https://doi.org/10.24036/jep/vol8-iss1/875

JEP (Jurnal Eksakta Pendidikan)
Volume 8, Issue 1, 62 - 72
ISSN: 2579-860X (Online), ISSN: 2514-1221 (Print)
https://jep.ppj.unp.ac.id/index.php/jep

Development of Mersics.com as an E-diagnostic Assessment with Rasch Model in Prerequisite Material for the Kinematics Chapter

Agun Gunawan[1*], Yudi Guntara[2], Asep Saefullah[3]
[1,2,3] Department of Physics Education, Sultan Ageng Tirtayasa University, Banten, Indonesia.

Received: 07 March 2024 ▪ Revised: 18 March 2024 ▪ Accepted: 25 May 2024

ABSTRACT

E-diagnostic Assessment is a diagnostic test instrument based on the Mersics.com website, which is used to determine students' initial level of understanding before carrying out physics learning. This research aims to determine the validity and reliability of the development of Mersics.com as an e-diagnostic assessment of the prerequisite material for the kinematics chapter. The limited trial subjects in this research were 75 class XI MIA students. Validity analysis uses the Item Response Theory (IRT) approach with one Logistic Parameter (PL), the difficulty level parameter, better known as the Rasch Model. The research results showed that 56 questions were declared valid by experts using Content Validity Ratio (CVR > 0) analysis. Meanwhile, the results of the analysis based on the Rasch Model showed 20 valid questions in the Understanding of Motion sub-chapter, nine valid questions and one invalid question item in the Motion with Constant Velocity and Motion with Constant Acceleration sub-chapters, ten valid questions in the Parabolic Motion sub-chapter, nine valid questions and one the question item is invalid and in the Circular Motion sub-chapter. The reliability of the questions as a whole is declared reliable. It can be concluded that E-diagnostic Assessment is suitable for use in physics learning.

Keyword: E-diagnostic Assessment, Kinematics, Mersics.com, Rasch Model

INTRODUCTION

Learning is a interaction process between students, teachers, and learning resources in a learning environment. Like other structured activities, the main goal of learning is to create an effective teaching process so that students can achieve optimal learning outcomes. Learning that can help students make the most development possible based on their comprehension is considered effective learning (Nurhayati et al., 2019). In this regard, the Ministry of Education and Culture has made rapid curriculum changes, which resulted in the Independent Curriculum. The main focus of the Merdeka Curriculum is to improve students' ability to understand concepts and provide flexibility for students to choose various learning aids that suit students' learning needs and interests (Fauzan et al., 2023).

The implementation of the Independent Curriculum follows a TARL (Teaching at the Right Level) methodology. The TaRL approach is a method of instruction that centers on

***Correspondence:**
Agun Gunawan, Department of Physics Education, Sultan Ageng Tirtayasa University, Banten, Indonesia.
✉ email : agungunawan413@gmail.com

students' aptitude levels (Dewi, 2022). TaRL is an approach to learning that prioritizes the ability level of the student over grade level (Mubarokah, 2022). This method acknowledges that students cannot be treated equally since they are all unique individuals (Wahyudi et al., 2023). This method acknowledges that no two students are the same and that they cannot all be treated equally. Before the Merdeka Curriculum and the TaRL approach were implemented, students and teachers carried out learning in the classroom by setting a target time to complete all material content, without paying attention to the students' initial level of understanding of the material (Mabsutsah et al., 2022). As a result, there are a large number of students who do not fully understand one material but they still have to move on to the next material.

based on the findings of a class interview In this instance, the sorting of students is based on their learning styles, with the assumption that all students have the same level of ability. This will result in a decrease in students' interest in learning and students' lack of understanding of learning material because students have not learned according to their level of understanding. Especially among class To implement the TaRL Approach, teachers must be able to recognize student interests and learning outcomes with the help of diagnostic assessments. Diagnostic assessments will help measure students' initial understanding of the physics prerequisite material they have studied (Rokhman et al., 2020).

Diagnostic assessments can be carried out via conventional paper-based methods or modern methods involving digital applications or websites. Using website-based diagnostic assessments can speed up the evaluation process and provide time efficiency, especially in correcting questions and scoring. However, ensuring that the diagnostic test instruments used are valid and reliable is also important (Maulana et al., 2023). To achieve this, analyzing the prerequisite material for the Kinematics Chapter, such as using the Rasch Model, is a very useful step. The Rasch Model can be used to gauge each question item's quality attributes, such as the questions' degree of difficulty. By analyzing the items using this method, you can identify which items need to be adjusted or improved to ensure that the diagnostic assessment provides accurate and relevant results. E-diagnostic Assessment can identify students' abilities. Identifying misconceptions is very beneficial for subject teachers because teachers can easily take steps to help students who experience misconceptions (Rohmah et al., 2019). implementing diagnostic tests at the beginning of learning can make it easier for teachers to identify students' weaknesses and strengths (Jannah, 2020).

Based on the explanation above, in this research a more valid and reliable digital diagnostic assessment will be developed using a website called Mersics.com to save time, especially in correcting questions and scoring. The multiple-choice diagnostic assessment tool was created using the logarithmic function to produce measurements with an interval scale. Additional question item analysis using the Rasch Model analysis determines the probabilistic relationship between the respondent's (student's) ability and the difficulty level of the question items. the same as well as to examine the degree of validity and reliability of the test questions in order to produce a useful diagnostic assessment question instrument.

METHOD

Research Design

By modifying the Borg & Gall development paradigm, this study is Research and Development (R&D) (Borg & Gall, 1989).The development steps were modified into seven

out of 10 steps according to research needs. The development steps are as follows: Initial stage, data collection, and literature study. At this stage, an interview was conducted with the class XI physics teacher at one of the senior high schools in Tangerang Regency. Then, review journals, books, and various references related to the Website-based E-diagnostic Assessment With the Rasch Model. Product design and description comprise the second phase. Right now, the output is the creation of a digital diagnostic assessment tool in the form of a grid of questions on the physics prerequisite content for the High School Kinematics Chapter. This tool is being developed using learning outcomes as question indicators, and it will be made available through the Mersics.com website. The third stage is the validation of products by experts. At this stage, two expert lecturers and one physics teacher validated the question items to assess and provide input regarding the suitability between the question indicators and the question items created and evaluating the product before limited trials. The results of the content validity analysis are seen from the scores obtained from the validity assessment by experts, which are calculated using the following equation:

$$CVR = \frac{\left(n_e - \frac{N}{2} \right)}{\frac{N}{2}}$$

Where CVR is the Content Validity Ratio, ne is the number of experts who provided essential responses, and N is the total number of experts. The reference for the calculation results is CVR > 0, then the question item is said to be valid (Widhiarso, 2010).

The fourth stage is revision stage I. At this stage, improvements are made regarding the content of the questions and the suitability of the question indicators based on expert validity before testing in the field. The fifth stage is limited trial. At this stage, approximately 100 students were involved as respondents. The limited trial analysis or empirical validity results are seen using the ministep program. The terms of the item fit order can determine the validity value. In the ministep program, a question item is said to fit if it meets the criteria, namely the MNSQ value is in the range of 0.5 < MNSQ < 1.5, the ZSTD value is in the range -2.0 < ZSTD < +2.0 and the Point Measure Correlation value is in the range of 0.4 < Pt Measure Corr < 0.85. If one question item is found where the MNSQ and Point Measure Correlation values do not meet the criteria, but the ZSTD value does, then the question item is still considered fit (Boone et al., 2014). The results of the reliability analysis are seen using the ministep program. The reliability value of the ministep program is viewed from the estimated item values and case (respondent) estimated values. Requirements for reliability values between respondents and question items are presented in Table 1.

Table 1. Classification of Reliability Values (Widhiarso, 2010).

Value	Information
<0,67	Weak
0,67-0,80	Enough
0,80-0,90	Good
0,91-0,94	Very Good
>0,94	Special

The ministep application displays the analysis findings for the questions' difficulty level. Each question item's logit value, which is displayed in the measure and standard deviation columns, indicates the degree of difficulty. Table 2 displays the question group categories according to their level of difficulty.

Table 2. Question Group Categories based on Difficulty Level (Sumintono & Widhiarso, 2015).

Logit Value	Category
> + 1SD	Very Difficult
0,0 Logit + 1SD	Difficult
0,0 Logit - 1SD	Medium
< -1SD	Easy

In the sixth stage, stage II revision. At this stage, questions are improved based on the results of the limited trial analysis. The final stage is dissemination and implementation. The test instrument developed is an e-diagnostic assessment using the Rasch Model on the prerequisite material for the kinematics chapter in multiple choice form and packaged using the mersics.com website.

RESULTS AND DISCUSSION

Results

The development of an E-diagnostic Assessment on the prerequisite material for the kinematics chapter obtained several results, including:

Content Validity

Based on the analysis of validation sheets by expert lecturers and teachers, it was concluded that several items on the prerequisite material for the Kinematics chapter that had been developed were valid for use. This is proven by the results of the assessment calculation using the CVR equation, which is more than zero (CVR > 0). So, 56 questions are valid for use in limited trials.

Empirical Validity and Reliability

Validity and reliability analysis of the results of this empirical trial used the Rasch model application assisted by the Ministep program. The results of this analysis are presented in Table 3.

Table 3. Overall Validation and Reliability Analysis Results

Sub Chapter	Amount		Item Reliability	Amount	
	Items	Valid Items (Rasch Model)		Item repaired	Item discarded
Understanding Movement	20 Items	20 Items	Very Good (Reliable)	-	-
Motion with Constant Velocity	10 Items	9 Items	Special (Reliable)	-	1 Items

and Motion with Constant Acceleration					
Parabolic Motion	10 Items	10 Items	Good (Reliable)	-	-
Circular Motion	10 Items	9 Items	Very Good (Reliabel)	-	1 Items

As for reliability, it can be seen in the value of item reliability and person reliability. The following are the reliability results using the Rasch model assisted by the Ministep program, which can be seen in Table 4.

Table 4. Results of Item Reliability and Person Reliability Analysis

Sub Chapter	Person Reliability	Item Reliability
Understanding Movement	0,62 (Weak)	0,94 (Very Good)
Motion with Constant Velocity and Motion with Constant Acceleration	0,23 (Weak)	0,95 (Special)
Parabolic Motion	0,02 (Weak)	0,88 (Good)
Circular Motion	0,23 (Weak)	0,92 (Very Good)

Based on the results of the analysis of questions using the Rasch Model with the help of the ministep program as presented in table 3 and 4, the results showed that there were two questions that were invalid, namely 1 question item (S10) in sub-chapter Motion with Constant Velocity and Motion with Constant Acceleration and 1 question item (S8) in sub-section circular motion chapter. The question items were declared invalid because they did not meet the three criteria the Rasch model determined: Outfit Means Square (MNSQ), Outfit Z-standard (ZSTD), and Point Measure Correlation. So, this question will be discarded. Meanwhile, in the sub-chapter on understanding motion and parabolic motion, all questions are declared valid because they meet the Rasch model criteria. As for the overall reliability of the items, each sub-chapter is declared reliable according to the criteria specified in Table 1.

Difficulty Level of Question Items

Each question item's logit item measure value, which has been ranked from highest to lowest based on the item logit value displayed in the jmle measure column, indicates the difficulty level of the question items. A high logit number denotes a challenging level of the questions (Sumintono & Widhiarso, 2015). The complete distribution data on the difficulty level of the questions is presented in Table 5.

Table 5. Results of Grouping Analysis of Question Item Difficulty Levels

Sub Chapter	Category			
	Very Difficult	Difficult	Medium	Easy
Understanding Movement	15, 20	3, 8, 10, 13, 16	2, 4, 5, 6, 9, 11, 12, 14, 17, 18, 19	1, 7
Motion with Constant Velocity and Motion with Constant Acceleration	2, 10	3, 7	5, 6, 8, 9	1, 4
Parabolic Motion	1	3, 4, 9, 10	2, 5, 6, 8	7
Circular Motion	3, 8	4, 6, 7, 10	9	1, 2, 5

Discussion

The final product developed in this research is the E-diagnostic Assessment with the Rasch Model on the prerequisite material for the kinematics chapter, which is packaged using the Mersics.com website which has been developed following the steps of the Borg and Gell development model. This final product has been adjusted to the learning outcomes and adapted to the indicators of physics subject questions, especially the prerequisite material for the kinematics chapter that has previously been studied. Thus, the instruments that have been created can be used to facilitate teachers in diagnosing and finding out students' initial understanding. So that in its implementation, teachers can more easily group students into mastered and not yet mastered categories. In the development of this instrument, it has actually gone through three revisions. The first revision was carried out based on input provided by the supervisor. The second revision was implemented based on commenters, suggestions and validator corrections. The third revision was carried out based on empirical tests on the results of limited trials.

The E-diagnostic Assessment's empirical validity and reliability analysis is predicated on the outcomes of a restricted trial conducted through the mersics.com website, enabling respondents to access it online. Analysis of limited trial results was carried out using Rasch modeling which involved one logistic parameter (1PL) or also called the Rasch Model, namely the level of difficulty with the help of the ministep program. The weakness of classical tests is what led to the birth of item response theory. Classical theory pays less attention to how respondents respond to an item. The ability of the answerer is assessed based on the total score of the answerer's correct answers, regardless of whether the questions answered correctly are easy or difficult (Nurcahyono, 2016). Classical theory only emphasizes the real score (T) which is considered a person's ability. In the meantime, item suitability analysis is produced through analysis using the Rasch model item response theory approach, which indicates whether or not the final data best shows that people typically have a high ability to provide feedback models for elements in accordance with their degree of ability (Misbach & Sumintono , 2014). According to the Rasch model, a person's ability is determined by the estimated level of difficulty of the questions and the respondent's answers influence the questions' level of difficulty (Sumintono & Widhiarso, 2015).

Based on the results of the empirical validation stage on all the question items in each sub-chapter of the prerequisite material for the kinematics chapter, several items were found that did not fit (misfit). In the Motion with Constant Velocity and Motion with Constant

Acceleration sub-chapters, one question item was found that was not fit (invalid), namely in question item number 10, the MNSQ outfit value (1.83) was not accepted, the ZSTD outfit value (2.76) was not accepted and the Pt Measure Corr value (-0.03) is not accepted. In the circular motion sub-chapter, one question item was found that was not fit (not valid), namely in question item number 8, the MNSQ outfit value (1.84) was not accepted, the ZSTD outfit value (2.53) was not accepted and the Pt Measure Corr value (0.15) not accepted. Question items that are declared unfit (invalid) because they do not meet the three criteria determined by the Rasch Model, then those question items must be discarded. As for the sub-chapter on understanding motion and the sub-chapter on parabolic motion, all the question items only do not meet one criterion, so the final conclusion is that there are no questions that need to be changed or replaced, meaning that the question items are valid.

Another factor that causes items to be unfit is the student's answer pattern which does not match the probability of ability with the level of difficulty of the item. Many students think that assessment does not affect achievement in class, causing many students to guess (lucky guess) and copy each other in answering questions. This is in accordance with the theory which explains that the existence of invalid questions and low reliability can be caused by students' high guessing ability in answering the questions given (Aida et al., 2017). Apart from that, it is influenced by the form of the questions which are made in multiple choice form. The form of questions where the answers are readily available tends to reveal only memory and recognition power, is difficult to measure higher mental processes, provides opportunities to play luck or guess, and opportunities for cooperation become more open (Purnomo, 2015).

For reliability calculations using the Rasch model application assisted by the Ministep program, you can see the Output Sumarry Statistics. This output can measure Item Reliability and Person Reliability. Based on this table 4, it shows that the reliability of the question instrument is always superior to the reliability of respondents. This is in accordance with previous research conducted by Angraeni et al., (2020) declaring that the question instrument's dependability is suitable for usage on the same subject by various people at different times and locations. Put differently, the Rasch Model maintains that the quality of the question instrument is independent of the respondent's aptitude. Even if the students' talents are low, the question instruments' quality will not degrade.

Then, for the level of difficulty of the questions, the results showed that in the sub-chapter Understanding Motion, items S15 and S20 were items that were very difficult for students to complete because they had very high logit values. Meanwhile, items S7 and S1 are questions with the lowest level of difficulty because they have very small logit values and there were 60 students who answered correctly. The same logit value for each question indicates a level of difficulty that is not much different. Likewise for other sub-chapters.

CONCLUSION

Based on the results of limited trial analysis using Rasch Model Analysis assisted by the Ministep Program which shows 20 valid questions in the Understanding of Movement sub-chapter, nine valid questions and one invalid question in the Motion with Constant Velocity and Motion with Constant Acceleration sub-chapters, ten valid questions in the Motion sub-chapter Parabola, nine valid questions and one invalid question in the Circular Motion sub-chapter. Items that are declared invalid are items whose outfit means-square (MNSQ), outfit z-standard (ZSTD) and point measure correlation (Pt Measure Corr) values are not fit or do

not meet the criteria set in Rasch modeling and Invalid question items will be discarded. Then, the overall reliability results show that the reliability of the items in each sub-chapter is categorized as good and the questions are suitable for use on the same subject even at different times and places and by different people. Because the quality of the question instrument will remain good even if the student's abilities are low. So, overall it can be concluded that the E-diagnostic Assessment with Rasch Model in the Prerequisite Material for the Kinematics Chapter which was developed through the Mersics.com website is suitable for use in physics learning in schools. And can help teachers diagnose or identify students' initial level of understanding before studying the Kinematics Chapter.

REFERENCES

Adams, R. J.and Khoo, S-T. (1996). ACER Quest: the interactive test analysis system. [Computer software]. Version 2.1. Camberwell, Victoria: Australian Council for Educational Research. https://research.acer.edu.au/measurement/3/.

Aida, N., Kusaeri, K., & Hamdani, S. (2017). Karakteristik Instrumen Penilaian Hasil Belajar Matematika Ranah Kognitif yang Dikembangkan Mengacu pada Model PISA [Characteristics of the Cognitive Domain Mathematics Learning Outcome Assessment Instrument Developed Referring to the PISA Model]. *Suska Journal of Mathematics Education*, *3*(2), 130. https://doi.org/10.24014/sjme.v3i2.3897.

Angraeni, D. N., Suherman, A., & Guntara, Y. (2020). Aplikasi Rasch Model: Pengembangan Fluids Assessment (Fass) Berdasarkan Taxonomy of Introductory Physics Problems (TIPP) [Rasch Model Application: Development of Fluids Assessment (Fass) Based on Taxonomy of Introductory Physics Problems (TIPP)]. *Jurnal Penelitian Pembelajaran Fisika*, *11*(2), 135–143. https://doi.org/10.26877/jp2f.v11i2.5903.

Bond T.G. dan Fox C.M. (2007). Applying the Rasch Model Fundamental Measurement in the Human Science. 2nd Ed, New Jersey: Lawrence Erbaum Associates, 11-14. https://doi.org/10.4324/9781410614575.

Boone, W.J., Staver, J.R., & Yale, M.S. (2014). Rasch Analysis in the Human Sciences. Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-007-6857-4.

Borg, W. R., & Gall, M. D. (1989). *Educational Research: An Introduction, Fifthy Edition*. New York: Longman. https://books.google.co.id/books/about/Educational_Research.html?.

Dewi Cahyono, S. (2022). Melalui Model Teaching at Right Level (TARL) Metode Pemberian Tugas untuk Meningkatkan Motivasi dan Hasil Belajar Peserta Didik [Through the Teaching at Right Level (TARL) Model, Assignment Method to Increase Student Motivation and Learning Outcomes]. *Jurnal Pendidikan Tambusai*, *6*(2), 12407–12418. https://doi.org/10.31004/jptam.v6i2.4431.

Fauzan, F., Ansori, R. A. M., Dannur, M., Pratama, A., & Hairit, A. (2023). The Implementation of the Merdeka Curriculum (Independent Curriculum) in Strengthening Students' Character in Indonesia. *Aqlamuna: Journal of Educational Studies*, *1*(1), 136–155. https://doi.org/10.58223/aqlamuna.v1i1.237

Fitriatun, A., & Sukanti. (2016). Analisis Validitas, Reliabilitas, Dan Butir Soal Latihan Ujian Nasional Ekonomi Akuntansi Di MAN Maguwaharjo [Analysis of Validity, Reliability and Practice Questions for the National Accounting Economics Examination at MAN Maguwaharjo]. *Jurnal Kajian Pendidikan Akuntansi Indonesia*, (3), 1–11. https://journal.student.uny.ac.id/index.php/kpai/article/viewFile/5801/5554.

Jannah, R. & I. R. (2020). Pengembangan E-Diagnostic Four Tier Test Untuk Mengidentifikasi

Miskonsepsi Peserta Didik [Development of E-Diagnostic Four Tier Test to Identify Student Misconceptions]. *Natural Science: Jurnal Penelitian Bidang IPA Dan Pendidikan IPA*, *6*(2), 151–160. https://doi.org/10.15548/nsc.v6i2.1721.

Julianda, R., Saminan, & Halim, A. (2022). Analisis Miskonsepsi Siswa dengan Two Tier Diagnostic Test pada Materi Gerak Lurus di SMA Negeri 3 Banda Aceh [Analysis of Student Misconceptions with Two Tier Diagnostic Test on Rectilinear Movement Material at SMA Negeri 3 Banda Aceh]. *Journal of Technology and Literacy in Education*, 1(1), 14–20. https://jurnal.serambimekkah.ac.id/index.php/jtle/article/view/9.

Koimah, N., & Muchtar, Z. (2022). Pengembangan tes diagnostik berbasis web pada materi konsep redoks [Development of web-based diagnostic tests on redox concept material]. *Jurnal Ilmiah Pendidikan*, 1(6), 636–644. https://doi.org/10.55904/educenter.v2i1.218.

Kurniawan, U., Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model [Analysis of Multiple Choice Questions with the Rasch Model]. *Jurnal Statistika*, 6(1), 1-6. https://doi.org/10.26714/jsunimus.6.1.2018.%25p.

Mabsutsah, N., Yushardi, Y. (2022). Analisis Kebutuhan Guru terhadap E Module Berbasis STEAM dan Kurikulum Merdeka pada Materi Pemanasan Global [Analysis of Teacher Needs for STEAM-Based E Modules and Independent Curriculum on Global Warming Material]. *Jurnal Pendidikan MIPA*. 12(2). https://doi.org/10.37630/jpm.v12i2.5.

Maulana, S., Rusilowati, A., Nugroho, S. E., & Susilaningsih, E. (2023). Implementasi Rasch Model dalam Pengembangan Instrumen Tes Diagnostik [Implementation of the Rasch Model in Diagnostic Test Instrument Development]. *Prosiding Seminar Nasional Pascasarjana (PROSNAMPAS)*, *6*(1), 748–757. https://proceeding.unnes.ac.id/index.php/snpasca/article/view/2214

McDermott dkk, L. C. (2017). Development and application of afour-tier test to assess pre-service physics teachers'misconceptions about geometrical optics. Research in Science and Technological Education 35(4):1-23. http://dx.doi.org/10.1080/02635143.2017.1310094z.

Misbach, I. H., & Sumintono, B. (2014). *Pengembangan dan Validasi Instrumen "Persepsi Siswa terhadap Karakter Moral Guru" di Indonesia dengan Model Rach* [Development and Validation of the Instrument "Students' Perceptions of Teacher Moral Character" in Indonesia using the Rach Model]. In Seminar Nasional Psikometri: Pengembangan Instrumen Penliaian Karakter yang Valid. http://repository.upi.edu/59545/1/S_FIS_1406454_Title.pdf.

Mubarokah, S. (2022). Tantangan Implementasi Pendekatan TaRL (Teaching at the Right Level) dalam Literasi Dasar yang Inklusif di Madrasah Ibtida'iyah Lombok Timur [Challenges of Implementing the TaRL (Teaching at the Right Level) Approach in Inclusive Basic Literacy at Madrasah Ibtida'iyah East Lombok]. *BADA'A: Jurnal Ilmiah Pendidikan*, *4*(1), 165–179. https://doi.org/10.37216/badaa.v4i1.582

Nurhayati, N., Alsagaf, S. L. H., & Wahyudi, W. (2019). Pengembangan Tes Diagnostik Three-Tier Multiple Choice Untuk Mengukur Konsepsi Fisika Siswa SMA [Development of a Three-Tier Multiple Choice Diagnostic Test to Measure High School Students' Conceptions of Physics]. *Jurnal Pendidikan*, *4*(2), 47–54. https://doi.org/10.26740/jp.v4n2.p47-54.

Pujayanto, P., Budiharti, R., Adhitama, E., Nuraini, A., & Putri, H. V. (2018). The development of a web- based assessment system to identify students' misconception automatically on linear kinematics with a four-tier instrument test. *Journal Physics Education*, 53(4).1-8.

DOI: 10.1088/1361-6552/aac695.

Purnomo. (2015). *Dasar-Dasar dan Perancangan Evaluasi Pembelajaran* [Basics and Design of Learning Evaluation]. Fakultas Keguruan dan Ilmu Pendidikan: Universitas Lampung.

Purwanto. (2013). *Evaluasi Hasil Belajar* [Evaluation of Learning Outcomes]. Yogyakarta: Pustaka Pelajar. https://inlislite.uin-suska.ac.id/opac/detail-opac?id=21358.

Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen Penilaian Hasil Pembelajaran Kognitif pada Tes Uraian dan Tes Objektif [Instrument for assessing cognitive learning outcomes in description tests and objective tests]. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, 4(2), 139–148. https://doi.org/10.36232/jurnalpendidikandasar.v4i2.2649

Rohmah, Z.-, Handika, J.-, & Huriawati, F.-. (2019). E-Diagnostic Test untuk Mengungkap Miskonsepsi Kinematika [E-Diagnostic Test to Reveal Kinematics Misconceptions]. *SPEKTRA : Jurnal Kajian Pendidikan Sains*, *5*(1), 22. https://doi.org/10.32699/spektra.v5i1.86

Rokhman, O., Ningsih, A. N., Augia, T., Dahlan, H., Rosyada, Amrina, Putri, Dini Arista, Fajar, N. A., Yuniarti, E., Vinnata, N. N., Pujiwidodo, D., Ju, J., Wei, S. J., Savira, F., Suharsono, Y., Aragão, R., Linsi, L., Editor, B., Reeger, U., Sievers, W., Michalopoulou, C., Mimis, A., … Devita, M. (2020). Pengaruh Komunikasi Pemasaran melalui Heuristik dan Sistematik Mode. *Jurnal Berkala Epidemiologi* [The Influence of Marketing Communications through Heuristic and Systematic Modes. Periodical Journal of Epidemiology]. *5*(1), 90–96.https://core.ac.uk/download/pdf/235085111.pdf%0Awebsite:http://www.kemkes.go.id%0Ahttp://www.yankes.kemkes.go.id/assets/downloads/PMK No. 57 Tahun 2013 tentang PTRM.pdf%0Ahttps://www.kemenpppa.go.id/lib/uploads/list/15242-profil-anak-indonesia_-2019.pdf%0Aht

Sugiyono. (2012). *Memahami Penelitian Kualitatif* [Understanding Qualitative Research]. Bandung: ALFABETA. http://repository.iainkudus.ac.id/5258/6/6.%20BAB%20III.pdf.

Suharyani, S., Suarti, N. K. A., & Astuti, F. H. (2023). Implementasi Pendekatan Teaching At The Right Level (Tarl) Dalam Meningkatkan Kemampuan Literasi Numerasi Anak Di SD IT Ash-Shiddiqin [Implementation of the Teaching At The Right Level (Tarl) Approach in Improving Children's Numeracy Literacy Skills at Ash-Shiddiqin IT Elementary School]. *Jurnal Teknologi Pendidikan : Jurnal Penelitian Dan Pengembangan Pembelajaran*, 8(2), 470- 479. https://doi.org/10.33394/jtp.v8i2.7590.

Sumintono, B dan Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan* [Application of Rasch Modeling in Educational Assessment]. Cimahi: Trim Komunikata Publishing House. https://www.researchgate.net/publication/282673464_Aplikasi_Pemodelan_Rasch_pada_Assessment_Pendidikan.

Suseno, E. dan Susongko, P. (2021). *Mengukur Validitas Tes* [Measuring Test Validity]. Edisi ke-1. Pemeral Edukreatif. Jawa Timur. https://books.google.co.id/books/about/MENGUKUR_VALIDITAS_.

Suyoso, S., Istiyono, E., & Subroto, S. (2017). Pengembangan Instrumen Asesmen Pengetahuan Fisika Berbasis Komputer Untuk Meningkatkan Kesiapan Peserta Didik Dalam Menghadapi Ujian Nasional Berbasis Komputer [Development of Computer-Based Physics Knowledge Assessment Instruments to Improve Students' Readiness in Facing Computer-Based National Examinations]. *Jurnal Pendidikan Matematika Dan Sains*, 5(1), 89–97. https://doi.org/10.21831/jpms.v5i1.12461.

Utari, G. P., Liliawati, W., & Utama, J. A. (2021). Design and validation of six-tier astronomy diagnostic test instruments with Rasch Model analysis. Journal of Physics: Conference Series, 1806(1). https://doi.org/10.1088/1742- 6596/1806/1/012028.

Wahyudi, S. A., Siddik, M., & Suhartini, E. (2023). Analisis Pembelajaran IPAS dengan Penerapan Pendekatan Pembelajaran Berdiferensiasi dalam Kurikulum Merdeka [Analysis of Science and Technology Learning with the Application of a Differentiated Learning Approach in the Independent Curriculum]. *Jurnal Pendidikan Mipa*, *13*(4), 1105–1113. https://doi.org/10.37630/jpm.v13i4.1296.

Wahyuni, N., Bhakti, Y. B., Mutakin, T. Z., & Astuti, I. A. D. (2021). The Development of Four-Tier Diagnostic Test Instrument to Identify the Learners' Misconception on Circular Motions. Impulse: *Journal of Research and Innovation in Physics Education*, 1(1), 24–31. https://doi.org/10.14421/impulse.2021.11-03.

Weeden, P., Winter, J., & Broadfoot, P. (2002). Assessment: What's in it for schools? (K. Myers & J. MacBeath (eds.)) Routledge Falmer: London. http://dx.doi.org/10.4324/9780203468920.

Widhiarso, W. (2010). *Prosedur Pengujian Validitas Isi melalui Indeks Rasio Validitas ISi (CVR)* [Content Validity Testing Procedure via the Content Validity Ratio Index (CVR)]. Retrieved February 22, 2024, from http://wahyupsy.blog.ugm.ac.id/2010/06/16/prosedur-pengujian-validitas-isi-melalui-indeks-rasio-validitas-isi-cvr/.