International Journal on Emerging Mathematics Education (IJEME)
Vol. 7, No. 1, March 2023, pp. 12-19
P-ISSN: 2549-4996, E-ISSN: 2548-5806, DOI: http://dx.doi.org/10.12928/ijeme.v7i1.22677

The Characteristics of Four-Tier Diagnostic Test: Classical Test Theory Perspective

Anggit Prabowo[1], Tatang Herman[2], Siti Fatimah[2]
[1]Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Universitas Pendidikan Indonesia, Bandung, Indonesia
e-mail: anggit.prabowo@pmat.uad.ac.id

Abstract

This study is part of developing a four-tier online diagnostic test for geometry in junior high schools in Indonesia. After wrote the items and validated by experts, items were tested on 249 students in Yogyakarta, one of the Special Regions in Indonesia. This study aims to identify the characteristics of a four-tier online diagnostic test, from a classical test theory (CTT) perspective. This test contains 20 items of four-tier multiple choice with four options. The students' responses were analyzed using Item and Test Analysis (ITEMAN) package. Item difficulty index, discriminant index, and reliability coefficient were three attributes analyzed. The average of item difficulty index is 0.384 (moderate category). The average item discriminant index is 0.513 (good category). The estimation of the reliability coefficient using the Cronbach-Alpha formula is 0.706 (good category).

Keywords: diagnostic test, four-tier, CTT

How to Cite: Prabowo, A., Herman, T., & Fatimah, S. (2023). The Characteristics of Four-Tier Diagnostic Test: Classical Test Theory Perspective. *International Journal on Emerging Mathematics Education, 7*(1), 12-19. http://dx.doi.org/10.12928/ijeme.v7i1.22677

INTRODUCTION

Misconceptions arise when there is a gap between the concepts understood by the participants and the actual concepts. Misconceptions are misunderstandings and misinterpretations based on the wrong meaning (Ojose, 2015). Misconceptions that are not handled will have a bad impact because it makes students have an understanding that is far from the correct concept. This is because they build their own concept (Aydin, Keles, & Hasiloglu, 2012). Therefore, the students' misconceptions need to be identified by a teacher urgently (Kula & Güzel, 2014).

Students have the opportunity to experience misconceptions in various subjects, including mathematics. Various studies have focused on analyzing students' misconceptions about learning mathematics (Gooding & Metz, 2011; Roselizawati, Sarwadi, & Shahril, 2014; Mohyuddin & Khalil, 2016; Aliustaoğlu, Tuna, & Biber, 2018) Students who have misconceptions in mathematics materials will have cognitive development obstructed (Kusmaryono, Basir, & Saputro, 2020). This is because the characteristics of mathematics learning materials are interrelated and continuous with other materials, so a misconception results in a continuous misconception. Studying any of the mathematical topics at an advanced level must be based on reasoning from basic knowledge or prior prerequisite knowledge. If someone experiences mathematical misconceptions in lower grade learning and is not immediately addressed, it will have an impact on mathematics learning in higher grades (Flevares & Schiff, 2014). Thus, the identification and handling of misconceptions needs to be done as early as possible.

Received December 29, 2021; Revised May 17, 2022; Accepted March 30, 2023

Mathematics is a subject studied by Indonesian students at various levels of education. Mathematics has a major role in various fields of science (Shim, Shakawi, & Azizan, 2017), such as physics, mechanical engineering, and statistics (Muijs & Reynolds, 2005). One indicator of mastery of competence in the field of mathematics can be seen based on the results of the national examination. The last three years of the implementation of the national examination (2017 to 2019), showed that the average mastery of junior high school students in mathematics subjects tended to be below 50% (Prabowo, Anggoro, Rahmawati, & Rokhima, 2019).

The mathematics material tested at the national examination of junior high school in Indonesia consists of numbers, algebra, geometry and measurement, and statistics and probability (Prabowo, Anggoro, Rahmawati, & Rokhima, 2019). The percentage of correct answers for the four materials is presented in Table 1.

Table 1. Percentage of Students Who Answered Correctly on the Junior High School Mathematics National Examination

No	Material	2016	2017	2018	2019	Average
1	Number	52.74%	51.05%	44.99%	39.71%	49.59%
2	Algebra	52.97%	48.60%	41.88%	51.24%	47.82%
3	Geometry and measurement	47.19%	48.57%	41.40%	42.27%	45.72%
4	Statistics and Probability	46.73%	56.40%	45.71%	55.60%	49.61%

Based on Table 1, it can be seen that in the last four years, geometry and measurement materials have become the material with the smallest average percentage to be answered correctly by junior high school students participating in the national examination. The low mastery of the material is closely related to the occurrence of misconceptions about the material (Kusmaryono et al., 2020). Thus, it is possible that junior high school students in Indonesia tend to have high misconceptions about geometry and measurement materials. This is in line with studies (Retnawati, Arlinwibowo, & Sulistyaningsih, 2017) which show that students in Indonesia experience difficulties and misconceptions in understanding geometric concepts. Misconceptions in geometry and measurement materials also occur in students in various countries in the world such as Cyprus (Özerem, 2012), Pakistan (Mohyuddin & Khalil, 2016), and Brunei Darussalam (Roselizawati, Sarwadi, & Shahril, 2014).

Misconceptions identified early on will be easier to correct. One way to find out students' misconceptions is through a diagnostic test (Fariyani, Rusilowati, & Sugianto, 2017). Diagnostic tests were developed into several models. The model that is able to provide the most complete information in making a diagnosis is the four-tier model. In this model, an item is equipped with answer choices, reasons for choosing answers, and the level of confidence in each answer and reason(Caleon & Subramaniam, 2010). The addition of the level of confidence in each answer and reason can measure the difference in the level of student's knowledge so that it will help in detecting the level of student misconceptions. This test model was initially developed on physical materials, such as optics (Caleon & Subramaniam, 2010), fluida (Diani, Alfin, Anggraeni, Mustari, & Fujiani, 2019), energy (Anggrayni & Ermawati, 2019), and Newton Law (Sundaygara et al., 2021). The most informative diagnostic test in diagnosing students' strengths and weaknesses is the four-tier model that has not been developed in mathematics, especially geometry.

The four-tier diagnostic test for geometry was developed prior to this study. It has been implemented for 249 students in Indonesia. Each item contains stem, alternatives, degree of belief to choose the alternatives, reasons why they choose the alternative, and degree of belief they choose the reason. The quality of this test has been known until the characteristics of the items are identified. This study aims to identify the characteristics of the items consisting of item difficulty index, discriminant index, and reliability from the classical test theory (CTT) perspective. It has been widely used in the item analysis process (Ravand & Baghaei, 2020). The fundamental tenet of classical test theory is that the observed score is a random variable that includes the measurement error, a random latent variable (Traub, 1997).

RESEARCH METHOD

This research used a descriptive approach as a part of the research and development of a four-tier diagnostic test of geometry for students in grade VII. This test has been developed based on the procedures of test development. After creating the items based on the blueprint and judging by experts as valid tests, it was implemented on 249 students in Yogyakarta, Indonesia. The pilot study compiled responses to four-tier diagnostic tests from respondents. The response analyzed only the students' first answer, not including the degree of belief, reasons, and degree of belief in choosing the reason. The type of response contains options A, B, C, or D. It was analyzed using ITEMAN to identify the difficulty index and discriminant index and use SPSS to estimate the reliability coefficient. The formula used to estimate reliability was a Cronbach-Alpha.

RESULTS AND DISCUSSION

This research was one part of the research and development of a four-tier diagnostic test of geometry. This test focuses on geometry material for students of junior high school, grade VII. The developer chose geometry because it was one of the most difficult subjects for junior high school students in Indonesia (Prabowo, Anggoro, Astuti, & Fahmi, 2017; Prabowo, Anggoro, et al., 2019). All the processes to develop were planning, writing items, assembling the test, providing directions, preparing the scoring key and marking scheme, and reviewing the test. One of the procedures at the last step is to review the test based on empirical data. It means the quality of the test was reviewed from the data of responses based on the pilot study. Data collected from the pilot study were students' answers (A, B, C, D), reasons, and degree of belief in choosing the answer and reason (sure, not sure).

Items Difficulty index

Items difficulty index means the proportion of respondents who correctly answered an item. It usually was represented by p_i. Table 2 presents the items' difficulty index from the pilot study of 249 students.

Table 2. Items Difficulty Index

No	Difficulty Index	Category
1	0.386	Moderate
2	0.478	Moderate
3	0.394	Moderate
4	0.450	Moderate
5	0.494	Moderate
6	0.598	Moderate
7	0.345	Moderate
8	0.313	Moderate
9	0.398	Moderate
10	0.418	Moderate
11	0.486	Moderate
12	0.474	Moderate
13	0.606	Moderate
14	0.329	Moderate
15	0.301	Moderate
16	0.321	Moderate
17	0.141	Too difficult
18	0.257	Too difficult
19	0.177	Too difficult
20	0.321	Moderate

The average of the items difficulty index is 0.384 (moderate). It is ideal because the expected item difficulty index is in interval 0.3 to 0.7 (Musa, Shaheen, & Elmardi, 2018).

Of 20 items, 17 are moderate and 3 are too difficult. Item 17 is the most difficult. Only 35 students answer it correctly. This item needs an analysis thinking to compile the information provided to answer the question. Here is the stem of item 17. Given the sketch of the house below.

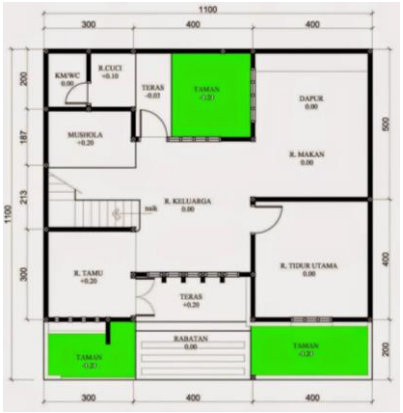

Figure 1. Sketch of House

Taken from https://www.99.co/blog/indonesia/gambar-denah-rumah-sederhana/

The developer want to tile all of the floor. The owner want that tiles in main sleeping room (ruang tidur utama) and guess room (ruang tamu) are same. The question is how m^2 of tiles needed to tile the floor. It is not easy for students because to solve the problem students need to identify the area of both room. To identify it, they need to analyze the sketch. After identify the dimension, they must measure the booth area.

They also need to convert of unit from centimeter (cm) to meter (m). In other hand, the stem is not clear. By asking how m^2, it is actually has same meaning with asking the area of both room. So, by revise the stem it is expected to make the item more clearly for students.

Discriminant index

Discriminant index (DI) means the ability of an item to distinguish test takers based on their level of ability. Table 3 presents the items discriminant index from the pilot study of 249 students. Items are good when they have a discriminant index of more than 0.2 (Fernandes, 1984).

Table 3 Items Discriminant Index

No	Discriminant index	Category
1	0.531	good
2	0.545	good
3	0.524	good
4	0.658	good
5	0.517	good
6	0.612	good
7	0.290	good
8	0.424	good
9	0.539	good
10	0.608	good
11	0.717	good
12	0.654	good
13	0.581	good
14	0.347	good
15	0.528	good
16	0.355	good
17	0.528	good
18	0.244	good
19	0.639	good
20	0.488	good

The average of the items discriminant index is 0.513 (good). It is ideal because the expected item discriminant index is more than 0.25 (Dhakne-Palwe, Gujarathi, & Almale, 2015). Even, 0.2 is enough (Fernandes, 1984). All items in this test are in good category in distinguishing students' ability.

Reliability

Reliability means consistency in test results. The test is reliable if the observed score has a high correlation with the actual score. There are various methods to estimate reliability. The first method is external consistency, and the second is internal consistency. The estimation of coefficient reliability of the test was estimated using internal consistency, Cronbach Alpha. Analysis using the SPSS package showed the coefficient reliability of the test is 0.706. It has already enough to state that the test is reliable because the minimum coefficient of reliability is 0.70 (Reynolds, Livingston, & Willson, 2010). It means this test will provide consistent results.

CONCLUSION

This study has identified the characteristics of the four-tier online diagnostic test, from a classical test theory perspective. This test contains 20 items of four-tier multiple choice with four options. The average item difficulty index is 0.384 (moderate category) The average item discriminant index is 0.513 (good category). The estimation of the reliability coefficient using the Cronbach-Alpha formula is 0.706 (good category). From a CTT perspective, this test is appropriate to be used to diagnose students' understanding in geometry and measurement.

DECLARATION

Author Contribution

All authors contribute in the research process, such as collecting the data, analyzing the data, and writing the manuscript. All authors approved the final manuscript.

Funding Statement

Not stated.

Conflict of Interest

Both authors declare that they have no competing interests.

Ethics Declaration

We as authors acknowledge that this work has been written based on ethical research that conforms with the regulations of our institutions and that we have obtained permission from the relevant institutes when collecting data. We support the International Journal on Emerging Mathematics Education (IJEME) in maintaining high standards of personal conduct, practicing honesty in all our professional practices and endeavors.

REFERENCES

Aliustaoğlu, F., Tuna, A., & Biber, A. Ç. (2018). Misconceptions of sixth grade secondary school students on fractions. *International Electronic Journal of Elementary Education*, *10*(5), 591–599. https://doi.org/10.26822/iejee.2018541308

Anggrayni, S., & Ermawati, F. U. (2019). The validity of Four-Tier's misconception diagnostic test for Work and Energy concepts. *Journal of Physics: Conference Series*, *1171*(1). https://doi.org/10.1088/1742-6596/1171/1/012037

Aydin, S., Keles, U., & Hasiloglu, A. (2012). Establishment for misconceptions that science teacher candidates have about geometric optics. *The Online Journal of New Horizons in Education*, *2*(3), 7–15.

Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, *32*(7), 939–961. https://doi.org/10.1080/09500690902890130

Dhakne-Palwe, S., Gujarathi, A., & 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 & Ethics, *5*(3), 254–259. https://doi.org/10.5958/2231-6728.2015.00052.9

Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). Four-Tier Diagnostic Test with Certainty of Response Index on the Concepts of Fluid. *Journal of Physics: Conference Series*, *1155*(1). https://doi.org/10.1088/1742-6596/1155/1/012078

Fariyani, Q., Rusilowati, A., & Sugianto, S. (2017). Four-Tier Diagnostic Test To Identify Misconceptions in Geometrical Optics. *Unnes Science Education Journal*, *6*(3), 1724–1729. https://doi.org/10.15294/usej.v6i3.20396

Fernandes, H. J. X. (1984). *Testing and Measurement*. Jakarta: Planning, Evaluation, and Development.

Flevares, L. M., & Schiff, J. R. (2014). Learning mathematics in two dimensions: A review and look ahead at teaching and learning early childhood mathematics with children's literature. *Frontiers in Psychology*, *5*(May), 1–13. https://doi.org/10.3389/fpsyg.2014.00459

Gooding, J., & Metz, B. (2011). From Misconceptions to Conceptual Change: Tips for identifying and overcoming students' misconceptions. *The Science Teacher*, (April/May), 34–37.

Kula, S., & Güzel, E. B. (2014). Misconceptions emerging in mathematics student teachers' limit instruction and their reflections. *Quality and Quantity*, *48*(6), 3355–3372. https://doi.org/10.1007/s11135-013-9961-y

Kusmaryono, I., Basir, M. A., & Saputro, B. A. (2020). Ontological Misconception in Mathematics. *Infinity*, *9*(1), 15–30.

Mohyuddin, R. G., & Khalil, U. (2016). Misconceptions of students in learning mathematics at primary level. *Bulletin of Education and Research*, *38*(1), 133–162.

Muijs, R. D., & Reynolds, D. (2005). *Effective Teaching. Evidence and Practice. (2nd ed.)*. Paul Chapman Publishing Ltd.

Musa, A., Shaheen, S., & Elmardi, A. (2018). Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. *Khartoum Medical Journal*, *11*(2), 1477–1468.

Ojose, B. (2015). Students' Misconceptions in Mathematics: Analysis of Remedies and What Research Says. *Ohio Journal of School Mathematics*, *72*(Fall), 30–34.

Özerem, A. (2012). Misconceptions In Geometry And Suggested Solutions For Seventh Grade Students. *Procedia - Social and Behavioral Sciences*, *55*, 720–729. https://doi.org/10.1016/j.sbspro.2012.09.557

Prabowo, A., Anggoro, R. P., Rahmawati, U., & Rokhima, N. (2019). Android-based teaching material for straight-sides solid. *Journal of Physics: Conference Series*, *1321*(3). https://doi.org/10.1088/1742-6596/1321/3/032097

Ravand, H., & Baghaei, P. (2020). Diagnostic Classification Models: Recent Developments, Practical Issues, and Prospects. *International Journal of Testing*, *20*(1), 24–56. https://doi.org/10.1080/15305058.2019.1588278

Retnawati, H., Arlinwibowo, J., & Sulistyaningsih, E. (2017). The Students' Difficulties in Completing Geometry Items of National Examination. *International Journal on New Trends in Education and Their Implications*, *8*(4), 28–41. Retrieved from https://ci.nii.ac.jp/naid/40021243259/

Reynolds, C. R., Livingston, R. B., & Willson, V. (2010). *Measurement and assessment*

in education (2nd ed.). New Jersey: Pearson Education.

Roselizawati, Sarwadi, & Shahril, M. (2014). Understanding Students' Mathematical Errors and Misconceptions: The Case of Year 11 Repeating Students. *Mathematics Education Trends and Research*, *2014*(June), 1–10. https://doi.org/10.5899/2014/metr-00051

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), 364–371. https://doi.org/10.5539/jel.v6n4p364

Sundaygara, C., Gusi, L. A. R. P., Pratiwi, H. Y., Ayu, H. D., Jufriadi, A., & Hudha, M. N. (2021). Identification students' misconception using four-tier diagnostic test on Newton Law subject. *Journal of Physics: Conference Series*, *1869*(1). https://doi.org/10.1088/1742-6596/1869/1/012157

Traub, R. E. (1997). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice*, *16*(4), 8–14. https://doi.org/10.1111/j.1745-3992.1997.tb00603.x