International Journal of Multicultural and Multireligious Understanding

http://ijmmu.com
editor@ijmmu.com
ISSN 2364-5369
Volume 12, Issue 5
April, 2025
Pages: 225-239

Diagnostic Test Research Trends In Science Education: A Systematic Review

Nurdiyanti Nurdiyanti; Supahar Supahar; Siti Maryam Ulfa; Irfandi Rudin

Department of Physics Education, Universitas Negeri Yogyakarta, Indonesia

http://dx.doi.org/10.18415/ijmmu.v12i5.6679

Abstract

This study examines research trends related to diagnostic tests in science education from 2018 to 2023, emphasizing their role in enhancing educational quality. The study employs a systematic literature review (SLR) following PRISMA guidelines, utilizing the Scopus database as the primary source. A total of 673 articles were initially identified, and 53 met the inclusion criteria for further evaluation. The findings underscore the critical role of diagnostic tests in identifying misconceptions, improving conceptual understanding, and assessing students' reasoning skills and affective dimensions. Research trends indicate a notable increase in publications on this topic in 2023. Further analysis using VOSviewer reveals that misconceptions are the most frequently associated aspect with diagnostic tests, emphasizing the importance of accurate diagnostic assessments in educational research.

Keywords: *Diagnostic Test; Science Education; PRISMA; Systematic Literature Review; VOSviewer*

Introduction

Science learning is an integral part of education that involves understanding scientific concepts. To produce good science learning, quality assessments are needed. Assessment holds very important power in the field of education (Day et al., 2018) because it is not only a tool for measuring student learning progress, but also an instructional instrument that influences the teaching and learning process. Assessments in science education not only measure students' knowledge but must also be able to evaluate students' understanding of scientific concepts and their ability to apply them in real-life situations (Osborne & Dillon, 2014). Assessment preferences are also related to learning strategies and orientations (Birenbaum, 1997), thus they must be carefully structured for science education. Furthermore, the enhancement of educational quality through well-designed assessments is crucial in the context of sustainable development. By ensuring that students not only grasp scientific concepts but also develop the ability to apply them in real-world scenarios (Bramwell-Lalor, 2019), assessments can foster a generation equipped to address global challenges, thereby contributing to the goals of sustainable development.

Educational assessment is a big agenda at both international and national levels. Implementation of assessments such as PISA (Program for International Student Assessment), PIRLS (Progress in International Reading Literacy Study), and TIMSS (Trends in International Mathematics and Science Studies) continues to develop and encourage the emergence of assessment innovation at the national level. The progression of the assessment revealed several deficiencies in assessment practices, including extended delays between test administration and feedback, the limited utility of summative test results for personalized intervention, and a general absence of student-specific feedback (Csapó & Molnár, 2019). This has an impact on learning patterns that are only test-oriented and test score inflation (Koretz, 2017), as well as the harmful influence on the educational atmosphere in schools and stress on educators (von der Embse et al., 2016).

Existing deficiencies have prompted a shift in learning assessment patterns. Assessment now focuses more on formative assessments or diagnostic assessments rather than summative assessments (Csapó & Molnár, 2019). There is no clear-cut distinction between diagnostic and formative assessments, nor is there a universally accepted definition of diagnostic assessment. However, diagnostic assessments are typically characterized by their focus on identifying problems, exploring potential difficulties, and determining whether students are prepared for learning tasks, thereby measuring prerequisite knowledge. Additionally, diagnostic assessments are often accompanied by interventions, such as compensatory instruction to address identified obstacles and the provision of various supportive activities (e.g., in mathematics, which facilitates data-driven decision-making (Brendefur et al., 2018), or in reading (Filderman et al., 2018)).

To increase the effectiveness of science learning, the role of diagnostic tests is increasingly recognized as a critical tool in identifying and responding to specific student learning needs. Diagnostic tests or assessments are tools used to identify students' learning difficulties (Miller et al., 2009). Apart from that, diagnostic assessments also indicate whether students have a correct understanding of a concept or not (Istiyono et al., 2023), and uncover students' misconceptions and the reasons behind their misconceptions of scientific understanding (Gierl, 2007). A teacher can use diagnostic tests to find errors in answers given by students so that they can plan efforts to correct students' knowledge problems.

Several studies have been conducted regarding assessment in physics education. Studies that focus on the use of four-tier tests to detect misconceptions in physics education (Çelikkanlı & Kızılcık, 2022) Other research on diagnostic tests also discusses more general aims to evaluate research related to diagnostic tests for educational assessment, possible future developments, and the impact of implementing evaluation in education (Maulana et al., 2023). Meanwhile, the research provides information about general topics that students often experience in science misconceptions, and diagnostic assessments used to identify student misconceptions in science and provide a comparison of each instrument along with its weaknesses and strengths which are reviewed. Diagnostic tests are believed to have many uses besides determining misconceptions in students (Soeharto et al., 2019). However, currently there is not much research on diagnostic tests in science learning. Therefore, this research will reveal the various benefits of diagnostic tests in science learning so that it is hoped that it can provide a comprehensive picture of diagnostic tests for future educational development. Research questions in this study include: (a) What types of diagnostic tests are used in science learning? (b) What is the role of diagnostic tests in science learning?

Method

This research is a systematic literature review (SLR). This research method aims to minimize systematic error (bias) by identifying, evaluating, and synthesizing all relevant studies, regardless of design, to address a specific question or set of questions (Petticrew & Roberts, 2006). In addition, SLR can benefit the field of education by providing better information on practice and identifying important new directions for research (Borrego et al., 2014). Researchers followed PRISMA guidelines to ensure that this systematic literature review was systematic as follows (Page et al., 2021; Sun et al., 2023).

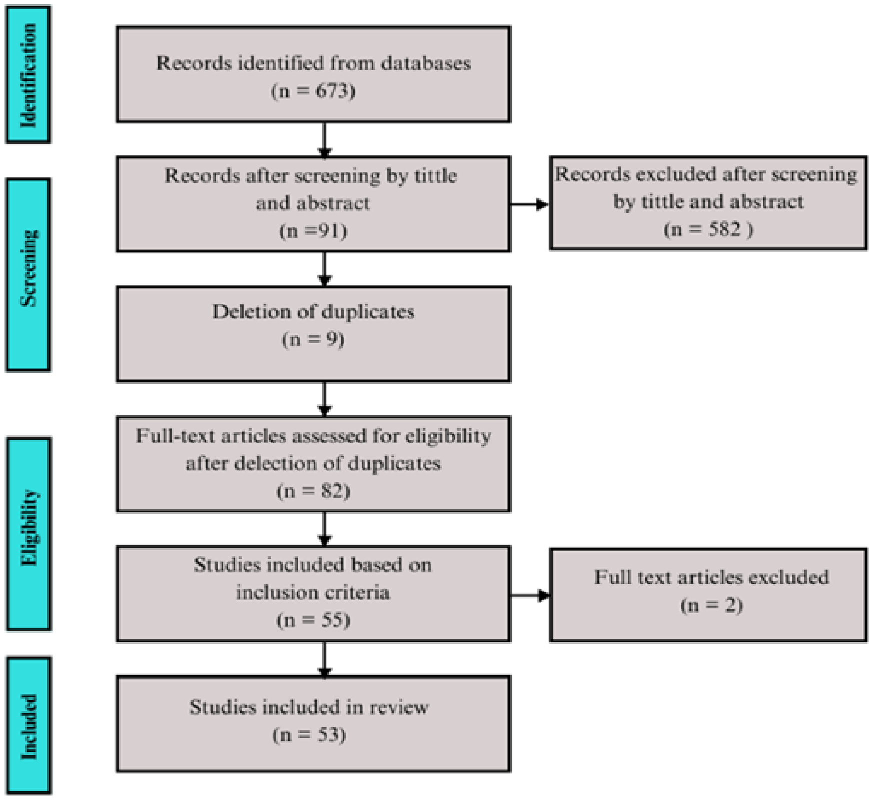

Figure 1. Flowchart Illustrating the Review Selection Process

The electronic scientific database used to identify relevant research is Scopus. The search process was completed in early November 2023. The search terms selected based on their relevance to the research objectives included: diagnostic tests (“diagnostic test” OR “diagnostic assessment” OR “diagnose test” OR “diagnose assessment”); education (“science education” OR “science learning” OR “science teaching” OR “science study”); and multitier (“two tier” OR “three-tier” OR “four tier” OR “five tier” OR “multitier”). Literature searches in scientific databases were carried out using the PRISMA method (Page et al., 2021; Sun et al., 2023). This procedure is divided into four stages: (1) identification, (2) screening, (3) eligibility, and (4) included as in Figure 1. The first stage in this research was a systematic search of electronic databases that met the inclusion criteria. These criteria include: a) this research is empirical research related to diagnostic tests in the field of education, b) the subject matter in the research includes science learning in physics, chemistry, biology, and mathematics. c) participants involved from elementary school, middle school to university level. d) research published between 2018 and 2023. e) research written in English. Researchers excluded studies that examined special populations, such as students with intellectual disabilities, and studies that were not directly related to the use of diagnostic tests.

This search process yielded 673 articles. Subsequently, the second stage entailed reviewing the titles and abstracts of chosen research articles. Altogether, this screening process yielded 91 pertinent articles, with 9 duplicates identified upon comprehensive scrutiny of the screening outcomes. The third stage involved evaluating the appropriateness of the articles, resulting in 82 articles being selected. This assessment involved thorough reading of the articles to determine their adherence to the specified inclusion criteria: (a) the research includes information relevant to various types of diagnostic tests, (b) this research includes information related to the benefits of diagnostic tests in learning, (c) The research explains the research methods and instruments, as well as various findings from the use of diagnostic tests. Next, the researchers considered the 53 articles extracted in the final stage.

The final selected articles were thoroughly reviewed and codified. Coding is done by collecting similar articles to answer research questions. First, researchers grouped articles according to year of publication and country of research location. Second, researchers identified research findings in these articles such as the type of diagnostic test used and the benefits of the diagnostic test. Third, researchers use tools to help analyze research trends regarding diagnostic tests using the Vosviewer application. Finally, the researcher revealed recommendations for future research from each article reviewed.

Results and Discussion

The results of searching for articles in the Scopus database produced 53 articles and a codification process was carried out. After that, researchers analyzed the contents of each article more deeply to process it into data that was easy to understand. First, the articles reviewed from 2018-2023 show the distribution in Figure 2.

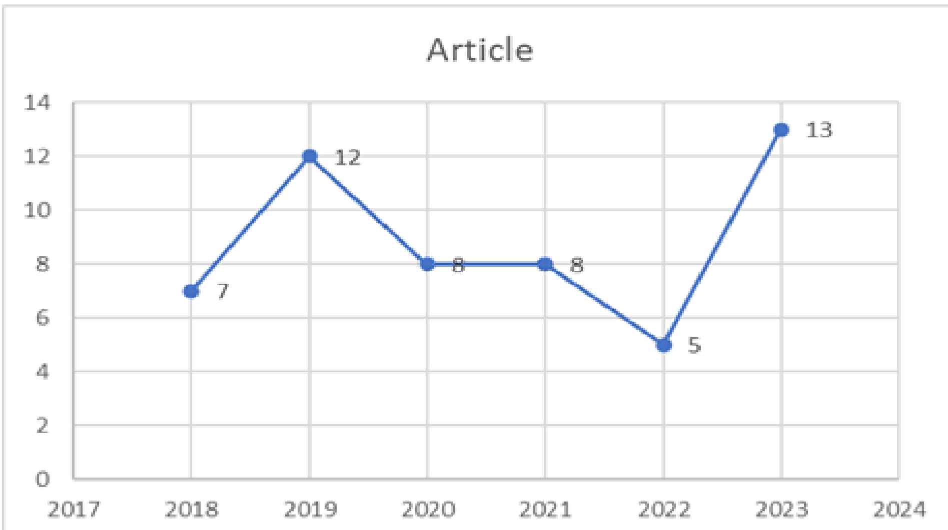

Figure 2. Classification of Publication Years of Diagnostic Tests

Research on diagnostic tests has a fluctuating frequency every year. In the last 5 years, the research trend shows a significant increase in 2023 followed by 2019. The lowest number of published articles was in 2022, with one distinct study (Phanphech et al., 2022) examines the impact of student anxiety resulting from online learning in different learning environments, namely asynchronous (YouTube) and synchronous (Zoom) learning environments and aims to compare students' conceptual understanding of electrical circuits.

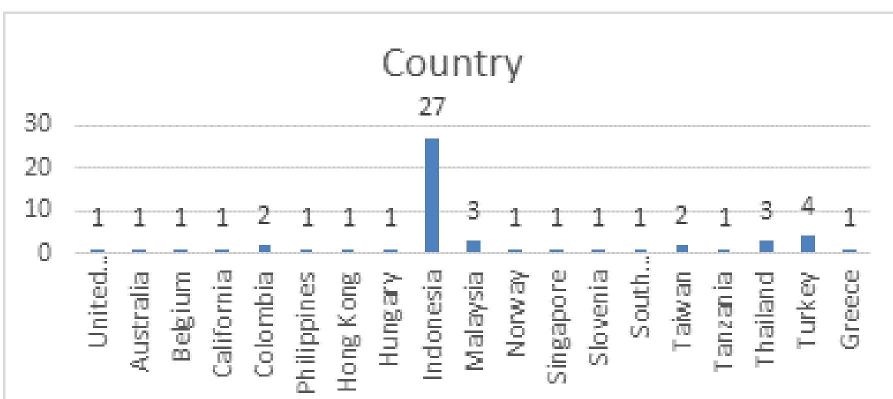

Figure 3. Distribution of Research Location Countries

The largets distribution of research location regarding diagnostic test is in Indonesia. Numerous studies in Indonesia focus on diagnostic assessments because the Ministry of Education and Culture has instructed teachers to conduct these assessments before starting lessons, particularly during the pandemic. This highlights the importance of diagnostic assessments for teachers and encourages researchers to explore the process and use of feedback from these assessments from the teachers' perspective (Yusron et al., 2024).

The next analysis involves mapping related articles using the VOSviewer application. VOSviewer is a powerful tool designed to create and visualize bibliometric networks, allowing researchers to identify key trends, influential authors, and significant topics within a given field (van Eck & Waltman, 2010). In this study, VOSviewer was employed to map the research related to diagnostic tests, providing a comprehensive overview of the research landscape. The results of this mapping, which are illustrated in Figure 4, help in understanding the relationships and connections between various studies, facilitating more informed and targeted research efforts.

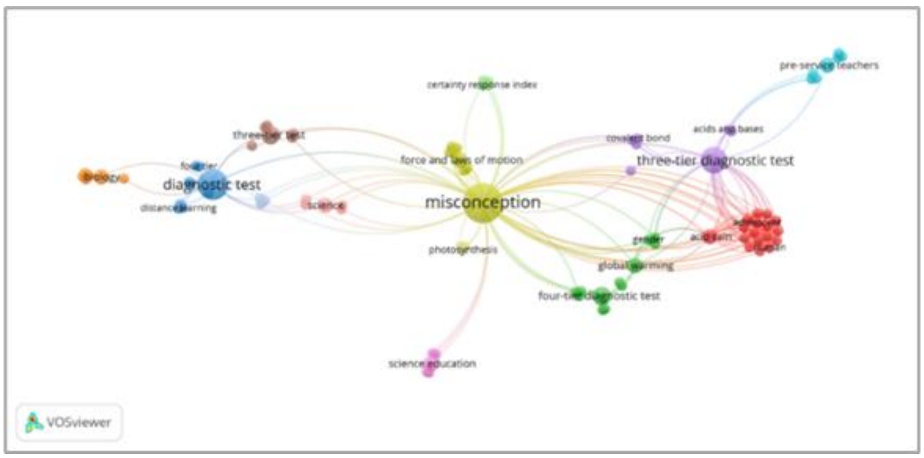

Figure 4. Mapping from VosViewer Related Diagnostic Tests

Based on network analysis from Vosviewer, it shows that misconceptions are the aspect most related to diagnostic tests. These misconception keywords have a strong relationship with various types of diagnostic tests such as multitier diagnostic tests (two-tier, three-tier, four-tier). In addition, there is a range of participants from elementary to middle school students, university students, and pre-service teachers. Furthermore, articles about diagnostic tests in Vosviewer show the field of science education studies from various materials in biology, physics and chemistry. The distribution of publications according to subject matter is shown in the following Table 1:

Table 1. Subjects addressed by diagnostic tests

Subjects	Frequency
Biology	6
Chemistry	7
Mathematics	3
Physics	28
Science	11

Types on Diagnostics Tests

Based on the results of this study, various types of diagnostic tests were found to be used. Each diagnostic test comes from diverse research methodologies, including instrument development studies for three-tier, four-tier, five-tier, six-tier, and two-tier diagnostic test types. These different tiers represent varying levels of complexity and depth in assessing students' understanding, providing a more nuanced and detailed evaluation of their learning progress and misconceptions. Such variety in diagnostic tests highlights the ongoing efforts to improve educational assessment tools and tailor them to specific educational needs and contexts.

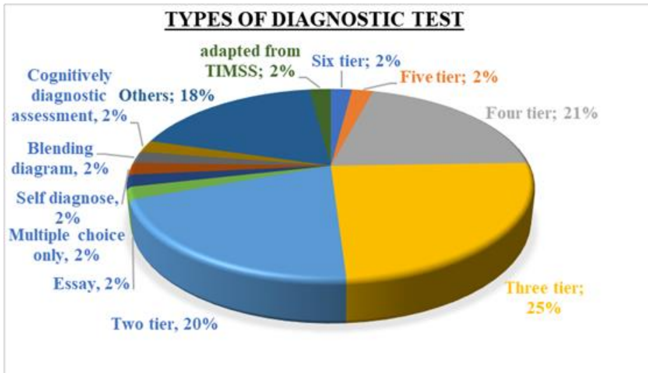

Figure 5. Diagram Types of Diagnostic Test

Based on the distribution of this data, the type of diagnostic test instrument used most is the three-tier diagnostic test (n=12). The next most common types of research (n=10) used were four-tier diagnostic tests and two-tier diagnostic tests. Furthermore, the diagnostic tests used once each include: essay, multiple choice only, self-diagnosis, blending diagram, cognitively diagnostic assessment, diagnostic adapted from TIMSS, six tier, and five-tier diagnostic tests. Apart from that, there are 9 other types of studies.

Role of Diagnostic Test in Science Education

The studies that have been reviewed are then identified based on the role of the diagnostic test used. The author classifies these studies into several groups, each focusing on different purposes of diagnostic tests: identifying misconceptions, understanding concepts, assessing reasoning skills, evaluating attitude aspects, and determining class placement before learning. This classification helps to highlight the diverse applications and benefits of diagnostic tests in educational research. By analyzing the distribution of diagnostic test use, as shown in Table 2 below, researchers can gain insights into how these tests are utilized to address specific educational needs and challenges. This systematic categorization not only provides a clearer understanding of the existing research landscape but also underscores the importance of tailored diagnostic assessments in improving educational outcomes and guiding instructional practices.

Table 2. Distribution of the Role of Diagnostic Tests in Science Education

Aspect	Role	References	Freq.
Miscon-ception	Detecting/identifying/exploring misconceptions	(Aksoy & Erten, 2022; Anam et al., 2019; Andariana et al., 2020; Astuti et al., 2023; Hadinugrahaningsih et al., 2020; Imaduddin et al., 2023; Istiyani et al., 2018; Kaniawati et al., 2019; Kiray & Simsek, 2021; Laeli et al., 2023; Lim & Poo, 2021; Mubarokah et al., n.d.; Nahadi, 2022; Prodjosantoso et al., 2019; Soeharto & Csapó, 2021; Suprapto et al., 2018; Taban & Kiray, 2022; Wancham et al., 2023b, 2023a)	24
	Identify students at risk of low achievement	(Kang & Yoon, 2022)	
	Tested reducing levels of misconceptions using narrative feedback, e-learning modules, and realistic videos.	(Djudin, 2021; Halim et al., 2021)	
Student understan-ding	Identify conceptual understanding	(Adimayuda et al., 2020; Hadinugrahaningsih et al., 2020; Istiyono et al., 2023; Kahraman, 2019; Majer et al., 2019; Mbwile et al., 2023; Nasir et al., 2023; Phanphech et al., 2022; Ping et al., 2020; Sopandi & Sukardi, 2020; Yeo et al., 2022)	15
	Identify differences in understanding of the genders	(Kahraman, 2019)	
	Identify various student practices, knowledge, and beliefs in all three aspects of computational literacy (material,cognitive, and social aspects)	(Odden et al., 2019)	
	Diagnose HOTS	(Ramadhan et al., 2019)	
Affective	Understand student beliefs and attitudes	(Miranda & Rada, 2018)	4
	Knowing the development of students' disciplinary literacy	(Chubko et al., 2019)	
	Understand students' beliefs and attitudes, determine the development of students' disciplinary literacy, identify creative thinking skills	(Batlolona et al., 2019)	
	Testing the effect of collaborative mind mapping	(Fung & Liang, 2023)	
Student reasoning	Diagnosing difficulties in students' reasoning.	(Van Den Eynde et al., 2020)	2
	To diagnose unipolar reasoning in electricity.	(Métioui & Trudel, 2020)	
Placement test	For physics placement, a predictor of student success in learning	(Burkholder et al., 2021)	1
Others		(Canlas, 2021; Irmak et al., 2023; Jumadi et al., 2023; Nurfalah et al., 2020; Safadi & Saadi, 2021)	5

The studies reviewed are classified based on the role of the diagnostic test used. This classification helps to highlight the diverse applications and benefits of diagnostic tests in educational research. By analyzing the distribution of diagnostic test use, as shown in Table 2 below, researchers can gain insights into how these tests are utilized to address specific educational needs and challenges. The objectives of the successive studies are: identifying misconceptions (47%), understanding conceptual knowledge (29.4%), evaluating affective aspects (7.8%), assessing student reasoning (3.9%), placement tests (1.98%), and others/not specifically studied (9.8%). This distribution reflects the primary focus areas within the field of diagnostic assessments and underscores the importance of a comprehensive approach to evaluating student learning and instructional effectiveness. This systematic categorization not only provides a clearer understanding of the existing research landscape but also emphasizes the value of tailored diagnostic assessments in enhancing educational outcomes and guiding teaching practices.

Research often aims to identify student misconceptions, such as studies focused on developing Android-based assessment applications using four-tier diagnostic test instruments to detect physics misconceptions among State Senior High School students. This research resulted in a product that includes a test question grid, test questions, answer sheets, instructions for answering questions, answer keys, assessment guidelines, and guidelines for interpreting results. The test questions are structured into four levels: questions with one correct answer and three distractors, and a measure of the confidence in the answers. This app can identify students' physics misconceptions from grade 10 to grade 12 in physics subjects. Another study researched the development of diagnostic assessments for teachers to identify students at risk of experiencing low achievement by collecting data from a total of 355 responses from primary and lower secondary school students (Kang & Yoon, 2022). Students' perceptions of most items align with the level assessment criteria, reflecting micro-level learning progress. However, some inconsistencies offer insights for adjusting learning progress to better match students' thinking patterns. Additionally, students exhibit learning challenges that necessitate additional support. Meanwhile, research has identified various profiles of students' misconceptions about buoyancy, noting a 54.71% decrease in these misconceptions following treatment (Djudin, 2021). Significant conceptual changes regarding buoyancy were observed among students after the intervention, indicating that the treatment was highly effective. (mean gain = 0.82).

Further analysis in the use of diagnostic tests to identify conceptual understanding. The Four-Tier Diagnostic Test (FTDT) Based on the Modern Test Theory instrument they developed can describe high school students' conceptual understanding of physics (Istiyono et al., 2023). Conducting exploratory studies on literacy has yielded a framework that can be theoretically applied to diagnose students' computing difficulties, differentiate educational approaches that emphasize either material or cognitive aspects of computational literacy, and underscore both the advantages and limitations of open projects like computational essays for student learning (Odden et al., 2019).

The use of diagnostic tests is also done to understand the beliefs and attitudes of engineering students from a Colombian-accredited university about physics and how physics should be taught in this context (Miranda & Rada, 2018). The results of this study highlight several students' answers and examine the impact of their beliefs on their attitudes so that physics courses should be oriented or prepared in such a way that students can solidify their beliefs and attitudes. Then, research developed a two-tier test to diagnose unipolar reasoning in electricity in 100 students aged 20-25 years who are registered as third-year students at the University in the Baccalaureate Program in Science Education (Métioui & Trudel, 2020). Finally, diagnostic tests can also be used for placement tests, such as the study of students taking physics mechanics classes at Stanford University . The study results indicated that most students adhered to the initial testing process, and diagnostic test scores effectively predicted performance in both introductory physics courses for physics majors and courses aimed at science and engineering majors. However, these scores were not good predictors for performance in the other two courses, likely due to a misalignment between course objectives and the diagnostic measures. The researcher also discovered that many individual questions were highly effective in predicting performance across different courses, highlighting specific topics emphasized in teaching. Of the 38 questions, 27 were predictors of performance in at least one course, while only 8 questions were neither predictive of course performance nor indicative.

This research reviews the study of trends in the use of diagnostic tests in science education in the last 5 years (2018-2023). The search is limited to the specified database as in the methodology section. In addition, the study includes a wide range of studies published only in English because of the language that the author can understand. Therefore, future research can be expanded by investigating studies from languages other than English. Previously, another research also reviewed 58 four-level diagnostic test articles in physics education from 2010 to 2012 in English and Turkish (Çelikkanlı & Kızılcık, 2022).

In terms of subject matter, it was found that the most dominant diagnostic tests were studied in physics learning. This shows the author's limitations in reviewing all subjects equally. However, on the other hand, it provides a more comprehensive picture for diagnostic test research in physics learning and can be used further by researchers in the future. In multitier diagnostic tests, the number of questions must be increased and made more efficient. Developing efficient tests across various physics topics will be beneficial in diagnosing misunderstandings. Additionally, strategies that focus on developing a high level of self-confidence in students should be considered in future research. The future research should also explore how diagnostic tests can be used to promote equitable learning opportunities and enhance the overall quality of education in diverse subjects.

Conclusion

This study has examined the use of diagnostic tests in science learning. The types of diagnostic test instruments used are two-tier diagnostic test, essay, three-tier diagnostic test, multiple choice, four-tier diagnostic test, self-diagnosis, blending diagram, cognitively diagnostic assessment, diagnostic adapted from TIMSS, six-tier and five-tier diagnostic tests. Diagnostic tests play a big role in identifying student misconceptions, identifying students at risk of low achievement, and testing for reduced levels of misconceptions using narrative feedback, e-learning modules, and realistic videos. In addition, his role includes identifying conceptual understanding, identifying differences in understanding between genders, identifying various knowledge, practices, and beliefs of students in the three aspects of computational literacy (material, cognitive, and social aspects), and diagnosing HOTS. Other aspects that can be measured from diagnostic tests are understanding students' beliefs and attitudes, knowing the development of students' disciplinary literacy, identifying creative thinking skills, testing the effects of collaborative mind mapping, diagnosing difficulties in students' reasoning, diagnosing unipolar reasoning in electricity and for physics placement, predictors. student success in learning. Diagnostic tests support sustainable development in education by promoting fair learning opportunities, improving teaching practices, and enhancing science education quality, which helps build more resilient and effective educational systems.

References

Adimayuda, R., Aminudin, A. H., Kaniawati, I., Suhendi, E., & Samsudin, A. (2020). A Multitier Open-Ended Momentum And Impulse (MOMI) Instrument: Developing And Assessing Quality Of Conception Of 11 th Grade Sundanese Students With Rasch Analysis. *International Journal of Scientific & Technology Research*, *9*, 2. www.ijstr.org.

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.

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.

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.

Batlolona, J. R., Diantoro, M., Wartono, & Latifah, E. (2019). Creative thinking skills students in physics on solid material elasticity. *Journal of Turkish Science Education*, *16*(1), 48–61. https://doi.org/10.12973/tused.10265a.

Birenbaum, M. (1997). Assessment Preferences and Their Relationship to Learning Strategies and Orientations. *Higher Education*, *33*, 71–84.

Borrego, M., Foster, M. J., & Froyd, J. E. (2014). Systematic literature reviews in engineering education and other developing interdisciplinary fields. *Journal of Engineering Education*, *103*(1), 45–76. https://doi.org/10.1002/jee.20038.

Bramwell-Lalor, S. (2019). Assessment for Learning on Sustainable Development. In *Encyclopedia of Sustainability in Higher Education* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-319-63951-2_1-1.

Brendefur, J. L., Johnson, E. S., Thiede, K. W., Strother, S., & Severson, H. H. (2018). Developing a Multi-Dimensional Early Elementary Mathematics Screener and Diagnostic Tool: The Primary Mathematics Assessment. *Early Childhood Education Journal*, *46*(2), 153–157. https://doi.org/10.1007/s10643-017-0854-x.

Burkholder, E., Wang, K., & Wieman, C. (2021). Validated diagnostic test for introductory physics course placement. *Physical Review Physics Education Research*, *17*(1). https://doi.org/10.1103/PhysRevPhysEducRes.17.010127.

Canlas, I. P. (2021). Using visual representations in identifying students' preconceptions in friction. *Research in Science and Technological Education*, *39*(2), 156–184. https://doi.org/10.1080/02635143.2019.1660630

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, *19*(4), 1291–1311. https://doi.org/10.36681/tused.2022.175.

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

Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01522.

Day, I. N. Z., van Blankenstein, F. M., Westenberg, M., & Admiraal, W. (2018). A review of the characteristics of intermediate assessment and their relationship with student grades. *Assessment and Evaluation in Higher Education*, *43*(6), 908–929. https://doi.org/10.1080/02602938.2017.1417974.

Djudin, T. (2021). Promoting Students' Conceptual Change by Integrating The 3-2-1 Reading Technique with Refutation Text in The Physics Learning of Buoyancy. *Journal of Turkish Science Education*, *18*(2), 290–303. https://doi.org/10.36681/tused.2021.66.

Filderman, M. J., Toste, J. R., Didion, L. A., Peng, P., & Clemens, N. H. (2018). Data-Based Decision Making in Reading Interventions: A Synthesis and Meta-Analysis of the Effects for Struggling Readers. In *Journal of Special Education* (Vol. 52, Issue 3, pp. 174–187). SAGE Publications Inc. https://doi.org/10.1177/0022466918790001.

Fung, D., & Liang, T. (2023). The Effectiveness of Collaborative Mind Mapping in Hong Kong Primary Science Classrooms. *International Journal of Science and Mathematics Education*, *21*(3), 899–922. https://doi.org/10.1007/s10763-022-10279-1.

Gierl, M. J. (2007). Making Diagnostic Inferences About Cognitive Attributes Using the Rule-Space Model and Attribute Hierarchy Method. In *Journal of Educational Measurement Winter* (Vol. 44, Issue 4).

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.

Imaduddin, M., Shofa, M. M., Riza, M. F., & Fikri, A. A. (2023). Exploring the pre-service basic science teachers' misconceptions using the six-tier diagnostic test. *International Journal of Evaluation and Research in Education*, *12*(3), 1626–1636. https://doi.org/10.11591/ijere.v12i3.24603.

Irmak, M., Inaltun, H., Ercan-Dursun, J., Yaniş-Kelleci, H., & Yürük, N. (2023). Development and Application of a Three-Tier Diagnostic Test to Assess Pre-service Science Teachers' Understanding on Work-Power and Energy Concepts. *International Journal of Science and Mathematics Education*, *21*(1), 159–185. https://doi.org/10.1007/s10763-021-10242-6.

Istiyani, R., Muchyidin, A., & Rahardjo, D. H. (2018). Analysis of Student Misconception On Geometry Concepts Using Three-Tier Diagnostic Test. *Cakrawala Pendidikan*, *37*(2). https://doi.org/http://dx.doi.org/10.21831/cp.v37i2.14493.

Istiyono, E., Sunu Brams Dwandaru, 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.

Kahraman, S. (2019). Evaluating University Students' Understanding of Atmospheric Environmental Issues Using a Three-Tier Diagnostic Test. *International Electronic Journal of Environmental Education*, *9*(1), 1–17.

Kang, N. H., & Yoon, H. (2022). Identifying Students at Risk in Learning Science: Diagnostic Assessment in Physical Science for Primary and Lower Secondary School Grades. *Asia-Pacific Science Education*, *8*(2), 331–360. https://doi.org/10.1163/23641177-bja10049.

Kaniawati, I., Fratiwi, N. J., Danawan, A., Suyana, I., Samsudin, A., & Suhendi, E. (2019). Analyzing students' misconceptions about Newton's Laws through Four-Tier Newtonian Test (FTNT). *Journal of Turkish Science Education*, *16*(1), 110–122. https://doi.org/10.12973/tused.10269a.

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.

Koretz, D. (2017). *Moving Beyond the Failure of Test-Based Accountability*.

Laeli, C. M. H., Gunarhadi, & Muzzazinah. (2023). The 3 Tiers Multiple-Choice Diagnostic Test for Primary Students' Science Misconception. *Pegem Egitim ve Ogretim Dergisi*, *13*(2), 103–111. https://doi.org/10.47750/pegegog.13.02.13.

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.

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.

Maulana, S., Rusilowati, A., Nugroho, S. E., & Susilaningsih, E. (2023). *Diagnostic Test for Analyze Students' Conception in Science Education: A Review Article*. https://proceeding.unnes.ac.id/index.php/iset698.

Mbwile, B., Ntivuguruzwa, C., & Mashood, K. K. (2023). Exploring the Understanding of Concept Inventories for Classroom Assessment by Physics Tutors and Pre-Service Teachers in Tanzania. *African Journal of Research in Mathematics, Science and Technology Education*, *27*(1), 36–46. https://doi.org/10.1080/18117295.2023.2183607.

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.

Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and Assessment in Teaching* (10th ed.). Pearson.

Miranda, J. C., & Rada, T. (2018). An examination of the beliefs about physics and learning physics among engineering students. *International Journal of Engineering Education*, *34*(4), 1427–1435.

Mubarokah, F. D., Mulyani, S., & Indriyanti, N. Y. (n.d.). *Identifying Students' Misconceptions of Acid-Base Concepts Using a Three-Tier Diagnostic Test: A Case of Indonesia and Thailand*. https://doi.org/10.12973/tused.10256a.

Nahadi, N. (2022). Development and Application of A Two-Tier Acid-Base Misconception Diagnostic Test Based On Pictorial To Identifying Student Misconceptions In Chemistry. In *Journal of Engineering Science and Technology Special Issue on ISCoE2022*.

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.

Odden, T. O. B., Lockwood, E., & Caballero, M. D. (2019). Physics computational literacy: An exploratory case study using computational essays. *Physical Review Physics Education Research*, *15*(2). https://doi.org/10.1103/PhysRevPhysEducRes.15.020152.

Osborne, J., & Dillon, J. (2014). *Science Education in Europe: Critical Reflections*. https://www.researchgate.net/publication/252404504.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. In *The BMJ* (Vol. 372). BMJ Publishing Group. https://doi.org/10.1136/bmj.n71.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.

Phanphech, P., Tanitteerapan, T., Mungkung, N., Arunrungrusmi, S., Chunkul, C., Songruk, A., Yuji, T., & Kinoshita, H. (2022). An Analysis of Student Anxiety Affecting on Online Learning on Conceptual Applications in Physics: Synchronous vs. Asynchronous Learning. *Education Sciences*, *12*(4). https://doi.org/10.3390/educsci12040278.

Ping, I. L. L., Halim, L., & Osman, K. (2020). Rural students' conceptual understanding of diffusion and osmosis. *Asia Pacific Journal of Educators and Education*, *35*(1), 93–109. https://doi.org/10.21315/apjee2020.35.1.6.

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.

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The development of an instrument to measure the higher order thinking skill in physics. *European Journal of Educational Research*, *8*(3), 743–751. https://doi.org/10.12973/eu-jer.8.3.743.

Safadi, R., & Saadi, S. (2021). Learning from Self-Diagnosis Activities when Contrasting Students' Own Solutions with Worked Examples: the Case of 10th Graders Studying Geometric Optics. *Research in Science Education*, *51*(2), 523–546. https://doi.org/10.1007/s11165-018-9806-8.

Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649.

Soeharto, S., & Csapó, B. (2021). Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. *Heliyon*, *7*(11). https://doi.org/10.1016/j.heliyon.2021.e08352.

Sopandi, W., & Sukardi, R. R. (2020). Using four tier diagnostic tests to understand the conceptions held by pre service primary school teachers about sea pollutant migration. *Review of International Geographical Education Online*, *10*(2), 13–29. https://doi.org/10.33403/rigeo.629388.

Sun, L., Kangas, M., Ruokamo, H., & Siklander, S. (2023). A systematic literature review of teacher scaffolding in game-based learning in primary education. In *Educational Research Review* (Vol. 40). Elsevier Ltd. https://doi.org/10.1016/j.edurev.2023.100546.

Suprapto, N., Abidah, A., Dwiningsih, K., Jauhariyah, M. N. R., & Saputra, A. (2018). Minimizing misconception of ionization energy through three-tier diagnostic test. *Periodico Tche Quimica*, *15*(30), 387–396. https://doi.org/10.52571/ptq.v15.n30.2018.390_periodico30_pgs_387_396.pdf.

Taban, T., & Kiray, S. A. (2022). Determination of Science Teacher Candidates' Misconceptions on Liquid Pressure with Four-Tier Diagnostic Test. *International Journal of Science and Mathematics Education*, *20*(8), 1791–1811. https://doi.org/10.1007/s10763-021-10224-8.

Van Den Eynde, S., Schermerhorn, B. P., Deprez, J., Goedhart, M., Thompson, J. R., & De Cock, M. (2020). Dynamic conceptual blending analysis to model student reasoning processes while integrating mathematics and physics: A case study in the context of the heat equation. *Physical Review Physics Education Research*, *16*(1). https://doi.org/10.1103/PhysRevPhysEducRes.16.010114

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3.

von der Embse, N. P., Pendergast, L. L., Segool, N., Saeki, E., & Ryan, S. (2016). The influence of test-based accountability policies on school climate and teacher stress across four states. *Teaching and Teacher Education*, *59*, 492–502. https://doi.org/10.1016/j.tate.2016.07.013.

Wancham, K., Tangdhanakanond, K., & Kanjanawasee, S. (2023a). Development of the automatic item generation system for the diagnosis of misconceptions about force and laws of motion. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(6). https://doi.org/10.29333/ejmste/13251.

Wancham, K., Tangdhanakanond, K., & Kanjanawasee, S. (2023b). Sex and Grade Issues in Influencing Misconceptions about Force and Laws of Motion: An Application of Cognitively Diagnostic Assessment. *International Journal of Instruction*, *16*(2), 437–456. https://doi.org/10.29333/iji.2023.16224a.

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., Istiyono, E., Hidayati, K., Retnawati, H., Hasan, A., & Karimah, U. (2024). Teacher's Perspective on Diagnostic Assessment Process During the Pandemic Covid-19. *Pegem Journal of Education and Instruction*, *14*(4), 400–409. https://doi.org/10.47750/pegegog.14.04.38.