technology platforms, employ two to five levels to provide nuanced insights into students' conceptual understanding. For instance, a five-tier computer-based chemistry diagnostic test developed by (Suparman et al., 2024) demonstrates how digital tools can enhance the accuracy and accessibility of assessments. Similarly, (Astuti et al., 2023) introduced an Android-based four-tier physics test app, emphasizing the shift towards mobile technology for practical and scalable diagnostic purposes. This trend underscores the growing reliance on technology to improve the precision of diagnostic tools and their implementation in diverse educational contexts, aligning with the need to make assessments both efficient and widely accessible. However, the limited application of such tools in foundational education remains a significant gap that warrants further exploration.

Despite the diversity of subjects, the central goal of Multi-Tier Diagnostic Tests is consistent: identifying and remediating student misconceptions to enhance learning outcomes. Studies such as those by (Aksoy & Erten, 2022), (Astuti et al., 2023), (Halim et al., 2021), (Jumadi et al., 2023), and (Suparman et al., 2024) demonstrate the use of these diagnostic tools across disciplines, utilizing two to five tiers to pinpoint specific areas of misunderstanding. For instance, (Ariyani & Rusilowati, 2023) and (Banawi et al., 2022) highlight how Multi-Tier Diagnostics can effectively identify gaps in physics and chemistry learning, while (Fikri et al., 2023) and (Wulandari et al., 2018) showcase similar applications in biology and environmental science. Improving the performance of public elementary schools through evidence-based interventions, including the use of diagnostic tools, has been emphasized as a key strategy to enhance educational outcomes (Kamaludin, 2023). The complexity of the tools and their applicability vary across educational levels, with higher-tier diagnostics predominantly used for advanced topics in secondary and higher education. Conversely, studies in foundational science and mathematics often rely on fewer tiers to address basic misconceptions. This adaptability across subjects and levels underscores the crucial role of Multi-Tier Diagnostics in enhancing teaching strategies. By offering detailed insights into specific misconceptions, these tools empower educators to design targeted interventions that directly address student learning challenges.

The review revealed that research on multiple-tier diagnostic tests with a focus on mathematics remains scarce, with only a few notable contributions in this domain. For instance, (Yang & Sianturi, 2019) developed a four-tier diagnostic test specifically designed to evaluate mathematical reasoning, targeting and addressing misconceptions in students' problem-solving approaches. Similarly, (Wardani et al., 2020) employed a three-tier diagnostic test utilizing triangulation analysis to assess deductive reasoning in mathematics. These studies demonstrate the potential of Multi-Tier Diagnostics to uncover intricate misconceptions in mathematical learning processes. However, the limited number of studies highlights the urgent need for further research to design comprehensive diagnostic tools specifically tailored to mathematical concepts, particularly at the primary education level. Expanding research in this area is essential to address misconceptions early, enabling targeted interventions that build strong foundational mathematical understanding.

Multiple-tier Mathematics Diagnostic Tests Can Be Developed in the Future

A review of studies on Multi-Tier diagnostic tests highlights a significant gap in the development of these tools for mathematic learning for primary school. Although extensive research and interventions have been conducted at higher levels of education and in disciplines such as chemistry and physics, the focus on younger students and foundational mathematical concepts remains limited. For instance, (Bhakti et al., 2022) demonstrated the use of four-tier diagnostic tests in thermodynamics to identify misconceptions effectively, while (Suparman et al., 2024) developed a computer-based five-tier diagnostic test for chemical concepts. However, comprehensive tools tailored to the unique needs of primary education are still lacking. In the context of mathematics, (Wardani et al., 2020) explored the role of inductive and deductive approaches in enhancing reasoning skills, and (Yang & Sianturi, 2019) utilized three-tier diagnostic tests to assess students' conceptual understanding. These studies emphasize the need for more targeted research and development of diagnostic tools that address misconceptions in mathematics learning for primary school (Halimah & Wibowo, 2024; Awaliyah et al., 2024).

The development of tiered mathematics diagnostic tests for primary education can leverage foundational concepts outlined in Elementary and Middle School Mathematics by Van de Walle, Karp, and Bay-Williams (Walle et al., 2016). This seminal work highlights key areas such as early number concepts, number sense, the meaning of operations, and basic fact fluency. By designing diagnostic tests that focus on these core areas, educators can pinpoint specific misconceptions and address gaps in young students more effectively (Nurhikmah & Wibowo, 2024; Santoso et al., 2024; Efendi et al., 2024). For instance, a progressive diagnostic test could begin by assessing basic number recognition and gradually move to more complex tasks like comparing and ordering numbers, providing a detailed understanding of students' grasp of place value. Similarly, tailored tests can evaluate students' arithmetic operation skills, beginning with simple addition and subtraction and advancing to explore the properties and relationships inherent in these operations. Thematic-integrative assessment approaches, as demonstrated in previous studies, underscore the importance of aligning diagnostic tools with curriculum goals to ensure practicality and relevance in classroom settings (Prasetyo, 2017; Mustadi et al., 2024). Such a stepwise approach ensures targeted interventions, fostering a deeper and more accurate understanding of foundational mathematical concepts.

In addition to early number concepts and operations, Multi-Tier diagnostic tests can be expanded to include measurement and data concepts, geometry, spatial understanding, and algebraic thinking, which are fundamental to building a comprehensive mathematical foundation. For example, diagnostic tests in measurement could begin with basic concepts like length and weight, gradually advancing to more intricate ideas such as volume and area. Similarly, geometry diagnostics could evaluate students' understanding of shapes, their properties, and spatial relationships, while algebraic thinking diagnostics could focus on identifying patterns, relationships, and functions. Additionally, the assessment of psychomotor skills has gained attention, especially during distance learning, where accurate evaluation of multidimensional competencies posed unique challenges (Ambarwati et al., 2022; Harokah et al., 2024; Toding & Wibowo, 2024). The integration of digital resources, such as science digital libraries, has shown significant potential in enhancing literacy and supporting diagnostic tools by providing accessible and interactive content (Prastiti & Adi, 2024; Wibowo et al., 2024). By integrating these broad content areas into a multi-level diagnostic tool, educators can gain a detailed understanding of students' mathematical competencies and misconceptions. This approach not only enables more targeted and effective teaching strategies tailored to the diverse needs of elementary students but also bridges existing gaps in educational research (Brilatin & Wibowo, 2024; Viantorus et al., 2024). Consequently, it contributes to enhancing the overall quality of mathematics education at the primary school level, fostering stronger foundational skills that support long-term academic success.

Conclusion

Multi-tier diagnostic tests have emerged as essential tools in education for identifying and addressing student misconceptions. These tests, ranging from two-level to five-level formats, provide nuanced insights into student understanding, enabling educators to adapt teaching strategies effectively. The findings of this study highlight significant research activity in disciplines such as physics, chemistry, and mathematics, with Indonesia playing a leading role in advancing these diagnostic tools. However, a critical gap persists in the application of multi-tier diagnostic tests for mathematics learning for primary school, particularly in foundational concepts. Addressing this gap, future research should prioritize the development of diagnostic tools tailored to early mathematical concepts such as number sense, operations, and geometry. These tools will empower educators to detect misconceptions early and deliver targeted interventions, enhancing the overall quality of mathematics education in primary schools.

References

Aksoy, A. C. A., & Erten, S. (2022). A four-tier diagnostic test to determine pre-service science teachers' misconception about global warming. *Journal of Baltic Science Education*, *21*(5), 747–761. https://doi.org/10.33225/jbse/22.21.747

Ambarwati, D., Herwin, H., & Dahalan, S. C. (2022). How elementary school teachers assess students' psychomotor during distance learning? *Jurnal Prima Edukasia*, *10*(1), 58–65. https://doi.org/10.21831/jpe.v10i1.45040

Anam, R. S., Widodo, A., Sopandi, W., & Wu, H. K. (2019). Developing a five-tier diagnostic test to identify students' misconceptions in science: An example of the heat transfer concepts. *Elementary Education Online*, *18*(3), 1014–1029. https://doi.org/10.17051/ilkonline.2019.609690

Andariana, A., Zubaidah, S., Mahanal, S., & Suarsini, E. (2020). Identification of biology students' misconceptions in human anatomy and physiology course through three-tier diagnostic test. *Journal for the Education of Gifted Young Scientists*, *8*(3), 1071–1085. https://doi.org/10.17478/JEGYS.752438

Ariyani, A., & Rusilowati, A. (2023). Identification and remediation of misconceptions on thermodynamics law for students with interactive demonstration. Dalam S. null, S. K., C. E., M. null, L. C.P., & A. S. (Ed.), *AIP Conference Proceedings* (Vol. 2614). American Institute of Physics Inc. https://doi.org/10.1063/5.0125922

Astuti, I. A. D., Bhakti, Y. B., Prasetya, R., & Zulherman. (2023). Android-based 4-tier physics test app to identify student misconception profiles. *International Journal of Evaluation and Research in Education*, *12*(3), 1356–1363. https://doi.org/10.11591/ijere.v12i3.25536

Austin-Tse, C. A., Jobanputra, V., Perry, D. L., Bick, D., Taft, R. J., Venner, E., Gibbs, R. A., Young, T., Barnett, S., Belmont, J. W., Boczek, N., Chowdhury, S., Ellsworth, K. A., Guha, S., Kulkarni, S., Marcou, C., Meng, L., Murdock, D. R., Rehman, A. U., … Rehm, H. L. (2022). Best practices for the interpretation and reporting of clinical whole genome sequencing. *npj Genomic Medicine*, *7*(1), 27. https://doi.org/10.1038/s41525-022-00295-z

Awaliyah, N. P., Hastuti, W. S., Wibowo, S. E., & Hidayat, P. (2024). The effect of problem based learning model on students' critical thinking ability. *MIMBAR PGSD Undiksha*, *12*(1), 101–107. https://doi.org/10.23887/jjpgsd.v12i1.62280

Azis, D. K., Sarah, S., Febriana, M., & Fajarudin, S. (2024). Post-covid-19 trends in elementary science education: Dominant methods, media , and assessment practices. *Jurnal Prima Edukasia*, *12*(2), 300–314. https://journal.uny.ac.id/index.php/jpe/article/view/75735

Banawi, A., Rumasoreng, M. I., Basta, I., & Latuconsina, A. (2023). Are the final semester examination questions for basic science concepts made by lecturers have the quality according to the rasch model? *Jurnal Prima Edukasia*, *11*(1), 81–95. https://doi.org/10.21831/jpe.v11i1.53633

Banawi, A., Sopandi, W., Kadarohman, A., & Solehuddin, M. (2022). Five-tier multiple-choice diagnostic test development: Empirical evidence to improve students' science literacy. *Proceedings of the International Conference on Madrasah Reform 2021 (ICMR 2021)*, *633*. https://doi.org/10.2991/assehr.k.220104.020

Bhakti, Y. B., Astuti, I. A. D., & Prasetya, R. (2022). Four-tier thermodynamics diagnostic test (4t-tdt) to identify student misconception. *KnE Social Sciences*. https://doi.org/10.18502/kss.v7i14.11958

Brilatin, A., & Wibowo, S. E. (2024). Elementary teacher leadership of pioneer teacher program alumnae in Yogyakarta city. *Jurnal Prima Edukasia*, *12*(1), 129–139. https://doi.org/10.21831/jpe.v12i1.69044

Clement, J., Brown, D. E., & Zietsman, A. (1989). Not all preconceptions are misconceptions: Finding 'anchoring conceptions' for grounding instruction on students' intuitions. *International Journal of Science Education*, *11*(5), 554–565. https://doi.org/10.1080/0950069890110507

Sessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, *10*(2–3), 105–225. https://doi.org/10.1080/07370008.1985.9649008

Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education*, *5*(1), 61–84. https://doi.org/10.1080/03057267808559857

Efendi, U. P., Wibowo, S. E., Sartono, E. K. E., & Lena, M. S. (2024). The effectiveness of differentiated instruction in elementary school subject: Teachers' perceptions. *Jurnal Pendidikan Progresif*, *14*(3), 1784–1799. https://doi.org/10.23960/jpp.v14.i3.2024121

Fikri, A. A., Hasanah, I. U., Ahsani, E. L. F., Hanik, E. U., & Musdalifah, M. (2023). Pre-service biology teacher's misconception about ecological succession. Dalam S. A., U. R., S. M., & P. F.G. (Ed.), *AIP Conference Proceedings* (Vol. 2595). American Institute of Physics Inc. https://doi.org/10.1063/5.0123819

Greca, I. M., & Moreira, M. A. (2002). mental, physical, and mathematical models in the teaching and learning of physics. *Science Education*, *86*(1), 106–121. https://doi.org/10.1002/sce.10013

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Habiddin, & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, *18*(2), e1230. https://doi.org/10.1002/cl2.1230

Hadinugrahaningsih, T., Andina, R. E., Munggaran, L. R., & Rahmawati, Y. (2020). Analysis of students' alternative conceptions about electrolyte and non-electrolyte solutions using a two-tier diagnostic test for chemistry teaching improvement. *Universal Journal of Educational Research*, *8*(5), 1926–1934. https://doi.org/10.13189/ujer.2020.080529

Halim, A., Mahzum, E., Yacob, M., Irwandi, I., & Halim, L. (2021). The impact of narrative feedback, e-learning modules and realistic video and the reduction of misconception. *Education Sciences*, *11*(4). https://doi.org/10.3390/educsci11040158

Halimah, S., & Wibowo, S. E. (2024). Optimization of science learning outcomes for series and parallel electrical circuits through project-based learning methods for grade VI elementary school. *Journal of Education*, *9*(1). https://journal.stkipsingkawang.ac.id/index.php/JETL/article/view/6414

Harokah, S., Wibowo, S. E., Sudigdo, A., Yudianto, A., & Wulansari, I. Y. (2024). The effectiveness of problem-based learning methods in improving prospective elementary school teachers' critical thinking skills. *Al Ibtida: Jurnal Pendidikan Guru MI*, *11*(1). https://doi.org/10.24235/al.ibtida.snj.v11i1.14104

Helm, H. (1980). Misconceptions in physics amongst South African students. *Physics Education*, *15*(2), 92–105. https://doi.org/10.1088/0031-9120/15/2/308

Hueth, K. D., Prinzi, A. M., & Timbrook, T. T. (2022). Diagnostic stewardship as a team sport: interdisciplinary perspectives on improved implementation of interventions and effect measurement. *Antibiotics*, *11*(2), 250. https://doi.org/10.3390/antibiotics11020250

Istiyono, E., Dwandaru, S.B.W., Fenditasari, K., Ayub, M. R. S. S. N., & Saepuzaman, D. (2023). The development of a four-tier diagnostic test based on modern test theory in physics education. *European Journal of Educational Research*, *12*(1), 371–385. https://doi.org/10.12973/eu-jer.12.1.371

Jumadi, J., Sukarelawan, M. I., & Kuswanto, H. (2023). An investigation of item bias in the four-tier diagnostic test using Rasch model. *International Journal of Evaluation and Research in Education*, *12*(2), 622–629. https://doi.org/10.11591/ijere.v12i2.22845

Kamaludin, K. (2023). How to improve the performance of public elementary schools? An empirical evidence from Indonesia. *Jurnal Prima Edukasia*, *11*(2), 235–246. https://doi.org/10.21831/jpe.v11i2.60290

Kiray, S. A., & Simsek, S. (2021). Determination and evaluation of the science teacher candidates' misconceptions about density by using four-tier diagnostic test. *International Journal of Science and Mathematics Education*, *19*(5), 935–955. https://doi.org/10.1007/s10763-020-10087-5

Klammer, J. (1998). An overview of techniques for identifying, acknowledging and overcoming alternate conceptions in physics education. *Klingenstein Project Report, Teachers College-Columbia University.*, 40.

Lai, A. F., & Chen, D. J. (2010). Web-based two-tier diagnostic test and remedial learning experiment. *International Journal of Distance Education Technologies*, *8*(1), 31–53. https://doi.org/10.4018/jdet.2010010103

Liampa, V., Malandrakis, G. N., Papadopoulou, P., & Pnevmatikos, D. (2019). Development and evaluation of a three-tier diagnostic test to assess undergraduate primary teachers' understanding of ecological footprint. *Research in Science Education*, *49*(3), 711–736. https://doi.org/10.1007/s11165-017-9643-1

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, *10*(2), 241–252. https://doi.org/10.15294/jpii.v10i2.26944

Liu, J. H., & Zanotti, K. M. (2011). Management of the adnexal mass. *Obstetrics and Gynecology*, *117*(6), 1413–1428. https://doi.org/10.1097/AOG.0b013e31821c62b6

López-Garduza, F., Díaz, M. H. R., & Cabral-Rosetti, L. G. (2024). Engineering professors' conceptions on the conceptual field of electrostatics in Mexico. *International Journal of Innovation in Science and Mathematics Education*, *31*(6), 33–45. https://doi.org/10.30722/IJISME.31.06.003

Madlazim, & Izzah, N. (2019). Identification of students' misconception on Newton's law of gravitation concept using the four-tier diagnostic test instrument. *Journal of Science Education*, *20*(1).

Mahmud, M. N., Retnawati, H., Yusron, E., & Rachmayanti, E. (2021). Exploring the alternative assessment on Mathematical instruction in an era of uncertainty. *Jurnal Prima Edukasia*, *9*(2). https://doi.org/10.21831/jpe.v9i2.39212

Majer, J., Slapničar, M., & Devetak, I. (2019). Assessment of the 14- And 15-year-old students' understanding of the atmospheric phenomena. *Acta Chimica Slovenica*, *66*(3), 659–667. https://doi.org/10.17344/acsi.2019.5087

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science*, *210*(4474), 1139–1141. https://doi.org/10.1126/science.210.4474.1139

McDermott, L. C. (1993). Guest Comment: How we teach and how students learn—A mismatch? *American Journal of Physics*, *61*(4), 295–298. https://doi.org/10.1119/1.17258

Métioui, A., & Trudel, L. (2020). Unipolar reasoning in electricity: developing a digital two-tier diagnostic test. *WSEAS Transactions on Electronics*, *11*, 86–95. https://doi.org/10.37394/232017.2020.11.11

Muhajir, A. (2021). Relationship teaching experience and educator certificates against pedagogical competencies. *Review of International Geographical Education Online*, *11*(6), 866–874. https://doi.org/10.48047/rigeo.11.06.104

Murni, H. P., Azhar, M., Ellizar, E., Nizar, U. K., & Guspatni, G. (2022). Three levels of chemical representation-integrated and structured inquiry-based reaction rate module: Its effect on students' mental models. *Journal of Turkish Science Education*, *19*(3), 758–772. https://doi.org/10.36681/tused.2022.148

Mustadi, A., Wibowo, S. E., Zubaidah, E., Supartinah, S., Sugiarsih, S., & Sayekti, O. M. (2024). Needs analysis of project based teaching module development in the independent curriculum. *Jurnal Prima Edukasia*, *12*(1), 52–60. https://doi.org/10.21831/jpe.v12i1.66933

Nainggolan, Y. N., Permadani, K. G., & Prajoko, S. (2022). Analysis of student's misconceptions on the material of the immune system using a three-tier diagnostic test. *JPBIO (Jurnal Pendidikan Biologi)*, *7*(2), 158–166. https://doi.org/10.31932/jpbio.v7i2.1681

Nasir, M., Cari, Sunarno, W., & Rahmawati, F. (2023). The development of a two-tier diagnostic test for student understanding of light. *Journal of Turkish Science Education*, *20*(1), 150–172. https://doi.org/10.36681/tused.2023.009

Nurfalah, E., Arofah, I., Yuniwati, I., Haslinah, A., & Lestari, D. R. (2020). Construction a diagnostic test in the form of two-tier multiple choice on calculus material. *Mathematics and Statistics*, *8*(5), 577–582. https://doi.org/10.13189/ms.2020.080512

Nurhikmah, I., & Wibowo, S. E. (2024). Improving mathematics learning outcomes on integer materials through think pair and share learning models. *Journal of Education*, *9*(1). https://journal.stkipsingkawang.ac.id/index.php/JETL/article/view/6415

Osadchuk, O. L. (2021). Psychological mechanisms of student teachers' professional reliability training. *Education and Self Development*, *16*(3), 115–126. https://doi.org/10.26907/esd.16.3.11

Osborne, J. F., Black, P., & Meadows, J. (1993). Young children's (7-11) ideas about light and their development. *International Journal of Science Education*, *15*(1), 83–93. https://doi.org/10.1080/0950069930150107

Otaya, L. G., Kartowagiran, B., & Retnawati, H. (2020). The construct validity and reliability of the lesson plan assessment instrument in primary schools. *Jurnal Prima Edukasia*, *8*(2), 126–134. https://doi.org/10.21831/jpe.v8i2.33275

Palmer, A. L., & Sarju, J. P. (2022). Inclusive outreach activity targeting negative alternate conceptions of chemistry. *Journal of Chemical Education*, *99*(5), 1827–1837. https://doi.org/10.1021/acs.jchemed.1c00400

Prasetyo, T. (2017). Development of learning outcomes assessment tools in integrative thematic learning for grade V elementary school. *Jurnal Prima Edukasia*, *5*(1), 102–111. https://doi.org/10.21831/jpe.v5i1.7528

Prastiti, M. R., & Adi, B. S. (2024). Science digital library: innovation in improving science literacy of elementary school students. *Jurnal Prima Edukasia*, *12*(2), 216–227. DOI: https://journal.uny.ac.id/index.php/jpe/article/view/69478

Prodjosantoso, A. K., Hertina, A. M., & Irwanto. (2019). The misconception diagnosis on ionic and covalent bonds concepts with three tier diagnostic test. *International Journal of Instruction*, *12*(1), 1477–1488. https://doi.org/10.29333/iji.2019.12194a

Putri, O. D. A., Prastowo, T., & Sanjaya, I. G. M. (2023). Profile of students' misconceptions on substance pressure using a three-tier diagnostic test. *IJORER : International Journal of Recent Educational Research*, *4*(1), 1–15. https://doi.org/10.46245/ijorer.v4i1.267

Ramadhani, N. N., & Ermawati, F. U. (2021). Five-tier diagnostic test instrument for uniform circular motion concepts: Development, validity, reliability and limited trials. *Jurnal Pendidikan Fisika*, *9*(1). https://doi.org/10.26618/jpf.v9i1.4763

Ramanan, P., Bryson, A. L., Binnicker, M. J., Pritt, B. S., & Patel, R. (2018). Syndromic panel-based testing in clinical microbiology. *Clinical Microbiology Reviews*, *31*(1). https://doi.org/10.1128/CMR.00024-17

Retone, L. E., & Prudent, M. S. (2023). Assessing undergraduates' misconceptions on central dogma of molecular biology using a 3-tier diagnostic test. *Journal of Sustainability Science and Management*, *18*(10), 150–160. https://doi.org/10.46754/jssm.2023.10.010

Ribič, L., Devetak, I., & Slapničar, M. (2024). Assessing 15-year-olds' Understanding of chemical concepts in the context of the lithosphere and pedosphere. *Acta Chimica Slovenica*, *71*(1), 84–90. https://doi.org/10.17344/acsi.2024.8610

Roos, H., Fälth, L., Karlsson, L., Nilvius, C., Selenius, H., & Svensson, I. (2023). Promoting basic arithmetic competence in early school years–using a response to intervention model. *Journal of Research in Special Educational Needs*, *23*(4), 313–322. https://doi.org/10.1111/1471-3802.12602

Santoso, A., Wibowo, S. E., & Raharja, S. (2024). Improving the mathematics ability using quantum learning method for elementary students. *Journal of Education*, *9*(1). https://journal.stkipsingkawang.ac.id/index.php/JETL/article/view/6475

Shamigulova, O. A., Kalimullina, G. I., & Yangirova, V. M. (2022). Diagnostics of methodological competencies of a teacher. *Proceedings of the International Conference «Social and Cultural Transformations in the Context of Modern Globalism» (SCTCMG 2022), 19-21 April, 2022, Grozny, Chechen Republic, Russia*, *128*, 553–560. https://doi.org/10.15405/epsbs.2022.11.76

Shim, G. T. G., Shakawi, A. M. H. A., & Azizan, F. L. (2017). Relationship between students' diagnostic assessment and achievement in a pre-university mathematics course. *Journal of Education and Learning*, *6*(4), 364. https://doi.org/10.5539/jel.v6n4p364

Sujinah, Isnah, E. S., Jalis, F. M. M., & Kharis, M. (2024). Utilizing cognitive diagnostic assessments to identify and address student needs in differentiated classrooms. *Journal of Higher Education Theory and Practice*, *24*(1), 94–100. https://doi.org/10.33423/jhetp.v24i1.6763

Suparman, A. R., Rohaeti, E., & Wening, S. (2024). Development of computer-based chemical five-tier diagnostic test instruments: a generalized partial credit model. *Journal on Efficiency and Responsibility in Education and Science*, *17*(1), 92–106. https://doi.org/10.7160/eriesj.2024.170108

Tamba, K. P., Cendana, W., & Adegbite, A. (2021). Pre-service primary school mathematics teachers' beliefs on mathematics assessment. *Jurnal Prima Edukasia*, *9*(2). https://doi.org/10.21831/jpe.v9i2.36815

Toding, D. A., & Wibowo, S. E. (2024). Enhancing learning outcomes in science subjects through demonstration method for fifth grade at inpres seringgu elementary school. *Journal of Education*, *9*(1). https://journal.stkipsingkawang.ac.id/index.php/JETL/article/view/6421

Treagust, D. (1986). Evaluating students' misconceptions by means of diagnostic multiple choice items. *Research in Science Education*, *16*(1), 199–207. https://doi.org/10.1007/BF02356835

Tuut, M. K., Burgers, J. S., van der Weijden, T., & Langendam, M. W. (2022). Do clinical practice guidelines consider evidence about diagnostic test consequences on patient-relevant outcomes? A critical document analysis. *Journal of Evaluation in Clinical Practice*, *28*(2), 278–287. https://doi.org/10.1111/jep.13619

Uyulgan, M. A., Akkuzu, N., & Alpat, Ş. (2014). Assessing the students' understanding related to molecular geometry using a two-tier diagnostic test. *Journal of Baltic Science Education*, *13*(6), 839–855. https://doi.org/10.33225/jbse/14.13.839

Viantorus, V. A. V., Wibowo, S. E., & Firdaus, F. M. (2024). Enhancing fourth-grade elementary students mathematical reasoning skills through the SAVI model. *Jurnal Pendidikan Progresif*, *14*(3), 2076–2087. https://doi.org/10.23960/jpp.v14.i3.2024141

Walle, J. A. Van de, Karp, K. S., Bay-Williams, J. M., McGarvey, L. M., Sandra, F., & Wray, J. (2016). Elementary and middle school mathematics: Teaching developmentally. Dalam *Pearson Education, Inc.* (Vol. 10, Nomor 5). ERIC.

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research in alternative conceptions in science: Part II learning. in D. L. Gabel (Ed.), *Handbook of Research on Science Teaching and Learning* (hlm. 177–210). Macmillan.

Wardani, S., Kusuma, I. W., Liu, S. T., & Harjito. (2020). Comparison of learning in inductive and deductive approach to increase student's conceptual understanding based on international standard curriculum. *Jurnal Pendidikan IPA Indonesia*, *9*(1), 70–78. https://doi.org/10.15294/jpii.v9i1.21155

Wibowo, S. E., Mustadi, A., Supartinah, Sugiarsih, S., & Sayekti, O. M. (2024). Development of project based teaching module to increase students' creativity in designing Indonesian language learning. *Journal of Ecohumanism*, *3*(7). https://doi.org/10.62754/joe.v3i7.4632

Wibowo, S., & Rosalina, S. (2019). Pragmatic view on the inscription heritage of Tarumanegara kingdom. *Proceedings of the First International Conference on Advances in Education, Humanities, and Language, ICEL 2019, Malang, Indonesia, 23-24 March 2019*. First International Conference on Advances in Education, Humanities, and Language, ICEL 2019, Malang, Indonesia, 23-24 March 2019, Malang, Indonesia. https://doi.org/10.4108/eai.23-3-2019.2284952

Wulandari, P. S., Cari, C., Aminah, N. S., & Nugraha, D. A. (2018). Pre-service teachers' conceptual understanding of rolling friction coefficient. Dalam S. M.A. & N. D.A. (Ed.), *AIP Conference Proceedings* (Vol. 2014, hlm. 020060). American Institute of Physics Inc. https://doi.org/10.1063/1.5054464

Yang, D. C., & Lin, Y. C. (2015). Assessing 10- to 11-year-old children's performance and misconceptions in number sense using a four-tier diagnostic test. *Educational Research*, *57*(4), 368–388. https://doi.org/10.1080/00131881.2015.1085235

Yang, D. C., & Sianturi, I. A. J. (2019). Assessing students' conceptual understanding using an online three-tier diagnostic test. *Journal of Computer Assisted Learning*, *35*(5), 678–689. https://doi.org/10.1111/jcal.12368

Yeo, J.-H., Yang, H.-H., & Cho, I.-H. (2022). Using a three-tier multiple-choice diagnostic instrument toward alternative conceptions among lower-secondary school students in Taiwan: Taking ecosystems unit as an example. *Journal of Baltic Science Education*, *21*(1), 69–83. https://doi.org/10.33225/jbse/22.21.69

Yusron, E., & Sudiyatno, S. (2021). How is the impact of Assessment for Learning (AfL) on mathematics learning in elementary schools? *Jurnal Prima Edukasia*, *9*(1). https://doi.org/10.21831/jpe.v9i1.34865