Journal of Engineering Science and Technology
Special Issue on ISCoE2024 Vol. 20 No. 4 (2025) 39 - 46

A WEB-BASED FOUR-TIER DIAGNOSTIC TOOL FOR ASSESSING MATHEMATICS LITERACY

ANGGIT PRABOWO[1,2], TATANG HERMAN[1,*], SITI FATIMAH[1], AFIT ISTIANDARU[2], NUR ROBIAH NOFIKUSUMAWATI PENI[2]

[1]Universitas Pendidikan Indonesia, Bandung, Indonesia
[2]Universitas Ahmad Dahlan, Yogyakarta, Indonesia
*Corresponding Author: tatangherman@upi.edu

Abstract

This study developed a web-based diagnostic test application to assess mathematical literacy in geometry (lines and angles) using four-tier items. The purpose was to enhance diagnostic accuracy and efficiency in identifying student misconceptions. The ADDIE model guided the development process through analysis, design, development, implementation, and evaluation. The application was validated by experts in mathematics literacy and information technology and tested on 139 junior high school students in Yogyakarta. The results confirmed good item quality, including difficulty index, discriminant index, and reliability. The improvement occurred because the four-tier structure provided deeper insights into students' reasoning and confidence levels. The application also received positive user feedback for its clarity, usability, and comprehensive reporting. This study contributes an effective web-based diagnostic tool that supports mathematical literacy assessment in an accessible and scalable format.

Keywords: Assessment, Diagnosis, Four-tier, Mathematical literacy, Web-based.

1. Introduction

Web-based platforms have significantly transformed educational assessment, offering flexibility, accessibility, and efficiency [1, 2]. Many reports regarding web-based platforms have been well-documented [3-6]. One such advancement is diagnostic assessment, which identifies student misconceptions to inform targeted instruction [7, 8]. Various diagnostic instruments (such as interviews, open-ended questions, and multiple-tier tests) offer different strengths and limitations [9].

Table 1 explains previous diagnostic test studies, showing a predominance of three-tier paper-based formats, often focused on physics content [10-12]. Four-tier tests, incorporating confidence levels and reasoning, remain limited, particularly in mathematics literacy [13, 14]. This study developed a web-based diagnostic test with four-tier items to assess mathematical literacy in geometry (lines and angles), using the ADDIE model. The novelties included: (i) transitioning from paper-based to web-based diagnostics, (ii) applying the four-tier format in mathematics, and (iii) enhancing diagnostic accuracy through online platforms.

Table 1. Previous studies on diagnostic test.

No.	Title	Ref.
1	Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction.	[10]
2	A three-tier diagnostic test to assess pre- service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain.	[11]
3	Exploring pre-service elementary science teachers' conceptual understanding of particulate nature of matter through three-tier diagnostic test.	[15]
4	Assessing students' conceptual understanding using an online three-tier diagnostic test.	[16]
5	Analysing geometry misconception of prospective teachers using three-tier diagnostic test.	[17]
6	A review of studies about four-tier diagnostic tests in physics education.	[18]
7	Four-tier diagnostic test to identify misconceptions in geometrical optics.	[12]

2. Literature Review

Figure 1 shows the entity relationship diagram of a computer-based test (CBT) system, integrating administrators, teachers, and students within a web-based assessment platform [19, 20]. The CBT model enables efficient test management, immediate feedback, and supports eco-friendly digital assessment [21, 22].

Mathematical literacy, as defined by Stacey and Turner, involves the ability to formulate, use, and interpret mathematics in real-world contexts, enhancing reasoning and problem-solving [23, 24]. The PISA framework emphasizes literacy across quantitative, spatial, and data-based content.

The four-tier diagnostic test expands on multiple-choice formats by adding reasoning layers and confidence levels, improving diagnostic precision [13, 25].

Though widely used in physics [14, 26], its application in mathematics literacy remains limited, signaling a need for further development [27].

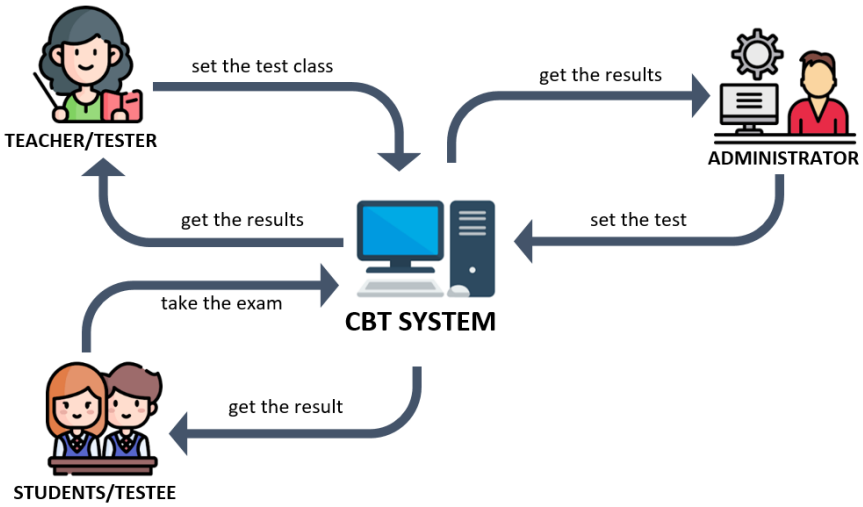

Fig. 1. Diagram for the CBT software.

3. Method

This study followed the ADDIE model, comprising analysis, design, development, implementation, and evaluation stages. Detailed information regarding this method is explained elsewhere [28]. The analysis phase identified the needs for diagnosing mathematical literacy in geometry (lines and angles). The design phase structured the web-based diagnostic test using four-tier items. During development, the application was built and refined through alpha testing with ICT experts. The implementation phase involved 139 junior high school students in Yogyakarta, Indonesia. Finally, the evaluation phase assessed the application's quality, including usability, item validity, difficulty index, discriminant index, and reliability, alongside student responses to ensure its effectiveness in diagnosing mathematical literacy.

4. Results and Discussion

Figure 2 shows the development flow of the web-based diagnostic test application, beginning from needs analysis, goal setting, website design, item blueprint, to assessment strategy formulation [19]. Figure 3 shows the website interface. We showed the homepage, account registration, and the test interface with timers and question indicators, corresponding to Figs. 3(a), (b), and (c), respectively, ensuring an organized and user-friendly experience [21].

The alpha testing phase involved ICT experts refining technical functionalities. We analyzed statistics to get a better understanding of the results. Detailed information on how to analyze using statistical analysis is reported elsewhere [29-31]. Expert validation confirmed the test items' quality, with an average difficulty index of 0.55 (good), a discriminant index of 0.43 (good), and a reliability coefficient of 0.76 (reliable). These results occurred because the four-tier item

structure enhanced the diagnostic accuracy, offering clear reports and accessible feedback [32-39].

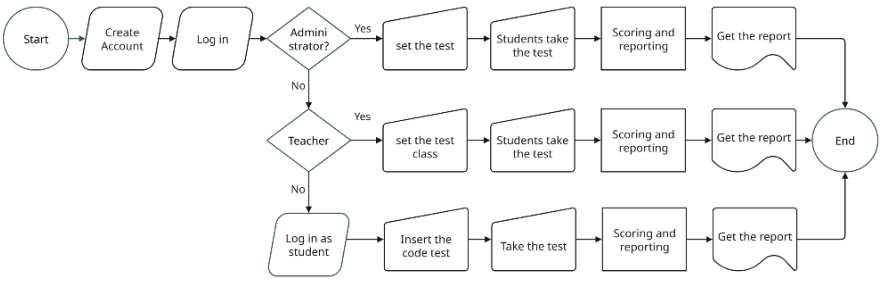

Fig. 2. Web-based diagnostic test application flowchart.

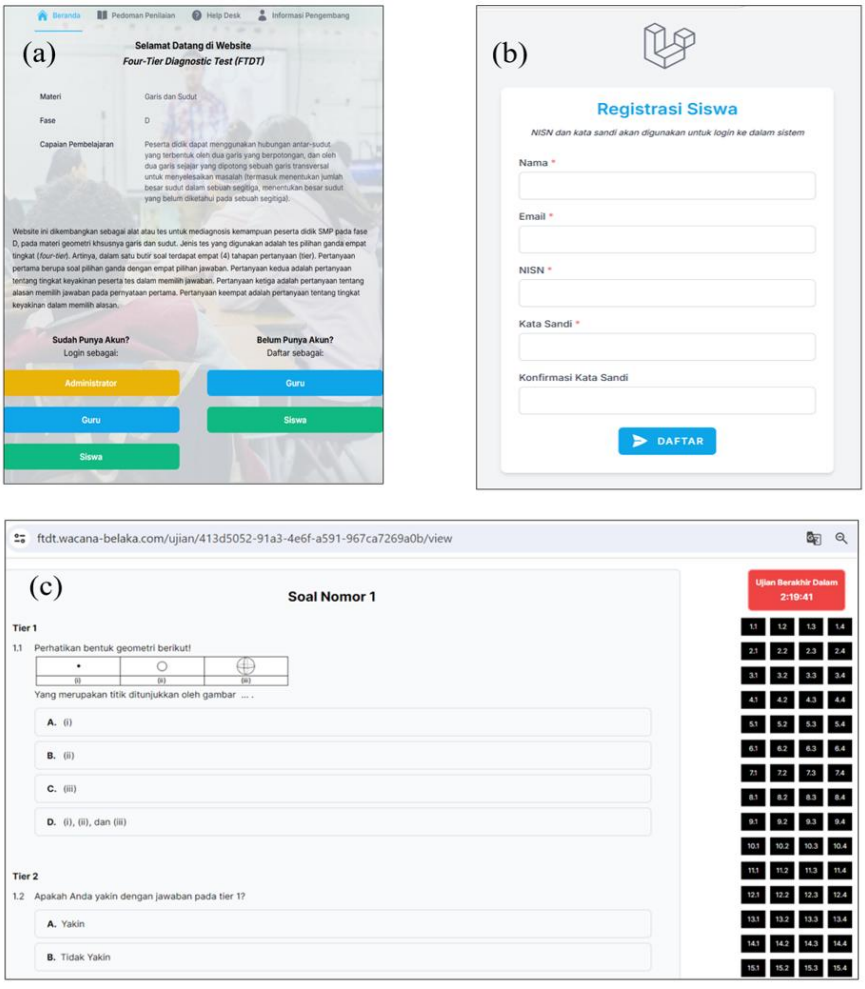

Fig. 3. Website appearance: (a) Home page; (b) Account registration; (c) Test appearance.

Consistent with prior research, these findings reinforce the effectiveness of four-tier diagnostic tests in pinpointing student misconceptions and ensuring reliable measurement in mathematics literacy assessments [13, 14, 26, 35]. This new model test has been shown to accurately identify specific misunderstandings and provide consistent results, which is crucial for effective evaluation.

5. Conclusion

This study developed a web-based diagnostic test application using four-tier items to assess mathematical literacy in geometry (lines and angles). The application was validated by experts and tested on 139 students, confirming good item characteristics (difficulty, discriminant, and reliability). These results occurred because the four-tier structure improved the diagnostic depth by assessing both reasoning and confidence levels. The application also provided clear reports, was user-friendly, and easily accessible. This research offers a scalable tool for effectively diagnosing mathematical literacy, supporting targeted interventions in educational assessments.

Acknowledgement

This study was supported by the Directorate of Research, Technology, and Community Service Ministry of Education, Culture, Research and Technology, Indonesia under the Grant No.107/E5/PG.02.00.PL/2024, 0609.12/LL5-INT/AL.04/2024, 014/PFR/LPPM-UAD/VI/2024.

References

1. Doğan, S.; Betin-Can, A.; and Garousi, V. (2014). Web application testing: A systematic literature review. *Journal of Systems and Software*, 91(1), 174-201.
2. Dahalan, H.M.; and Hussain, R.M.R. (2010). Development of web-based assessment in teaching and learning management system (e-ATLMS). *Procedia - Social and Behavioral Sciences*, 9, 244-248.
3. Hartati, A.D.; Maryanti, R.; Azizah, N.N.; Al Husaeni, D.F.; Wulandary, V.; and Irawan, A.R. (2023). Webtoon comic media to improve reading comprehensions for students with hearing impairment in special primary schools. *ASEAN Journal of Community and Special Needs Education*, 2(1), 9-16.
4. Nuhu, K.M.; and Onojah, A.O. (2022). Effect of webinar multimedia platform on students' academic performance in selected educational technology concepts in University of Ilorin. *Indonesian Journal of Multidiciplinary Research*, 2(1), 9-20.
5. Nugraha, M.G. (2023). Development of web-based radioactivity teaching materials oriented-on character education for highschool students. *ASEAN Journal of Science and Engineering Education*, 3(1), 95-102.
6. Sumitra, I.D.; Wirawan, E.; Putra, M.; Kusumaningrat, I.; and Maulana, H. (2021). Design of webinar information system for people with hearing impairments. *Indonesian Journal of Community and Special Needs Education*, 1(2), 77-86.
7. Lin, Y.C.; Yang, D.C.; and Li, M.N. (2016). Diagnosing students' misconceptions in number sense via a web-based two-tier test. *Eurasia*

Journal of Mathematics, Science and Technology Education, 12(1), 41-55.

8. Gurel, D.K.; Eryilmaz, A.; and 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.
9. Ojose, B. (2015). Students' misconceptions in mathematics: Analysis of remedies and what research says. *Ohio Journal of School Mathematics*, 72, 30-34.
10. Odom, A.L.; and Barrow, L.H. (1995). Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *Journal of Research in Science Teaching*, 32(1), 45-61.
11. Arslan, H.O.; Cigdemoglu, C.; and Moseley, C. (2012). A three-tier diagnostic test to assess pre- service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, 34(11), 1667-1686.
12. Fariyani, Q.; Rusilowati, A.; and Sugianto, S. (2017). Four-tier diagnostic test to identify misconceptions in geometrical optics. *Unnes Science Education Journal*, 6(3), 1724-1729.
13. Caleon, I.S.; and Subramaniam, R. (2010). Do students know what they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, 40(3), 313-337.
14. Kaltakci-Gurel, D.; Eryilmaz, A.; and McDermott, L.C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science and Technological Education*, 35(2), 238-260.
15. Aydeniz, M.; Bilican, K.; and Kirbulut, Z.D. (2017). Exploring pre-service elementary science teachers' conceptual understanding of particulate nature of matter through three-tier diagnostic test. *International Journal of Education in Mathematics, Science and Technology*, 5(3), 221-221.
16. Yang, D.C.; and 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.
17. Prayitno, S.; Arjudin, A.; and Hapipi, H. (2020). Analyzing geometry misconception of prospective teachers using three-tier diagnostic test. *Advances in Social Science*, *Education and Humanities Research*, 465, 119-122.
18. Çelikkanlı, N.O.; and Kızılcık, H.S. (2022). A review of studies about four-tier diagnostic tests in physics education. *Journal of Turkish Science Education*, 19(4), 1291-1311.
19. Ajinaja, M. (2017). The design and implementation of a computer based testing system using component-based software engineering. *International Journal of Computer Science and Technology*, 8(1), 58-65.
20. Shaturaev, J.; and Khamitovna, K.K. (2023). A computer-based approach to teaching foreign languages. *ASEAN Journal of Educational Research and Technology*, 2(2), 89-98.

21. Rajiha, N. and Rini, S. (2023). Computer-based standardized testing (CBT) summative assessment: One institution experience. *JISAE*: *Journal of Indonesian Student Assessment and Evaluation*, 9(1), 1-7.
22. Cagiltay, N.; and Ozalp-Yaman, S. (2010). How can we get benefits of computer-based testing in engineering education?. *Benefit of Computer-Based Testing*, 21(2), 1-7.
23. Suciati, S.; Munadi, S.; Sugiman, S.; and Febriyanti, R.W.D. (2020). Design and validation of mathematical literacy instruments for assessment for learning in Indonesia. *European Journal of Educational Research*, 9(2), 865-875.
24. Husna, N.; Zulfahita, Mulyani, S.; and Rahman, M.S. (2020). Reading literacy, mathematical literacy, and self-confidence of junior high school students in Singkawang. *Journal of Education*, *Teaching and Learning*, 5(2), 253-262.
25. Sopandi, W.; and 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.
26. Kiray, S.A.; and 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.
27. Soeharto; Csapó, B.; Sarimanah, E.; Dewi, F.I.; and Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, 8(2), 247-266.
28. Susilawati, A.; Al-Obaidi, A.S.M.; Abduh, A.; Irwansyah, F.S.; and Nandiyanto, A.B.D. (2025). How to do research methodology: From literature review, bibliometric, step-by-step research stages, to practical examples in science and engineering education. *Indonesian Journal of Science and Technology*, 10(1), 1-40.
29. Afifah, S.; Mudzakir, A.; and Nandiyanto, A.B.D. (2022). How to calculate paired sample t-test using SPSS software: From step-by-step processing for users to the practical examples in the analysis of the effect of application anti-fire bamboo teaching materials on student learning outcomes. *Indonesian Journal of Teaching in Science*, 2(1), 81-92.
30. Fiandini, M.; Nandiyanto, A.B.D.; Al Husaeni, D.F.; Al Husaeni, D.N.; and Mushiban, M. (2024). How to calculate statistics for significant difference test using SPSS: Understanding students comprehension on the concept of steam engines as power plant. *Indonesian Journal of Science and Technology*, 9(1), 45-108.
31. Rahayu, N.I.; Muktiarni, M.; and Hidayat, Y. (2024). An application of statistical testing: A guide to basic parametric statistics in educational research using SPSS. *ASEAN Journal of Science and Engineering*, 4(3), 569-582.
32. Musa, A.; Shaheen, S.; and Elmardi, A. (2018). Item difficulty and item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. *Khartoum Medical Journal*, 11(2), 1477-1468.
33. Gugiu, C.; and Gugiu, M. (2018). Determining the minimum reliability

standard based on a decision criterion. *Journal of Experimental Education*, 86(3), 458-472.

34. Shete, A.; Kausar, A.; Lakhkar, K.; and Khan, S. (2015). Item analysis: An evaluation of multiple choice questions in physiology examination. *Journal of Contemporary Medical Education*, 3(3), 106-109.
35. Tang, W.; Cui, Y.; and Babenko, O. (2014). Internal consistency : Do we really know what it is and how to assess it ? *Journal of Psychology and Behavioral Science*, 2(2), 205-220.
36. Abdelraheem, I.; Ahmed, M.; and Moalwi, A.A. (2017). Correlation between difficulty and discrimination indices of MCQs type a in formative exam in anatomy. *IOSR Journal of Research and Method in Education* (*IOSR-JRME*), 7(5), 28-43.
37. Dhakne-Palwe, S.; Gujarathi, A.; and Almale, B. (2015). Item analysis of mcqs and correlation between difficulty index, discrimination index and distractor efficiency in a formative examination in community medicine. *Journal of Research in Medical Education and Ethics*, 5(3), 254-259.
38. Hingorjo, M.R.; and Jaleel, F. (2012). Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *Journal of the Pakistan Medical Association*, 62(2), 142-147.
39. Sim, S.M.; and Rasiah, R.I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals of the Academy of Medicine Singapore*, 35(2), 67-71.