Developing an Online Mobile Based Diagnostic Instrument to Identify Science Misconceptions for Junior High School Students

Puji Rahayu[1,*] Supahar[2,] Nurhasanah[1,] Djuniar Rahmatunnisa Haristy[1]

[1] *Master of Natural Science Education,, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia*
[2] *Department of Physics Education,, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia*
[*]*Corresponding author. Email:pujirahayu.2019@student.uny.ac.id*

ABSTRACT

This research aimed to develop a test instrument in identifying misconceptions of three tier skills in junior high school science learning.. This study used instrument development model by Mardapi. The data method used a three tier misconception diagnostic test via Google Form. The research was conducted at SMPN 1 Kepung, SMPN 2 Kepung, and SMPN 2 Pare Kediri, East Java, involving 99 students of class VII IPA program. Analysis of the content validity used the Aiken V. The dichotomous item analysis used the 1-parameter Logistic Model (1 PL) or Rasch model. The instrument developed to identify misconceptions with three tier is declared valid and suitable for use. The diagnostic misconception instrument assisted by Google Form has met the expert validation with high validity. 10 items were developed, item 1 has a low difficulty level, so the item is too easy and not good to use in item number 2. While the other 9 items have a good difficulty level and can be used. The 10 items developed, the highest misconception was item number 1.

Keywords: Instrument, Misconception, Three tier, Mobile online, Science assessment.

1. INTRODUCTION

The 21st century, modernization and globalization are increasingly developing, causing many changes to influence human life. There is a very high influence on the development of technology and information that has influenced human life, both in the fields of economy, politics, art, culture, and even in the world of education. The development of technology and information today cannot be avoided, because the development of technology and information develop in line with the advancement of science [1]. It is hoped that the world of education in Indonesia will be able to face this change. Since mid-March, Indonesia has reportedly been affected by the corona virus pandemic, known as COVID-19 [2]. COVID-19 can strike every individual with a very fast chain of transmission. The impact of this pandemic is expected to be long lasting [3]. The government's way of minimizing the impact in terms of the education sector is by implementing online or online learning [4]. Technology is one of the facilitators who help facilitate the world of education.

Learning outcomes are strongly influenced by several aspects. One important aspect is assessment [5] [6].

The advent of computers in educational assessments has led to a comprehensive shift from paper-pencil based assessments to assessments administered on computers [7] [8]. Most of the assessment processes carried out so far are still conventional in nature, namely using paper (*paper-based tests*. Tests using paper are less practical in their use. The assessment process cannot be done in *real time*, students cannot know the test results directly and quickly, so Students' understanding of concepts cannot be immediately detected.The teacher also experiences difficulties in correcting answers and providing feedback to each student [9].

The use of a *mobile phone* or computer as a medium for conducting tests is an effort to familiarize students with technology interaction. The use of electronic tests has spread rapidly in recent years. Electronic-based testing offers direct scoring and feedback, and facilitates the use of individual testing

methods [10]. Computer-assisted tests can streamline teacher assessment time, and are able to diagnose student learning difficulties quickly and accurately based on completeness of indicators [11] With its various advantages, electronic assessment can be a medium in facilitating a more effective evaluation process.

One of the problems that are often encountered in the world of education and included in science learning is misconception [12] [13]. Misconceptions stem from false initial prejudice and naive beliefs [14]. Misconceptions develop from the wrong interpretation of students or errors occur in teachers in transmitting knowledge to students [15]. Misconceptions become a barrier for students when trying to understand the concepts of Natural Science (IPA) and can consistently affect the effectiveness of the learning process [16] [17].

Misconceptions that occur in students can be detected and overcome by conducting tests. It tests the diagnosis of student misconceptions in an accurate manner, helping teachers to design, improve teaching and develop effective learning strategies [18] [19]. Diagnostic tests are used to determine student weaknesses and as a basis for providing treatment to overcome student difficulties in learning something [20]. The results of diagnostic tests can be used as a reference for the implementation of learning according to students' abilities, including learning difficulties experienced [17].

There are various types of diagnostic tests, including: using semi-structured interviews, multiple choice tests, essay, *teststwo-tier multiplen choice tests and three-tier multiple choice tests.* Each form of diagnostic test has its own advantages and disadvantages [21]. The multiple choice test is *three tier* more effective because it is equipped with a confidence level scale to measure the level of confidence in the answers and reasons selected for one question [22].

Table 1. Interpretation of the results of the diagnosis of three tier misconceptions [23]

Category	Combination of Answers		
	Answer	Reason	Confidence
Miscconeption	False	Correct	Sure
	False	False	Sure
	Correct	False	Sure

2. RESEARCH METHOD

This type of development research used in this study aims to develop a test instrument to identify misconceptions in science learning skills in junior high school. This study uses an instrument development model by Mardapi [24]. This research was conducted at SMPN 1 Kepung, SMPN 2 Kepung, and SMPN 2 Pare Kediri East Java as many as 33 students for each school in grade VII students.

The development of the instrument follows the following steps: 1. The stage of making the instrument includes: (1) determining the objectives of the test, (2) determining the competencies to be tested, (3) determining the materials to be tested, (4) preparing the tests, (5) compile tests according to digital literacy indicators; 2. The test analysis phase includes: (1) validating the items, (2) developing the test items. The testing phase of this test includes: (1) determining the test subject, (2) conducting the trial, (3) analyzing the results of the trial.

The expert's assessment of the validity of the instrument was analyzed using the Aiken validation index. Aiken validation aims to measure the extent to which the items in the instrument meet the specific objectives of the preparation of test items [25].

$$V = \frac{\sum s}{n\,(c-1)} \qquad (1)$$

In which V is an index of expert agreement regarding the validity of the items. s is the number given by an expert minus the lowest validity rating. There are many experts. c is the highest validity score. Interpretation of the value of V, if the value of V is less than 0.4 it means that the validity is low, if the value is between 0.4-0.8 it is said that the validity is medium, and if it is more than 0.8 then the validity is high [25].

Analysis of dichotomous items was carried out using the 1 Parameter Logistic Model (1 PL) or the Rasch Model. This model only has one item parameter to show the characteristics of the item, namely the item difficulty level parameter [25]. The specific sample sizes for the 1 PL model range from 30-300. Item analysis was performed using the QUEST program. The item analysis carried out included model fit, item reliability, and item difficulty level.

3. RESULTS AND DISCUSSION

3.1. Test Development Results

The test instrument was developed based on the analysis of the three-tier aspects, indicators and criteria. The specification of the assessment

instrument in the form of a multiple choice test refers to the science learning material indicators. A total of 10 graded questions assisted by Google form were given to students and worked independently.

Students can directly access the Google Form website that has been provided through a digital device (mobile phone or laptop) that is connected to the internet. Students can immediately provide answers on the form that has been given. All responses and answers will be stored quickly and the teacher can easily find out the answers given by students. The use of Google Forms can also be set by giving the processing time so that when the time limit is up, the Google Form is automatically closed and students can no longer work on it. Google forms can also be a medium for students to get used to interacting with technology.

3.2. The Validation of Items Result

Validity testing is used to show how accurate the instrument is. The validity test shows the level of test accuracy in measuring the target to be measured [26]. At the instrument assessment stage, the content validity or content validity test was carried out. Content validity relates to the ability of the assessment tool to measure the content it should be. Evaluation of content validity using the Aiken validation index for each item.

There are 10 multilevel multiple choice items validated by experts. Assessment of the three tier misconception diagnostic test instrument was reviewed based on the suitability of the test with indicators, construction, and language. Three aspects (substance, construct, and language) determine whether a science learning achievement test is good or not. The substance aspect must represent the competence to be assessed, the construction aspect must meet the technical requirements according to the type of instrument used, and the language aspects used must be good, correct and communicative following the level of development of students [27]. Each test item is assessed based on three aspects of the assessment. The results of item validation can be seen in Table 2.

In accordance with the provisions of the Aiken V index, the results of the evaluation of all validities by 5 experts on 10 items have high validity. Valid

instruments influence data or information collected in the field [28]. The content validity based on Aiken's V was said to be good with a value of more than 0.80 [29]. The validity values of Aiken ranging from items 1 to 10 are in the range 0.9 to 1. This means that all

items have good content validity so that each item of the instrument is declared fit for use.

3.3 Test Results

3.3.1 Goodnes fit

Goodnes fit testing is carried out on the whole test or on each item. In Rasch analysis, the fit item quality model with the shortened item fit model [30]. It is determined that an item or test is declared fit with the model in the INFIT MNSQ range range from 0.77 to 1.30 and using INFIT t with a limit of -2.0 to 2.0, it is obtained items that correspond to the goodness of fit [25]. The assumption of validity in the Rasch model refers to INFIT MNSQ with a value ranging from 0.6 to 1.21 and OUTFIT MNSQ with a value of 0.11 to 1.17 [31]. Table 3 shows the estimated value of INFIT MNSQ of 0.99 with INFIT t 0.15 and OUTFIT MNSQ with an estimated value of 1.12 with INFIT t 0.56. The results showed that the overall fit test was used to measure students' *diagnostic misconception* skills.
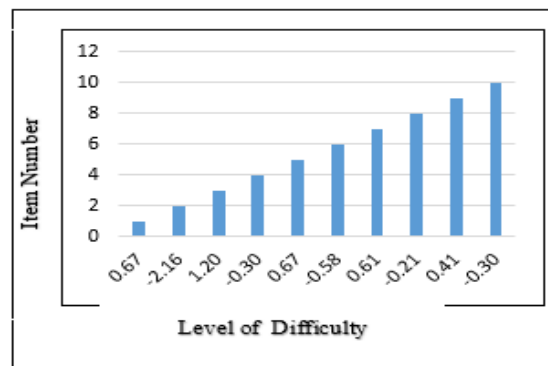

Figure 1 The level of difficulty item

Figure 1 also shown the reliability estimates with a value of 0.93. Reliability is one of the requirements for the instrument to be said to be feasible and can be used for measurements in field trials [23]. Reliability is adequate. The instrument has sufficient strength and reliability because it consists of items that have a high information function [27].

3.3.2 Difficulty levels of items

The difficulty level or difficulty on each item can be seen in the difficulty level index of the *Quest* software output shown in Figure 3.

Figure 3 shows the results of dividing the questions with difficulty levels ranging from easy questions to difficult questions. Items with a difficulty level of -2.00 indicate that the item is very easy, while items with a difficulty level of 2.00 indicate that the item is

very difficult [32]. Item 3 has a difficulty approaching +2 which means it has a high difficulty level. Problems 1, 4, 5, 7, 8 9, and 10 approach 0 means that the question has a difficulty level of 2 close to -2 indicating that it has an easy difficulty level.

Item is said to be good if the difficulty level is between -2 and 2 ($-2 \leq b \leq 2$) [25]. Item 2 has a very low difficulty level of -2.1 and 6 so the item is too easy and not good to use. At 1, 4, 5, 7, 8 9, and 10 because it has a difficulty level between -2 and 2, the item is said to be good and usable.

Based on the test results from Google Form, it can be identified that there are still many students who experience misconceptions on temperature and heat material. Students' answers with the highest percentage of misconceptions are shown in the indicator explaining the effect of heat absorbed by objects on temperature changes. In item 1 with cognitive domain C1, category of easy question. Students experience misconceptions is 56.4% in the first stage, for the reason level is 48.5%, and for the confidence level of 94.9% of students who answered confidently, because the students chose the wrong answer but gave the right reason with a confident level of confidence, this is because the student still not fully understand that when heating water continuously, the water temperature remains.

Table 2. Results of item analysis expert judgment assessment instrument

Number	Aiken V	Decision
1	1	high validity
2	0.95	high validity
3	0.95	high validity
4	1	high validity
5	1	high validity
6	1	high validity
7	1	high validity
8	1	high validity
9	0.95	high validity
10	1	high validity

Table 3. Goodnes fit and reliability analysis results

No.	Parameter	Item Estimates	Case Estimates
1.	INFIT MNSQ	0.99 ± 0.15	0.09 ± 0.33
2.	OUTFIT MNSQ	1.12 ± 0.56	1.12 ± 0.87
3.	Reliability estimates	0.93	

4. CONCLUSION

The test instrument was developed based on the analysis of the three-tier aspects, indicators and criteria. The specification of the assessment instrument be in the form of a multilevel multiple choice test refers with the science learning material indicators. A total of 10 Google form-assisted questions were given to students. The instrument developed to identify student misconceptions was declared valid and fit for use. The number of 10 items developed, item 1 has a very low difficulty level so that the item is too easy and not good to use in question number 2. While the other 9 items have a difficulty level between -2 and 2 which are categorized as good and usable. Of the 10 questions given the highest misconception occurred in item number 1.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Ngafifi, Kemajuan Teknologi Dan Pola Hidup Manusia Dalam Perspektif Sosial Budaya, Jurnal Pembangunan Pendidikan: Fondasi Dan Aplikasi 2(1) 2014 33-47. DOI: https://doi.org/10.21831/jppfa.v2i1.2616

[2] Gugus Tugas Percepatan Penangan COVID-19 Indonesia, Data COVID-19 Global dan Indonesia, 2020.

[3] Y.L. Hsu, Y.P. Chai, S. Vasoo, A midpoint perspective on the COVID-19 pandemic, Singapore Medika Journal 61(7) (2020) 1-10. DOI: 'https://doi.org/10.11622/SMEDJ.2020036

[4] C. Drane, L. Vernon, S. O'shea, The impact of 'learning at home' on the educational outcomes of vulnerable children in Australia during the CO VID 19 pandemic. Literature Review prepared by the National Centre for Student Equity in Higher Education, Curtin University, Australia, 2020.

[5] A.D. Astalini Kurniawan, U. Sulistiyo, and R. Perdana. E-Assement Motivation in Physics Subjects for Senior High School, Internasional Journal Of Online and Biomedical Engineering 15(11) 2019 2626-8493. DOI: https://doi.org/10.3991/ijoe.v15i11.10843

[6] M. Park, Effects of Simulation-based Formative Assesments on Students' Conceptions in Physics, EURASIA Journal Of Mathematics, Science and Technology Education15(7) 2019 1305–8223. DOI: https://doi.org/10.29333/ejmste/103586

[7] H. Budiman, Peran Teknologi Informasi dan Komunikasi dalam Pendidikan, Al- Tadzkiyah, Jurnal Pendidiman Islam 8(1) (2017) 31-43.

[8] S. Greiff, C. Niepel, R. Scherer, R. Martin, Understanding students' perfomance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files, Computer in Human Behavior 61 (2016) 36-46 DOI: https://doi.org/10.1016/j.chb.2016.02.095

[9] H. Sahidu, G. Gunawan, I. Indriaturrahmi, F. Astutik, Desain sistem e-assessment pada pembelajaran fisika di LPTK, Journal Pendidikan Fisika dan Teknologi 3(2) (2017) 255-265. DOI: https://doi.org/10.29303/jpft.v3i2.422

[10] H. Jeong, Behaviour & information technology a comparative study of scores on computer-based tests and paper-based tests, Taylor Fr 33(4) (2014) 37–41. DOI: https://doi.org/10.1080/0144929X.2012.710647

[11] U. Maier, N. Wolf, C. Randler, Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types, Computer Education 95 (2016) 85–98. DOI: https://doi.org/10.1016/j.compedu.2015.12.002

[12] D. S. Jubaedah, I. Kaniawati, I. Suyana, A. Samsudin, E. Suhendi, "Pengembangan Tes Diagnostik Berformat Four-Tier ntuk Mengidentifikasi Miskonsepsi Siswa Pada Topik Usaha Dan Energi," Seminar Nasional Ipa (E-Journal) 6(1) (2017) 32. DOI: 10.29303/jppipa.v6i1.314

[13] S. Anggrayni and F. U. Ermawati, "The Validity Of Four-Tier ' s Misconception Diagnostic Test For Work and Energy Concepts," in IOP Conf, Series: Journal Of Physics (2019). DOI: https://doi.org/10.1088/1742-6596/1171/1/012037

[14] C.P. Wijaya, H. Supriyono Koes, Muhardjito, The diagnosis of senior high school class X MIA B students misconceptions about hydrostatic pressure concept using three-tier,Jurnal Pendidikan IPA Indonesia 5(1) (2016) 14-21. DOI: https://doi.org/10.15294/jpii.v5i1.578

[15] N. Wahida, S. Saptono, Jurnal Of Innovative Science Education The Development of Three Tier Multiple Choice Test to Explore Junior High School Students' Science Literacy Misconceptions, Journal of Innovative 8(2) (2019) 190-198.

[16] P. Pujayanto, R. Budiharti, E. Adhitama, N. R. A. Nuraini, H. V. Putri, The development of a web-based asssessment system to indetify students' misconception automatically on linear kinematics with a four-tier instrument test, Physics Education 53(4) (2018). DOI: https://doi.org/10.1088/1361-6552/aac695

[17] T. Wahyuningsih, T. Raharjo, and D. Fitriana Masithoh, Pembuatan Instrumen Tes Diagnostik Ilmu Pengetahuan Alam (IPA) SMP Kelas VIII, Jurnal Pendidikan Pengetahuan Alam (IPA) 1(1), 2013) 111-117.

[18] D. Kaltakci-Gurel, A. Eryilmaz, L. C. McDermott, 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) (2017) 238-260. DOI: https://doi.org/10.1080/02635143.2017.1310094

[19] L. A. R. Laliyo, D. N. Botutihe, C. Panigoro, The development of two-tier instrument based on distractor to assess conceptual understanding level and student misconceptions in explaining redox reactions, International Journal of Learning, Teaching and Educational Research, 18(9) (2019) 216-237. DOI: https://doi.org/10.26803/ijlter.18.9.12

[20] D. Nugraeni, J. Jamzuri, and S. Sarwanto, Penyusunan tes diagnostik Ilmu Pengetahuan Alam (IPA) materi listrik dinamis, Jurnal Pendidikan Ilmu Pengetahuan Alam (IPA) 1(2) (2013) 12-16.

[21] P. Suparno, Miskonsepsi danPerubahan Konsep Dalam Pendidikan Ipa, Jakarta, PT. Grasindo, 2013.

[22] W. Bunawan, A. Setiawan, and, A. Rusli, Penilaian Pemahaman Representasi Grafik Materi Optika Geometri Menggunakan Tes Diagnostik, Jurnal Cakrawala Pendidikan 2(2) (2015) 257-267. DOI: https://doi.org/10.21831/cp.v2i2.4830

[23] H.O, Arslan, Cigdemoglu, and C.Moseley, A Three Tier Diagnostik Test to Asses Pre-Service Teacher' Misconception about Global Warming. Green Effect,Ozone Layer Depletion,and Acid Rain, International Journal of Science Education, 34(11) (2012) 667-1686.

[24] D. Mardapi, Pengukuran penilaian dan evaluasi pendidikan, Parama Publishing, 2008.

[25] E. Istiyono, Pengembangan instrumen penilaian dan analisis hasil belajar fisika, UNY Press, Yogyakarta, 2018.

[26] T. P. Siregar, E. Surya, E. Syahputra, Quality analysis of multiple choice test and clasical test at X grade students of senior high school, Journal IJARIIE Universitas Medan 3(2) (2017) 2153–2159.

[27] Supahar, Applying content validity ratios (cvr) to the quantitative content validity of physicslearning achievement tests, Yogyakarta (2015) 17–19.

[28] F. Kurnia, D. Rosana, Supahar, Developing evaluation instrument based on CIPP models on the implementation of portfolio assessment, in AIP Conference Proceedings, vol 12, AIP Publishing, College Park, Maryland, 2017, pp. 1999-2010. DOI: https://doi.org/10.1063/1.4995187

[29] S. Azwar, Penyusunan skala psikologi edisi kedua, Pustaka Belajar, Yogyakarta, 2017.

[30] S. Nuryanti, M. Masykuri, E. Susilowati, Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan, Jurnal Inovasi Pendidikan 4(2) (2018) 224–233. DOI: https://doi.org/10.21831/jipi.v4i2.21442

[31] C.Y. Huang, L.C. Tung, Y.T. Chou, W. Chou, K.L. Chen, C.L. Hsieh, Improving the utility of the fine motor skills subscale of the comprehensive developmental inventory for infants and toddlers: a computerized adaptive test, Disability Rehability 40(23) (2018) 2803–2809 DOI: https://doi.org/10.1080/09638288.2017.1356385

[32] E. Istiyono, D. Mardapi, S. Suparno, Pengembangan tes kemampuan berpikir tingkat tinggi fisika (PysTHOTS) peserta didik SMA, Jurnal Penelitan dan Evaluasi Pendidikan 18(1), (2014) 1–12. DOI: https://doi.org/10.21831/pep.v18i1.2120