DOI: 10.14421/impulse.2024.41-03

IMPULSE: Journal of Research and Innovation in Physics Education
Volume 4, Issue 1, 22 – 30

© ISSN (p): 2798–1762; ISSN (e): 2798-1754
http://ejournal.uin-suka.ac.id/tarbiyah/impulse

The Development of the Two-Tier Diagnostic Test Instrument with Google Form to Measure Student Misconceptions on Energy and Energy Forms

Ika Thalia Pratiwi[1], **Viyanti**[1*], **Dimas Permadi**[1]

[1] Department of Physics Education, Universitas Lampung, Indonesia
*Corresponding author: viyanti.1980@fkip.unila.ac.id

ABSTRACT

This study develops a two-tier diagnostic test instrument with Google Forms- to measure student misconceptions about energy and energy forms. The development research used the 4-D model with 4 stages: (1) define; (2) design; (3) develop; and (4) disseminate. The researchers validated the product to evaluate the construct, material, and language aspects. Expert validation results declared the two-tier diagnostic test instrument very valid at 87.7%. The researchers tested the two-tier diagnostic test instrument with Google form to measure student misconceptions on 30 students and then analyzed using the Rasch model with the assistance of Ministep 5.6.2 software. Based on the results of the Rasch analysis, 20 valid items were obtained. The questions on the two-tier diagnostic test instrument to measure student misconceptions were reliable, with a Cronbach alpha value of 0.66, categorized as adequate. The final product of the developed instrument met the standards of instrument validity and reliability.

INTISARI

Penelitian ini bertujuan untuk mengembangkan instrumen tes diagnostik two tier berbasis google form untuk mengukur miskonsepsi siswa pada materi energi dan bentuk-bentuk energi. Penelitian pengembangan menggunakan model 4-D dengan 4 tahapan yakni: (1) define; (2) design; (3) develop; (4) disseminate. Validasi produk untuk menilai aspek konstruk, materi, dan bahasa. Berdasarkan hasil validasi ahli instrumen tes diagnostik two tier dinyatakan sangat valid sebesar 87,7%. Instrumen tes diagnostik two tier berbasis google form untuk mengukur miskonsepsi siswa diujicobakan kepada 30 siswa dan selanjutnya dianalisis menggunakan model Rasch dengan berbantuan software Ministep 5.6.2. Berdasarkan hasil analisis rasch diperoleh sebanyak 20 butir soal yang valid. Soal-soal pada instrumen tes diagnostik two tier untuk mengukur miskonsepsi siswa dinyatakan reliabel dengan nilai alpha Cronbach sebesar 0,66 pada kategori cukup. Produk akhir dari instrumen yang telah dikembangkan telah memenuhi standar kelayakan instrumen yaitu valid dan reliabel.

ARTICLE HISTORY
Received: March 27, 2024
Accepted: July 01, 2024

KEYWORDS:
Assessment instrument, energy, google form. misconception

KATA KUNCI:
Energi, google form, instrumen tes, miskonsepsi

* Corresponding author:
Viyanti, Department of Physics Education, Universitas Lampung, Indonesia
✉ viyanti.1980@fkip.unila.ac.id

A. Introduction

Physics is the study of natural theories and concepts with human-interpreted explanations. Learning concepts, using them to solve physics questions, and doing science exercises are all crucial parts of studying physics [1]. Students find it challenging to understand physics principles because most students cannot connect between the learned materials and real-world applications. As a result, pupils make conceptual mistakes and come up with substitute ideas that lead to misunderstandings.

Misconceptions refer to discrepancies between the understanding of physics and the scientist's perception of the world. One of the challenges associated with studying physics is the presence of misconceptions, unnoticed by students [2]. The general procedures to identify the type of misconceptions are important to prevent misunderstandings such as identifying the cause, and choosing the best course of action [3]. A diagnostic test is one tool for dispelling myths. As a result, the creation of diagnostic exams that may gauge students' misconceptions is crucial to the assessment procedure and identifying misconceptions. The importance of this instrument development is crucial for SMAN 1 Kotabumi, which lacks a diagnostic test tool in terms of energy materials and forms of energy.

A diagnostic test identifies the pupils' areas of material understanding difficulty in a specific field of study. There are two types of diagnostic tests: cognitive and non-cognitive [4]. Diagnostic tests provide a comprehensive picture of students' cognitive preparation for learning, whereas non-cognitive diagnostics seek to ascertain students' psychological and social well-being as well as their learning preferences, personalities, and interests.

Education frequently uses technology, particularly for learning assessment and media [5]. Assessments examine the extent of changes in student learning outcomes and provide feedback to improve the learning process. The implementation procedure for assessments in the form of tests continues to rely on paper and stationery to support the assessment [5]. However, the high cost of money to duplicate questions and the prolonged time to prepare the answer sheets make the paper test ineffective. On the other hand, the implementation of the Internet for evaluation is excellent and useful such as with Google Forms. The Google Form is one tool available on the Google website for quizzing students or quickly and easily gathering information.

The capacity to do or create anything is known as energy [6]. Energy is necessary for all life. Humans use energy to drive motorcycles, cars, airplanes, and other vehicles in their daily lives since power can occur in these situations. Fitri & Oktavianty [7] and Fujiyati [8] revealed misconceptions about the correlation between effort and energy materials. Table 1 shows this matter.

Table 1. Misconceptions of Work and Energy

Literacy	Definition
Enterprises	Correlation between effort, force, and displacement
Kinetic Energy	Correlation between an object's position and kinetic energy
Potential Energy	Correlation between potential energy and mechanical energy
Law of Mechanical Energy Conservation	Objects have different amounts of energy

B. Method

This research and development created a diagnostic test instrument, consisting of questions with five answer choices. The first tier of the test measures students' misconceptions about energy and its forms. The second tier includes five choices of reasons that correspond to the first-tier answers. The researchers collected the data for this research using a Google Form. This research adopted Thiagarajan's 1974 4-D model, comprising four distinct stages: 1) Define the process, including the necessity analysis in terms of learners and learning objective; 2) Design the process, after the creation of diagnostic test in the form of a two-tier test on various energy materials and energy types; 3) Developing process, validating the materials by experts, revising, promoting limited trial run, and end-product revising; 4) Disseminating process, distributing products.

The data collecting instruments in this research included a questionnaire for teacher necessity analysis and a sheet for expert validity testing. The data collection strategies were obtaining data on validity and reliability test results. Three validators, specifically two physics education lecturers and one physics subject instructor, conducted the validation process. The researchers examined the data using a Likert scale score consisting of four levels: 1, 2, 3, and 4.

$$P = \frac{Total\ score\ obtained}{Sum\ of\ the\ highest\ scores} \times 100\% \qquad (1)$$

Description:
P = feasibility percentage

Table 2. Result Criteria Percentage of Feasibility

Percentage	Result Criteria
25% - 43.75%	Invalid
43.76% - 62.50%	Fairly Valid
62.51% - 81.25%	Valid
81.26 % - 100%	Very valid

The empirical validity test used the Rasch model with Ministep 5.6.2 software. This Rasch model could determine the interaction between respondents, items, and criteria for checking the suitability of items. Table 3 shows the item fit criteria.

Table 3. Item Fit Criteria

Percentage	Criteria
25% – 43.75%	Invalid
43.76% – 62.50%	Fairly Valid
62.51% – 81.25%	Valid
81.26 % – 100%	Very valid

Cronbach Alpha formula is useful to determine the reliability of the Rasch model. The researchers used the categories of reliability by the values of Cronbach Alpha. Table 4 shows the criteria.

Table 4. Item Reliability and Person Reliability Criteria

Value	Criteria
>0.94	Special
0.91 – 0.94	Very good
0.81 – 0.90	Good
0.67 – 0.80	Simply
>0.67	Weak

C. Result and Discussion

This research developed a two-tier diagnostic test instrument with Google Forms to measure student misconceptions about energy materials and evaluate the validity and reliability of various energy sources. The instrument product test was divided into two stages: the defining stage, including both theoretical and empirical investigations. The researchers conducted an empirical study at SMAN 1 Kotabumi to analyze the necessity of certain factors by distributing a questionnaire to three teachers at the school. The questionnaire comprised three analyzed aspects: the learning process, the online platform, and instrument development requirements. The preliminary study yielded field data that substantiated the research. Based on the questionnaire, the mean score of the development was 0.64, indicating the necessity of developing a diagnostic test instrument. Potential and present issues in the domain, specifically the failure of teachers to assess students' comprehension levels and devise efficient learning techniques, became the necessities of the development. The educators did not use internet tools for conducting evaluations. Teachers encountered challenges when using the two-tier diagnostic test instrument with Google Forms to assess student misconceptions.

The second step was designing stage. The researchers determined the instrument's structure in the design step by creating a two-tier diagnostic test instrument based on the learning objectives. The researchers designed the diagnostic

exam questions in a two-tier format, considering cognitive capacities derived from Bloom's taxonomy and concepts associated with energy and its various manifestations. The researchers divided the diagnostic exam questions into two tiers. The first tier consisted of questions with five answer choices, while the second tier with five corresponding options to the given answers in the first tier. Additionally, the test instrument provided rubrics, scoring criteria, and recommendations for answering questions. Therefore, the researchers divided the text into three sections: the first section consisted of the cover, preface, table of contents, and rationale. The content section consisted of a grid, instructions, instrument form, instrument rubric, instrument scoring criteria, and instrument recapitulation. The last section of the text comprised recommendations and a bibliography.

The next phase involved development. The stages of product development included preparing test instrument parts, consisting of three components. Instrument-shaped writing implements are digital assessments that include student identification, question instructions, and 2-tier 20 multiple choice questions put on Google Forms. The initial section comprises the personal information of the learners, while the subsequent section encompasses the questions.

An expert conducted a rigorous validity test after designing the instrument. This stage involved the evaluation of the instrument's validity by two highly knowledgeable physics education professionals with a specialization on the instrument development and also physics subject teachers. The evaluation covered three key aspects: construct, material, and language. The researchers used quantitative data in the form of Likert scale scores to assess the validity of the expert test results. Then, the researchers measured three different elements using the Likert scale, consisting of four response options: 1, 2, 3, and 4. Table 5 shows the results of the validity test.

Table 5. Test Instrument Expert Validity Results

Aspect	Expert			Max Score	Assessment Percentage	Category
	1	2	3			
Construct	27	30	30	36	80,3%	Highly valid
Material	32	34	35	36	93,4%	Highly valid
Language	9	10	12	12	89,6%	Highly valid
Average assessment percentage					87,7%	Highly valid

The validity test value from the constructed aspect is 80.3%, categorized as very high [9], and is valid with minor revisions. The material aspect is 93.4%, categorized as very high [9]; and is valid with minor revisions. The language aspect is 89.6%, categorized as very high [9]; and is valid with minor revision. The researchers revised the product based on the suggestions and improvements of the validator. After the revision, the researcher conducted a field trial to test the empirical validity and

reliability of the developed instrument. The researchers promoted the test after the three validators declared the product was valid. Then, the researchers involved 30 students of X-2 at SMAN 1 Kotabumi.

Analysis of empirical validity or the level of item fit using the Rasch model with the assistance of Ministep 5.6.2 software. Table 6 shows the results of the empirical validity analysis.

Table 6. Item Fit Analysis on Diagnostic Test Instruments

Measure	Outfit		PT-Measure Corr	Item
	MNSQ	ZSTD		
5.05	0.97	0.25	0.69	S18
0.68	1.26	0.71	0.43	S1
032	0.87	-0.12	0.45	S10
0.23	0.94	0.5	0.42	S14
0.18	1.26	0.64	0.29	S12
0.09	0.98	0.14	0.37	S4
-0.15	1.17	0.48	0.29	S15
-0.21	0.94	0.11	0.33	S5
-0.31	0.72	-0.28	0.33	S2
-0.31	1.01	0.23	0.29	S20
-0.37	0.83	-0.7	0.31	S6
-0.37	1.02	0.24	0.29	S17
-0.43	1.23	0.54	0.24	S11
-0.43	1.04	0.28	0.27	S16
-0.49	0.82	-0.5	0.30	S3
-0.49	0.89	0.6	0.26	S8
-0.49	0.74	-0.19	0.30	S13
-0.84	0.83	-0.6	0.26	S7
-0.84	0.66	-0.38	0.30	S9
-0.84	0.71	-0.28	0.26	S19

The outfit mean squared (Outfit MNSQ) value is in the interval 0.66 <MNSQ> 1.38, indicating the accurate measurement of the product to assess the students. The Outfit ZSTD value is -0.37<ZSTD>0.64, indicating the rational probability value of the data. The PT Measure Corr value is in the interval 0.24 <PT Measure Corr> 0.69, indicating the normal distribution of the data. Based on these three criteria, all items met the Rasch model.

The obtained analysis on item fit is in the form of item suitability based on Bond and Fox (2015): (1) the outfit mean square (MNSQ) value is accepted 0.5 < MNSQ < 1.5; (2) the Outfit Z-standard (ZSTD) value is accepted -2.0 < ZSTD < +2.0; (3) the Point Measure Corr value is accepted 0.4 < Pt Measure Corr < 0.85. If the test item meets at least one of the criteria, then the item or statement is applicable or valid. Azizah and Wahyuningsih also explain the determination of fit items must at least meet one of the criteria [10]. Based on these criteria, the items of the test instrument met one of the conditions so that the instrument was valid.

An instrument is reliable if the instrument shows the same results after being repeatedly used [14]. Reliability validity analysis using Rasch model with the assistance of Ministep 5.6.2 software. Table 7 shows the results of the empirical validity analysis.

Table 6. Item Reliability and Pearson Reliability Analyses

Analysis	Value	Question	Conclusion
Person Reliability	0.88	1-20	Average reliability
Item Reliability	0.95		
Alpha Cronbach	0.66		

The INFIT MNSQ and OUTFIT MNSQ values increase by 0.98 and 0.94 because these values are close to the ideal of 1.00 and INFIT ZSTD. The OUTFIT SZTD values increase by 0.06 and 0.12 because these values are close to the ideal of 0.00, so the reliability of the question items is excellent. Furthermore, the item reliability value is 0.95, meeting the special criteria [11]. The grouping of respondents on the item question is quite good, 4.27, indicating the capability of the item question to measure respondents with low to high abilities. Respondents' interaction with the question items is moderate because the Cronbach alpha value is 0.66, average. Tarigan et al also explain that high or low reliability is empirically indicated by a number called the reliability coefficient [12]. Based on item reliability, item reliability, and Conbach's alpha value, the two-tier diagnostic test instrument is acceptable. This is in line with Bond and Fox's statement that the item and respondent reliability index is acceptable if it is more than 0.8 [15]. The last stage is dissemination. After the test instrument is valid based on the experts and empirically valid and reliable, the developed instrument was declared a standardized product.

D. Conclusion

The development resulted in a two-tier diagnostic test instrument based on Google Forms, designed to measure student misconceptions about energy and various forms of energy. The instrument includes a grid, instructions for use, an instrument form, answer guidelines, and a score recapitulation. Experts declared the test instrument valid in terms of construct, substance, and language. The Minister 5.6.1 software found the two-tier diagnostic test instrument, based on Google Forms, to be empirically valid and reliable for measuring student misconceptions about energy and forms of energy based on the empirical validity standards in the very valid category and average category reliability.

Acknowledgments

The researcher would like to thank the 2020 physics education study program students who have provided assistance and motivation so that this research can be completed.

References

[1] Puspitasari, A. D. (2019). Penerapan media pembelajaran fisika menggunakan modul cetak dan modul elektronik pada siswa SMA. *Jurnal Pendidikan Fisika*, *7*(1), 17-25.

[2] Suparno, Paul. 2013. Miskonsepsi dan Perubahan Konsep dalam Pendidikan Fisika. Jakarta: PT Grasindo

[3] Ali, Muhammad. (2019). Analisis Miskonsepsi Siswa Berdasarkan Gender Dalam Pembelajaran Fisika Dengan Menggunakan Tes Diagnostik Two-Tier di Kotabaru. *Cendikia: Jurnal Ilmiah Pendidikan*. 7(1).59-66.

[4] Dasar, D. S. (2020). Asesmen Diagnostik. Direktorat Sekolah Dasar, Direktorat Jenderal PAUD Dikdas Dan Dikmen, Kementrian Pendidikan Dan Kebudayaan, Riset Dan Teknologi.

[5] Sahidu, H., Gunawan, G., Indriaturrahmi, I., & Astutik, F. (2017). Desain sistem e-assessment pada pembelajaran fisika di LPTK. *Jurnal Pendidikan Fisika dan Teknologi*, No 2. Vol 2.

[6] Khoiriyah, I. Z., Faizah, S. N., & Mubin, M. (2019). Efektivitas Metode Inquiry Learning Terhadap Hasil Belajar Ipa Pada Tema Energi dan Perubahannya. *At-Thullab: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, *3*(2), 52-62.

[7] Fitri, S., & Oktavianty, E. (2019). Remediasi Miskonsepsi Menggunakan Model Pembelajaran REACT Tentang Usaha Dan Energi Di SMA. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa (JPPK)*, *8*(9)

[8] Pujiyati, P. (2018). *Pengembangan instrumen tes diagnostik two tier untuk mendeteksi miskonsepsi fisika peserta didik sma 3 mataram tahun ajaran 2017/2018* (Doctoral dissertation, Universitas Mataram).

[9] Arikunto, S (2012) Dasar-Dasar Evaluasi Pendidikan Edisi 2. Jakarta: Bumi Aksara.

[10] Azizah, A., & Wahyuningsih, S. Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *Jupitek*, *3*(1), 45-50.

[11] Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch Pada Assessment Pendidikan. Cimahi: Tim komunikata.

[12] Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, *16*(2), 92-96

[13] Thiagarajan, S., Semmel S.D and Semmel M.I. (1974). Instructional Development for Training Teachers of Exceptional Children. Minneapolis, Minnesota: Leadership Training Institute Education, University of Minesota.

[14] Rosidin, U. (2017). *Evaluasi Dan Asesmen Pembelajaran*. Media Akademi.
[15] Bond, T. G., & Fox, C. M. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences (3rd ed.). Mahwah, NJ: L. Erlbaum.