Jurnal Phi

Article Type: *orginial research*

Development of E-Module Based Physics Test Instrument: Multiple Choice Test

Melani Dyah Ayu Sukma[1], Kuncoro Asih Nugroho[2], Bayu Setiaji[3]

[1,2,3]Department of Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta

Corresponding e-mail: melanidiah11@gmail.com

KEYWORD: Multiple Choice, E-Module, Instrumen Tes

Submitted/Received: 10 November 2024
Revised: 16 January 2025
Accepted: 21 January 2025
Published: 23 January 2025
First Available Online: 23 January 2025

ABSTRACT. The purpose of this study was to develop a question instrument and describe the validity of reliability, and item analysis. The research design used in developing the instrument is RnD with the 4D method. Based on the results of the analysis and discussion, it can be concluded that the results of expert validation meet the valid criteria. The empirical validity obtained states that it meets the validity criteria with a good category. The results of the reliability test on the product show that the test instrument is quite reliable. The results of the item analysis stated that the instrument was in the category of good difficulty level. In the analysis of item discrimination, there are 15 categories of questions that are feasible to maintain. In the analysis of item distractors, there are 15 questions with distractors that are accepted because they are good enough and there are 5 questions whose distractors are revised. This research is useful for knowing how the level of quality and effectiveness of the development of test instruments.

1. Introduction

The advancement of increasingly sophisticated technology has many positive impacts on the progress of the world of education. Education is a fundamental aspect of life for national development which plays an important role in preparing the quality of human resources capable of mastering science and technology (Sulistiyono, Mundilarto, 2019). Improving the quality of education is a process that is integrated with the process of improving the quality of human resources. This is because improving the quality of education and human development is a dominant factor in the ability of humans to deal with problems in everyday life. The learning process has three components, namely objectives, learning activities and evaluation. In the aspect of

assessment, it must be able to provide a comprehensive picture or information related to the development of students (Fatwa Tentamaa, Purwatib, Subardjoc, 2023; Reni Wahyuningsih, Sri Wahyuni, 2013). To get the maximum assessment, an assessment instrument is needed that can be used to obtain information related to students' understanding, namely instruments in the form of multiple choice questions.

One of the objectives of physics learning is to develop the ability to reason in deductive thinking by using physics concepts and principles to explain various natural phenomena and solve their problems (Dian Ratih Utama Sari, Sri Wahyuni, 2018). Therefore, teachers as teachers are required to be able to prepare students who have good concept understanding skills, especially in physics subjects. The implementation of the physics learning process often experiences errors in understanding concepts (Nadia, 2019; Pratiwi, 2016). Errors that occur in understanding concepts are acted out by the difference in concepts received by students against a concept that is agreed upon and considered correct by experts. This incorrect understanding will get worse when considering the learning process is a transfer of knowledge from teachers to students, even though students already have prior knowledge that forms the concept of a material, then the learning that takes place can be more meaningful and can be used to overcome learning difficulties. Therefore, a test instrument is needed that can be used to measure how deep the level of understanding in students.

The use of multiple-choice diagnostic tests can show the characteristics of concept understanding in students (Kang et al., 2010; Pratama, 2023). Diagnostic tests are used to find out the weaknesses of students so that they can be given appropriate treatment (Pratiwi, 2016). This is in line with research conducted by (Mita Rahmani, Kurnia Ningsih, 2015)mita ra that diagnostic tests can be used to see how the level of difficulty of students in the learning process. This shows that diagnostic tests can be used as instruments to describe the level of understanding of students' concepts in physics material. The researcher's question used in this study is to find out how the level of effectiveness of the e-module-based multiple choice instrument in physics learning.

2. Research Methode

This research is part of development research, namely the development of diagnostic tests using the 4-D development design. The research stages carried out in this study are the result of adaptation and modification of Thiagarajan. This research was conducted using 36 research subjects. The stages in this study are as follows:

The defining stage (Define), which includes front end analysis, student analysis, concept analysis task analysis, and formulation of learning objectives. In this study, the material used is limited to temperature and heat material.

The design stage, which includes the preparation of learning outcome test instruments that are in accordance with the indicators of students' concept understanding with the subject matter of temperature and heat, in addition to media selection, format selection, and initial design.

The development stage (Develop), which includes the validity of the instrument carried out by expert validators or in this study conducted by two lecturers of the Physics Education Department, and two high school teachers. Testing the results of empirical validity is done by comparing test scores based on research results and test scores from teachers as validators. The reliability test of the test instrument was analysed using KR-20. Item analysis includes the level of difficulty of the questions used, the differentiating power of the questions and the analysis of question exemptions. Analysis of the ability to understand concepts in students is analysed using data sources of research results that have been carried out by students.

The data analysis technique to measure expert validity was calculated using the equation:

$$\text{Average of validators in all aspects } V_a = \frac{\sum_{i=1}^{n} A_i}{n}$$

The validity criteria for determining the level or level of decision on the validity of the concept understanding test instrument can be seen in the table below:

Tabel 1. validation category by (Hobri, 2010:52)

Kategori Validitas	Interval
Tidak Valid	$1 \leq Va < 2$
Kurang Valid	$2 \leq Va < 3$
Cukup Valid	$3 \leq Va < 4$
Valid	$4 \leq Va < 5$
Sangat Valid	$= 5$

The data analysis technique used for empirical validity is the Current Validity method:

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} \quad (1)$$

The data analysis technique for item validity uses the correlation formula:

$$\gamma_{pbi} = \frac{M_p - M_t}{s_t}\sqrt{\frac{p}{q}} \quad (2)$$

The reliability data analysis technique uses the KR-20 equation with the following equation:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{s_t^2 - \sum p_i q_i}{s_t^2}\right) \quad (3)$$

Data analysis techniques to determine the level of difficulty of the question using the following equation:

$$P = \frac{\sum B}{N} \quad (4)$$

The data analysis technique for question differentiating power is as follows:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B \quad (5)$$

The data analysis technique for the question's exception index was analysed using the following equation:

$$IP = \frac{P}{(N-B)(n-1)} \times 100\% \quad (6)$$

The technique of analysing the ability to understand concepts in students has the highest score (correct value = 4) multiplied by the whole question (20 questions). So that the maximum score is 4×20 = 80, with a minimum score of 0. In this study, interval 4 was used with the following criteria:

Tabel 2. Criteria Interval by (Lewy, 2009)

Nilai Peserta Didik	Kriteria
61-80	Sangat Baik
41-60	Baik
21-40	Cukup
0-20	Kurang

The Disseminate stage, which was carried out by researchers only on a limited basis with the subject used was students in one of the high schools in Bantul, and the socialisation of the use of e-modules based on local wisdom.

3. Results and Disscusion

The results of this study are in the form of products from the development of multiple choice test instruments that are useful for measuring the level of understanding of the concepts of students on temperature and heat material. Based on the results of research and analysis obtained some results related to the results of validation, reliability, etc. related to research.

Based on the results of the expert validity analysis of the question as an instrument that is useful for measuring the level of understanding of the concept of students, quantitative and qualitative data are obtained which are contained in the figure below.

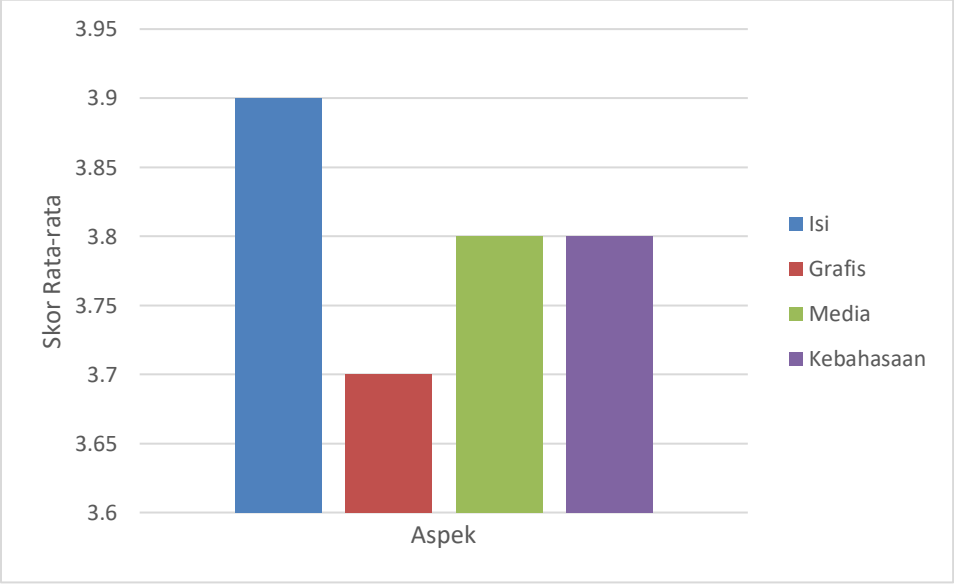

Figure 1. expert validation analysis results

Based on the results of expert validation data analysis, the average value is 3.8 with a valid category, so this indicates that the test instrument in the form of multiple choice questions meets the valid criteria for use. Questions with valid categories or have high validity criteria indicate that the questions used can measure the expected competencies (Dian Ratih Utama Sari, Sri Wahyuni, 2018) or in this study the expected competence is the understanding of the physics concepts of students. Meanwhile, questions that have an invalid question validity category or have low validity indicate that the question cannot be used to measure the expected

competencies (Mita Rahmani, Kurnia Ningsih, 2015; Prihono, 2019) .So it can be concluded that to measure the level of competence of students must use valid test questions or instruments.

The empirical validity data that has been analysed is used as a reference to interpret the test implementation using the test instrument in the form of multiple choice questions that have been developed. The result of the empirical validity analysis is r_xy=0.84 with valid criteria. The value that is getting closer to 1 means that the relationship is getting stronger which means that the instrument is able to measure what should be measured properly (Bismy Hafizha Mayara, Emma Yuniarrahmah, 2016; Fatwa Tentamaa, Purwatib, Subardjoc, 2023) .The empirical validity test used aims to compare the results of the evaluation conducted by the teacher with the results of the concept understanding test conducted by students. The high and low correlation coefficient obtained indicates the high and low validity of the test that will be assessed for quality (Dian Ratih Utama Sari, Sri Wahyuni, 2018).

The data on the results of the validity of the question items have four criteria that are used as a reference, namely very valid, valid, quite valid, and less valid. Based on the results of the analysis on the test instrument in the study, the following results were obtained:

Tabel 3. question item validation results

INFIT MNSQ	INTERPRETASI	INFIT MNSQ	INTERPRETASI
0,88	VALID	1,07	VALID
1,09	VALID	0,99	VALID
1,13	VALID	1,24	VALID
0,96	VALID	1,06	VALID
0,94	VALID	0,99	VALID
0,95	VALID	0,95	VALID
1,11	VALID	0,83	VALID
0,9	VALID	0,94	VALID
0,97	VALID	1,1	VALID
1,11	VALID	0,94	VALID
1,14	VALID	0,88	VALID
0,87	VALID	1,01	VALID
0,85	VALID	1,16	VALID
1,05	VALID	0,83	VALID
0,92	VALID	0,87	VALID
0,65	NOT VALID	0,76	NOT VALID
0,7	NOT VALID	0,73	NOT VALID
0,54	NOT VALID	0,62	NOT VALID
0,92	NOT VALID	0,7	NOT VALID
0,62	NOT VALID	0,75	NOT VALID

Based on the analysis, all numbers have represented questions that have valid criteria, which means that the questions can be maintained and are suitable for use as data collection instruments on tests and do not need any improvement. If there are questions that have invalid criteria then, the question must be corrected (Keumala et al., 2019). Based on the results of the data on the validity of the question items that have been analysed, it can be concluded that the question items in this study are included in the questions that are quite good, this is based on the level of validity produced.

The reliability test of the questions that have been carried out produces results that this reliability result data meets the reliability criteria with a reliable category with a value of 0.71. (Marthunis et al., 2015) stated that a reliable question instrument means that the instrument used is mature enough to be used in collecting research data, so that the data is able to capture reliable data results.

Item analysis which includes question differentiator analysis, exemption analysis and question difficulty level analysis. The results of the analysis on the question difficulty level data have the following values:

Tabel 4. question difficulty level

Persentase	Interpretasi
15%	Sulit
60%	Sedang
25%	Mudah

The results of the analysis as in the table above, the difficulty level of the questions should ideally be in the range of 30% to 70%. Questions with difficulty levels outside this range need to be revised to ensure the validity and reliability of the test (Abdul Qadir, Nuril Huda, 2024). Based on these results, it can be concluded that the question has good quality because most of the questions have a proportion of questions in the medium - difficult category. According to Sidabutar et al (2017), a question can be said to be good and feasible if it has a proportion value of 25% for questions with easy criteria, 50% for medium questions and 25% for difficult questions. Good question items are items that are not too difficult and also not too easy (Arikunto, 2011: 207).

The differentiating power of the questions based on the analysis results has a category with the following proportions:

Tabel 5. Interpretation of Question Distinguishing Power

Persentase	Interpretasi
30%	Baik
50%	Cukup
20%	Jelek

Question discriminating power is the ability of the question to distinguish students between those who are good and those who have low abilities (Arikunto, 2011: 211). Analysis of the question's differentiating power is carried out by dividing students into two groups, namely the upper group and the lower group (Dian Ratih Utama Sari, Sri Wahyuni, 2018). The results of the analysis divide the level of question discriminating power into three categories, namely good, fair and poor categories. Good discrimination power is discrimination power that has items that are not too easy and not too difficult (Ambarwati & Ismiyati, 2021). Question items that have a

poor discrimination index can be discarded immediately, while other items can be retained or revised for improvement (Ardhani, 2020). Thus, differentiating power analysis is important to ensure the quality of items in learning evaluation, so that they can effectively distinguish between high and low ability students.

The results of the analysis of the differentiating power of the questions as listed above show that the results of the interpretation of the good, sufficient and poor categories vary (Eliza Pradita, Priarti Megawanti, 2023). The good category interprets that the question is effective for distinguishing students with high and low abilities, the moderate category interprets that the question can distinguish the abilities of students with moderate abilities, and the poor category interprets that the question is unable to distinguish between students with high and low abilities.

The data on the results of the analysis of question exemptions in this study are divided into three categories, namely the good category, the sufficient category, and the poor category. The results of the analysis are as follows:

Tabel 6. Interpretation of Question Excerpts

Persentase	Interpretasi
30%	Baik
45%	Cukup
25%	Kurang Baik

A distractor can be said to function well if there is at least 1 person who chooses that answer or as many as 5% of the total test participants (Amelia, 2014; Wibawa, 2019).The effectiveness of a good distractor will make the item also have a good category. A good question is a question that has a distractor percentage of 1 per number of alternative choices provided (Muh Syahrul Sarea, 2022; Yani, 2018). Question items that are classified as poorly effective must be corrected by paying attention to distractors or other alternative answer choices so that they can function properly (Eliza Pradita, Priarti Megawanti, 2023). So that from the results of the analysis it is known that there are still questions that must be revised.

Conclussion

Based on the results and discussion in the research on the development of multiple choice test instruments on e-module-based physics learning in high school, it can be concluded that: (1) Data from expert validation results have valid criteria. The results of the reliability test on the product produced a reliable category; (2) Item analysis on the analysis of the difficulty level of the question produced a good category, with a proportion of 60% of questions in the moderate category, 25 in the difficult category, and 15% of questions in the easy category. In the analysis of the differentiating power of the questions based on the results of the analysis has a category with a proportion of good categories as much as 30%, sufficient categories as much as 50%, and poor categories have a proportion of 20% which indicates that as many as 15 questions are suitable for use. The results of the analysis of the question examiner found that there were 15 questions accepted with good enough distractors and as many as 5 questions were revised. This research is expected to be a source of reference in developing e-module-based test instruments.

Based on the research and development carried out, suggestions that can be given include the instrument to be used should be tested in many schools with different materials to determine the effectiveness of the product, and making questions with a number that is not tight so that when there are questions that are discarded there are still questions that are retained.

Acknowledgements

Our gratitude goes to the Department of Physics Education UNY and final thesis supervisors, as well as the high school where this research was conducted.

References

Abdul Qadir, Nuril Huda, D. H. (2024). Analisis Butir Tes : Tingkat Kesukaran, Daya Pembeda Dan Efektivitas Pengecoh. *Al Furqan: Jurnal Agama, Sosial Dan Budaya*, *3*(3), 1450–1467.

Ambarwati, Y. F., & Ismiyati, I. (2021). Analisis Butir Soal Pilihan Ganda Ulangan Akhir Semester Genap Mata Pelajaran Kearsipan. *1*(2), 64–75.

Amelia, M. A. (2014). Analisis Soal Tes Hasil Belajar High Order Thinking Skills (Hots) Matematika Materi Pecahan Untuk Kelas 5 Sekolah Dasar.

Ardhani, Y. (2020). Kualitas Butir Soal Penilaian Akhir Tahun Mata Pelajaran Teknologi Dasar Otomotif Kelas X Teknik Kendaraan Ringan Otomotif Di Smk Muhammadiyah Gamping Periode 2018/2019. *3*(November), 85–94.

Bismy Hafizha Mayara, Emma Yuniarrahmah, dan M. D. M. (2016). Hubungan kepercayaan diri dengan konformitas pada remaja. *Jurnal Ecopsy*, *3*(2).

Dian Ratih Utama Sari, Sri Wahyuni, R. W. B. (2018). *Pengembangan Instrumen Tes Multiple Choice High Order Thinking Pada Pembelajaran Fisika*. 100–107.

Eliza Pradita, Priarti Megawanti, Y. (2023). Analisis Tingkat Kesukaran, Daya Pembeda, dan Fungsi Distraktor PTS Matematika SMPN Jakarta. *HIMPUNAN: Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, *3*(80), 109–118.

Fatwa Tentamaa, Purwatib, Subardjoc, A. S. N. (2023). Validitas dan reliabilitas konstruk skala intensi berwirausaha untuk siswa sekolah menengah kejuruan: Analisis faktor konfirmatori. *Jurnal Psikologi Terapan Dan Pendidikan*, *5*(1), 23–34.

Kang, H., Scharmann, L. C., Kang, H., Scharmann, L. C., Kang, S., & Noh, T. (2010). *DigitalCommons University of Nebraska - Lincoln Cognitive conflict and situational interest as factors influencing conceptual change Cognitive conflict and situational interest as factors influencing conceptual change*.

Keumala, M., Khaldun, I., & Hanum, L. (2019). Analisis Butir Soal Ujian Semester Genap Pelajaran Kimia Kelas X SMK SMTI Banda Aceh Tahun Ajaran 2018 / 2019 Menggunakan Program Proanaltes. *4*(2), 27–32.

Marthunis, M., Khaldun, I., Mia, X., Mia, X., Mia, X., Mia, X., & Mia, X. (2015). Analisis Kualitas Butir Soal Ujian Semester Genap Mata Pelajaran Kimia Kelas X MAN Model Banda Aceh Tahun Pelajaran 2014 / 2015 Menggunakan Program Proanaltes Pendahuluan Latar Belakang Rumusan Masalah Tujuan Penulisan Manfaat Penulisan Membantu Penentuan Kebijakan Sekolah Selanjutnya . *Evaluasi Pendidikan*. *1*(4), 70–78.

Mita Rahmani, Kurnia Ningsih, A. N. (2015). *Jurnal Penelitian dan Evaluasi Pendidikan*. 19–36.

Muh Syahrul Sarea, S. H. (2022). *Analisis Kualitas Soal Ujian Akhir Semester Mata Pelajaran Kimia*. *April*.

Nadia, S. dan L. (2019). Pengembangan Modul Fisika Pada Materi Tekanan di MTsN. *Jurnal Phi*, *5*(April), 74–80.

Pratama, R. (2023). Pengembangan Tes Diagnostik Four-Tier Multiple Choice Pendeteksi Pemahaman Konsep Materi Sistem Koordinasi. *Journal of Education*, *05*(03), 7012–7019.

Pratiwi, H. Y. (2016). Pengembangan Instrumen Tes Pilihan Ganda. 842–850.

Prihono, E. W. (2019). Validitas Instrumen Kompetensi Profesional pada Penilaian Prestasi Kerja Guru Professional Competency Instrument Validity on The Assessment of Teacher Work Performance. *EKSPOSE: Jurnal Penelitian Hukum Dan Pendidikan*, *18*(2), 897–910.

Reni Wahyuningsih, Sri Wahyuni, A. D. L. (2013). *Pengembangan Instrumen Self Assessment Berbasis Web*. 338–343.

Sulistiyono, Mundilarto, H. K. (2019). Pengembangan Instrumen Penilaian Kerja Laboratorium Fisika untuk Mengukur Sikap dan Tanggung Jawab Siswa. *9*, 43–49.

Wibawa, E. A. (2019). *Karakteristik Butir Soal Tes Ujian Akhir Semester*. *XVII*(1).

Yani, M. (2018). *Efektivitas Distractor Pada Tes Pilihan Ganda Untuk Mendeteksi Kesalahan Siswa Dalam Menyelesaikan Soal Matematika*. *2*(2), 125–138.

Hobri, H. (2010). Metodologi Penelitian Pengembangan (Aplikasi Pada Penelitian Pendidikan Matematika). *Jember: Pena Salsabila*.

Sidabutar, G. D. U., Putrayasa, I. B., & Martha, I. N. (2017). Kualitas Butir Soal Ulangan Akhir Semester Ganjil Bahasa Indonesia Kelas IX SMP Negeri 2 Singaraja Tahun Pelajaran 2016/2017 Ditinjau Dari Segi Taraf Kesukaran, Daya Beda, dan Fungsi Pengecoh. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia Undiksha*, *7*(2).