Available online at: http://journal.uny.ac.id/index.php/jipi

Jurnal Inovasi Pendidikan IPA, 11 (2), 2025, 402-417

Diagnosing Students' Problem-Solving Challenges in Rotational Dynamics Using Two-Tier AR Flashcard Tests

Fiqran Isnaini [1] *, Haratua Tiur Maria Silitonga [1] , M. Musa Syarif Hidayatullah [1] Judyanto Sirait [1] , Renol Afrizon [2]

[1] Universitas Tanjungpura. Pontianak, Indonesia.
[2] Universitas Pendidikan Sultan Idris. Perak, Malaysia.
* Corresponding Author. E-mail: f1051211003@student.untan.ac.id

Received: 12 December 2024; Revised: 04 January 2025; Accepted: 11 June 2025

Abstract: The inadequate capacity of students to recognize and resolve difficulties, particularly with intricate concepts, has emerged as a significant concern in physics education. This project aims to develop a two-tier diagnostic assessment that utilizes augmented reality (AR) flashcards to identify students' challenges in problem-solving. This media-assisted assessment aims to visually represent concepts interactively while examining students' comprehension and the origins of their challenges. The study methodology employs the ADDIE development paradigm, encompassing analysis, design, development, implementation, and evaluation. Validation is performed by subject matter and media specialists, subsequently followed by trials with secondary education students. The study equipment comprises validation sheets, answer questionnaires, and diagnostic examinations supplemented by augmented reality flashcards. The study demonstrates that the two-tier diagnostic test augmented with AR flashcards effectively identifies students' challenges, particularly in grasping concepts and problem-solving processes, as indicated by a Cronbach's alpha reliability coefficient of 0.89. This media-assisted assessment enhances student engagement, as evidenced by survey findings indicating that students perceive it as easier to comprehend physics concepts through the instrument. This research is anticipated to motivate educators and institutions to create more creative AR-based learning media for evaluations and interactive learning.

Keywords: augmented reality, diagnostic assessment, problem resolution, rotational dynamics.

How to Cite: Isnaini, F., Silitonga, H. T. M., Hidayatullah, M. M. S., Sirait, J., & Afrizon, R. (2025). Diagnosing students' problem-solving challenges in rotational dynamics using two-tier AR flashcard tests. *Jurnal Inovasi Pendidikan IPA, 11*(2)., 1-xx. doi:https://doi.org/10.21831/jipi.v11i2.80461

INTRODUCTION

The period of globalization has significantly influenced all facets of society, including educational implementation issues. A significant difficulty is that education must cultivate human resources (HR) equipped with 21st-century competencies. Partono et al. (2021) assert that 21st-century competencies in the educational process encompass empathy, critical thinking, problem-solving, creativity, innovation, communication, and collaboration.

Problem-solving skill is a crucial 21st-century capability in science education, especially in physics. Problem-solving may entail the formulation and application of intricate and advanced rules to attain solutions (Gagne & Briggs, 1979). In problem-solving, learners must retrieve subordinate rules and pertinent information that they have previously acquired. Problem-solving cultivates critical, analytical, and creative thinking skills in learners as they confront complex challenges (Anggraeni & Listiaji, 2024; Anton & Trisoni, 2022; Partono et al., 2021).

Many students encounter challenges in comprehending fundamental physics concepts during learning activities, particularly those related to rotational dynamics and the equilibrium of rigid bodies. The erroneous vector representation observed when students address problems in rotational dynamics and inflexible body equilibrium signifies their difficulties in comprehending physics ideas (Syahrial et al., 2022). The issues arise from conceptual errors, insufficient comprehension, or the application of unsuitable problem-solving procedures (Ginting, 2021; Malik et al., 2023; Subekti et al., 2025). Asuri et al. (2021) asserted that students must navigate five stages in problem-solving: visualize the problem, describe the problem in physics description, plan the solution, execute the plan, and check and evaluate.

Students' challenges in problem-solving must be precisely diagnosed to identify the underlying causes, enabling the implementation of suitable interventions to enhance their comprehension and problem-solving abilities. A proficient approach to discerning kids' challenges in problem-solving is through diagnostic assessments. Diagnostic assessments are used by educators to determine pupils' strengths and weaknesses in learning, allowing the results to inform subsequent interventions. Widyaningsih & Aloysius (2023) assert that assessment may consist of a sequence of enquiries or directives to execute specific actions.

Diagnostic exam questions are classified into two categories: short-answer questions (essay/short answer) and multiple-choice questions (multiple choice with justification) (Habellia et al., 2021). This study employs a diagnostic assessment of multiple-choice items with reasoning, structured in a two-tier format. The two-tier format has become a preferred approach because it evaluates students' answers and the logic behind those answers (Mutlu & Sesen, 2015; Widyaningsih & Aloysius, 2023). Habellia et al. (2021) argue that employing a two-tier format can reduce speculation and facilitate the integration of multiple dimensions within a single phenomenon, with the first tier denoting the metacognitive domain and the second tier denoting the conceptual domain.

In the contemporary technological landscape, the utilisation of digital media, including augmented reality (AR) flashcards, might improve the efficacy of learning and evaluation (Arsyad, 2011; Pujiastuti & Haryadi, 2024; Siahaan & Putri, 2024). Augmented reality-based flashcards provide interactive and engaging visualisations, facilitating a more concrete understanding of physics ideas for pupils (Angraini et al., 2023; Fajriansyah & Widodo, 2022; F. Utami et al., 2021). Aminudin et al. (2024) assert that employing augmented reality flashcards enhances the learning experience by making it more engaging and offering visual context that facilitates the identification of students' challenges more efficiently.

Prior research findings suggest that diagnostic assessments primarily aimed at evaluating students' problem-solving skills remain significantly constrained (Rahmawati et al., 2019). Koimah & Muchtar (2023) show that technology-based diagnostic assessments, including web-based tests, enhance the precision of identifying students' misconceptions and facilitate more rapid and focused feedback. The two-tier exam method has demonstrated efficacy in revealing the conceptual rationale underlying students' responses, thereby offering a more thorough understanding of their misconceptions in physics. Nevertheless, these studies typically remain distinct in terms of the advancement of technology media and the dual-tier conceptual diagnostic methodology. To date, few advancements in diagnostic tests have emerged that integratively combine the two-tier framework with augmented reality (AR) technology within the realm of physics education.

This research addresses the theoretical gap by creating a two-tier diagnostic test instrument supported by AR-based flashcards. The objective is to more precisely discern students' challenges in resolving physics problems, while simultaneously providing a contextual visual experience to reconstruct their comprehension. This research expands upon prior findings while also differentiating itself through a cohesive and novel methodology that amalgamates the advantages of diagnostic assessment with immersive technology.

Based on observations in several high schools, it has been revealed that students often face challenges in problem-solving related to the dynamics of rotation and the equilibrium of rigid bodies. Additionally, educators emphasize that students need diagnostic evaluation instruments that can enhance engagement in the learning process. In response to this issue, the researchers designed a two-tier diagnostic test equipped with augmented reality flashcards to identify the challenges faced by students in problem-solving, particularly related to the dynamics of rotation and the equilibrium of rigid bodies.

This work presents a novel contribution by constructing a two-tier diagnostic test instrument using flashcards enhanced by augmented reality (AR), an area that has been underexplored in the context of rotational dynamics and rigid body equilibrium. This work theoretically enhances physics learning assessment research by utilizing AR technology to detect students' conceptual and procedural issues more thoroughly. The resultant product is anticipated to serve as an engaging and effective evaluation instrument for educators in the physics learning process across many curricula.

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

METHOD

This research employs the research and development (R&D) methodology utilizing the ADDIE development model, which is adapted to meet the research requirements across five primary stages: analysis, design, development, implementation, and evaluation. Sampling was executed using purposive sampling, focusing on schools that have adopted the Merdeka Curriculum and possess facilities conducive to the utilization of technology-based media, including mobile phones and laptops. The study involved 178 students from grades X and XI who had engaged with the content on rotational dynamics and the equilibrium of rigid bodies. Before initiating the investigation, the researcher secured authorization from the school authorities and got consent from the physics instructors at four institutions: two high schools in Kubu Raya, one high school in Bengkayang, and one high school in Pontianak.

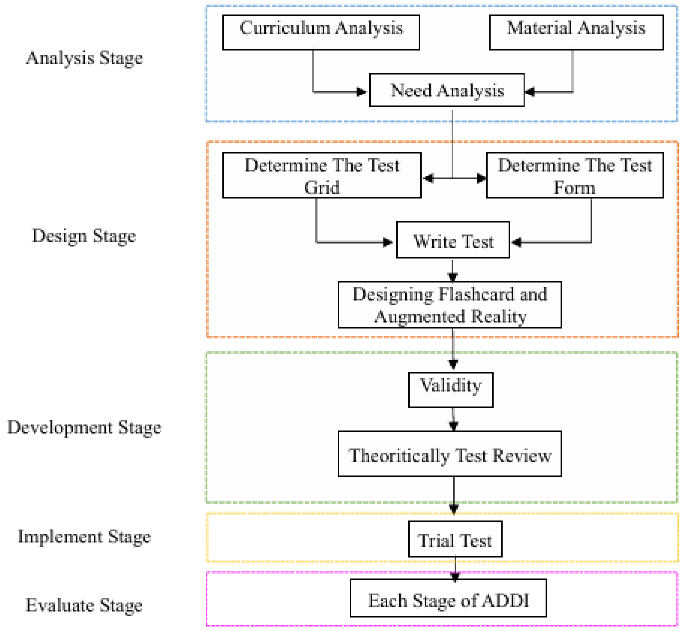


Figure 1. Development procedure

Figure 1 outlines the procedures for constructing a second-level test utilizing augmented reality flashcards based on the ADDIE stages of research development method. The analysis stage is initiated with the administration of a needs analysis questionnaire. The needs analysis questionnaire encompasses curriculum analysis, material analysis, and needs assessment. This investigation aims to assess the significance and immediacy of students' requirements about the curriculum, resources, augmented reality technology, and problem-solving.

The Design stage involves formulating questions derived from the learning outcomes of rotational dynamics and rigid body equilibrium, in conjunction with the educator's established learning objectives. Consequently, 10 questions were formulated as essays or short-answer inquiries requiring open-ended reasoning. The questions encompass cognitive elements aligned with the problem-solving steps delineated by Heller & Heller: (1) identify the problem, (2) articulate the physics, (3) formulate the solution, (4) implement the plan, and (5) assess the answer (Silitonga et al., 2024).

Ten items were formulated into a two-tier diagnostic assessment based on these cognitive features. Tier 1: multiple-choice inquiries about the students' conclusive responses. Tier 2: multiple-choice inquiries aimed at elucidating the rationale behind students' selections in Tier 1. Each question is designed to detect students' challenges in addressing problem-based physics problems, followed by answer keys, grading criteria, and guidance for interpreting diagnostic results to classify the types of errors or misconceptions that occur.

Furthermore, augmented reality (AR)-enhanced flashcard media were created utilizing the Paint 3D application, complemented by additional tools such as Asemblr EDU and Canva, which present interactive visual models for each subject. This graphic aims to help students understand the physical scenarios presented, including force distribution, pivot locations, and torque direction. To verify the

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

instrument's viability, expert validation questionnaires (content, media, and evaluation validators) and student response questionnaires were created, employing a five-point Likert scale to assess content, construct, language, and appearance. The outcomes of this questionnaire are utilized to amend and enhance the instrument before the field trial phase.

The Development stage consists of two phases: the validity test phase and the small group trial. The validity assessment is performed by five professional evaluators. The expert validators comprise media specialists, linguistic authorities, and subject area authorities. Expert validators evaluate the two-tier diagnostic test instrument, enhanced with augmented reality flashcards, to acquire evaluations and feedback. Following the acquisition of assessment and feedback from the validators, a data analysis was performed to examine the feasibility and validity of the constructed two-tier diagnostic test instrument enhanced with augmented reality flashcards. The evaluation was computed with the Aiken index formula (Sirait et al., 2023).

$$V = \frac{\sum S}{[n(C-1)]}$$

The Aiken index is preferable for an instrument feasibility study due to its simplicity, emphasis on content validity, suitability for ordinal data, and efficacy with a limited number of validators (Romero-Saldaña et al., 2024). The Aiken index readings are classified into three categories: low (0.00–0.40), moderate (0.41–0.80), and high (0.81–1.00) (L. Utami et al., 2024). Should the validation results indicate that the instrument satisfies the feasibility criteria, it would then be administered to students to ascertain challenges in problem-solving related to rotational dynamics and the equilibrium of rigid bodies. A small group trial was then conducted to determine the completion time and student responses to the use of this AR-assisted diagnostic test.

The Implementation phase involved testing the product on an extensive cohort. The trial was executed in person with 178 pupils across four distinct schools. Students were allotted 90 minutes to finalize the tasks. The data collected at this step were analyzed to ascertain the difficulty level, dependability, and discriminative power of the test items. The analysis of the gathered respondent data was performed using the Winsteps Rasch model software to ascertain the difficulty level of the test items, the validity of the test items, and the reliability of the test instrument employed. The Rasch model was employed because of its superior accuracy, comprehensive examination of student responses, and adaptability in effectively identifying conceptual challenges (Muliadi et al., 2025).

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

Table 1. Eligibility Aspect Rasch Model

Eligibility Aspects from the Rasch Model	Description
Unidimensionality	Percentage of "raw variance explained by measure" > 20%
Person dan Item Reliability	Special : > 0.94 Very Good : 0.91 – 0.94 Good : 0.81 – 0.90 Pretty Good : 0.67 – 0.80 Weak : < 0.67
Wright Map	Describe the distribution of persons and items over the map.
Item FIT	0.5 > Outfit MNSQ > 1,5 -2 > Outfit ZSTD > 2 0.4 > PT Measure Corr > 0.85
Item-Measure	Very Easy : b <-2,00 Easy : -2.00 < b < -1.00 Pretty Difficult : -1.00 < b < +1,00 Difficult : + 1.00 < b < +2.00 Very Difficult : b > +2.00

(Rahmawati et al., 2019; Sari & Mahmudi, 2024)

The Evaluation phase occurred at every stage of ADDI. The evaluation during the analysis phase was based on students' feedback concerning their comprehension of the implemented curriculum, challenging materials, the integration of augmented reality technology, and the resolution of Heller stage issues. The evaluation during the design phase was based on feedback and recommendations from validators concerning the design of learning indicators and the appropriateness of items in the flashcards and augmented reality. Assessment during the development phase was carried out based on the evaluations and recommendations of the validators. Evaluation during the implementation phase was performed using the outcomes of extensive group trials to assess the difficulty level, reliability, and discriminative power of the questions.

RESULT AND DISCUSSION

Analysis Stage

The phases conducted at this juncture encompass curriculum analysis, material analysis, and needs analysis. The study at this level involved administering questionnaires to 54 students. This questionnaire solicits information and perceptions from students concerning the implemented curriculum, challenging physics concepts, augmented reality-based flashcard learning tools, and problem-solving skills in accordance with Heller's phases.

According to the needs analysis conducted via Google Forms, the curriculum analysis depicted in Figure 2 reveals that 55.6% (30 students) indicated the adoption of the independent curriculum in learning activities, 40.7% (22 students) stated the adoption of Curriculum 2013, while 1.85% (one student) indicated the adoption of the emergency curriculum and 1.85% (one student) noted the adoption of the competency-based curriculum.

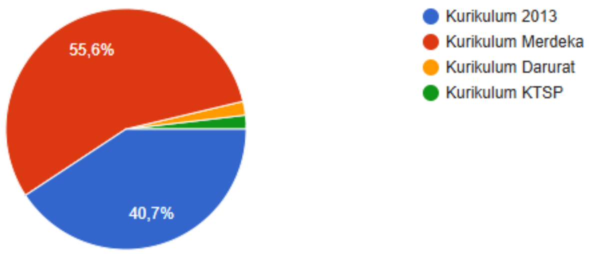


Figure 2. Curriculum Analysis Results

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

Figure 3 presents the findings of the material analysis: 11.1% (6 students) reported that the rotational dynamics material remained very difficult, 74.1% (40 students) indicated that it was somewhat tricky, and 14.8% (8 students) asserted that it was not difficult. This aligns with a study by Syarial et al. (2024), which indicates that rotational dynamics is a challenging physics topic, leading to student misconceptions.

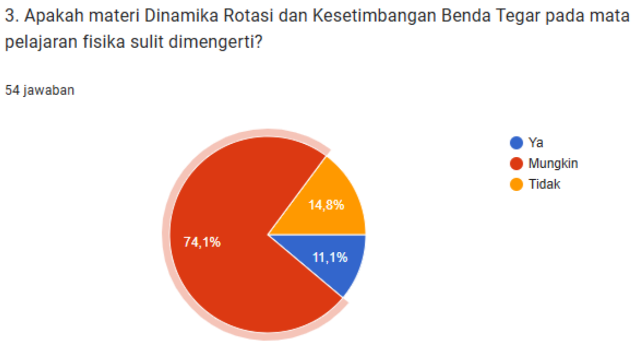


Figure 3. Results of the material difficulty analysis

An analysis of augmented reality-based flashcard media revealed that 64.8% (35 students) were unfamiliar with it, while 35.2% (19 students) were already acquainted with the concept. This aligns with a study by Elvina et al. (2024), which indicates that augmented reality flashcard media is rarely utilized in the educational process, particularly when introducing topics that are challenging to convey in tangible terms within the classroom.

Figure 4. Results of the media needs analysis

Figure 5 presents an analysis of problem-solving abilities based on Heller's stages: 37 % (20 students) reported utilizing the visualization stage, 50 % (27 students) indicated the application of the problem description stage, 33.3 % (18 students) acknowledged employing the solution planning stage, 31.5 % (17 students) confirmed the use of the solution implementation stage, and 9.3% (5 students) engaged in the evaluation stage. The findings of this analysis align with those of Malik et al. (2023), which indicate that students' problem-solving skills in physics are categorized as moderate and deficient when assessed using the physics problem-solving ability indicator.

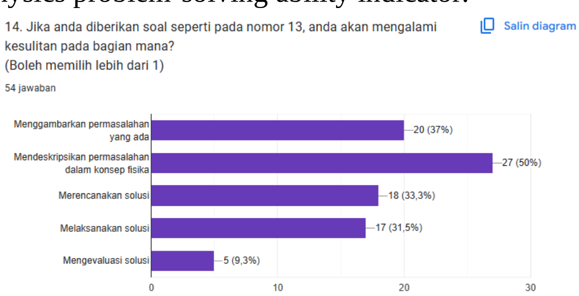


Figure 5. Result of problem solving difficulty analysis

The Google survey's findings are consistent with the findings of field observations and interviews. The concepts and issues in the questions are not fully comprehended by the students, who spend a significant amount of time studying the material on rotational dynamics and are preoccupied with mathematical computations. This is because their learning resources solely provide mathematical computations and theoretical issues (Malik et al., 2023). As a result, students are eager to use flashcards based on augmented reality to solve challenges about tangible learning exercises.

Design Stage

The second phase of this study is the design phase. The concept, content, and materials of the product to be generated are designed during the design phase in accordance with the findings of the curriculum and needs analysis. The design is as follows.

1. Creating test tools

The assessment tool is developed using Microsoft Word and comprises a two-tier format with various components, including topic, problem indicators, problem-solving stages, solutions, and justifications for answer selection.

Table 2. Stages, Topics, and Indicators of the Items

Problem-Solving Stage	Topic	Indicator	Item
Visualization	1. The Force Moment 2. Momentum in angle 3. Momentum in angle 4. Balance 5. Moment of Inertia 6. Equilibrium	Students can visualize physics problems in each topic of the questions.	1,2,9,10.17,18,25, 26,33,34,41,42
Description	1. The Force Moment 2. Momentum in angle 3. Momentum in angle 4. Balance 5. Moment of Inertia 6. Equilibrium	Students can describe the physics problems in each topic of the questions.	3,4,11,12,19,20.27, 28,35,36,43,44
Planning	1. The Force Moment 2. Momentum in angle 3. Momentum in angle 4. Balance 5. Moment of Inertia 6. Equilibrium	Students can plan solutions to physics problems on each topic.	5,6,13,14,21,22,29, 30. 37,38,45,46
Implementation	1. The Force Moment 2. Momentum in angle 3. Momentum in angle 4. Balance 5. Moment of Inertia 6. Equilibrium	Students can implement solutions to physics problems on every topic	7,8,15,16,23,24,31, 32,39,40.47,48

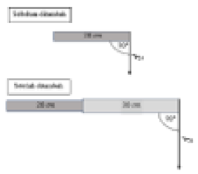
ISSN 2406-9205 (print), ISSN 2477-4820 (online)

First Question

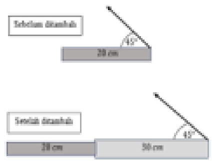
Cab driver encountered a flat tire while returning to the base. The motorist thereafter sought the nearest workshop to replace the flat tire on his vehicle Upon arrival at the workshop, the mechanic prepared two primary tools to loosen the nuts on the flat tire. The instruments employed consisted of a 2 cm wrench and a 30 cm wrench extender. The mechanic exerted a force of 10 N orthogonal to the wrench.

1.1. Select the suitable graphic depicting the technician loosening the bolt on the flat tire with a wrench, both prior to and subsequent to the integratio of the wrench with a connecting rod!

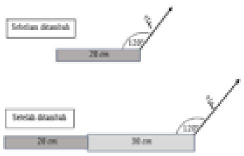
a.



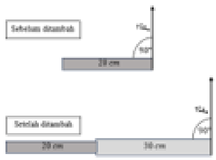
b.



c.



d.



Reason for choosing the answer ....

a. If the force applied to the shaft is increasingly tilted towards the northwest, then the torque produced i s maximized for loosening the bol
b. If the force applied to the shaft tilts more towards the northeast, the torque produced will be maximal for loosening the bolt.
c. If the force applied to the shaft tilts more towards the north, then the torque generated is maximal for loosening the bolt.
d. If the force applied to the shaft tilts further towards the south, the torque produced will be maximized for loosening the bolt.

1.2. Pay attention to the following statement!
1. The combined length of the wrench and iron is 50 cm. The combined length of the wrench and iron is 50 cm.
2. The magnitude of the force applied by the mechanic is 20 N.
3. the length of the wrench is 20 cm
4. The magnitude of the force applied by the mechanic is 10 N. The magnitude of the force applied by the mechanic is 10 N.

The correct statement based on the above narrative is ...
a. 1 and 2
b. 2 and 3
c. 2 and 4
d. 1 and 4

The reason for choosing the answer ...
a. Because the length of the key and iron joint is 50 cm, while the total force applied to the key joint is 20 N.
b. Because the length of the key and iron joint is 50 cm, while the force applied to both the key and the key and iron joint is 10 N
c. Because the total force applied to the key joint is 20 N while the force applied to the key is 10 N
d. Because the total force applied to the key combination is 20 N while the length of the key is 20 cm

1.3. How to determine the comparison of torque before and after the wrench is combined with the connecting iron
a. Determine the torque produced by the mechanic before and after the wrench is combined with the connecting iron using the equation $\tau = r + \vec{F}$, then compare the two produced torques.
b. Determine the torque produced by the mechanic before and after combining the wrench with the connecting iron using the equation $\tau = r - \vec{F}$, then compare the two produced torques.
c. Determine the torque produced by the mechanic before and after combining the wrench with the connecting iron using the equation $\tau = r \cdot \vec{F}$, then compare the two torques produced.
d. Determine the torque produced by the mechanic before and after combining the wrench with the coupling iron using the equation $\tau = r : \vec{F}$, then compare the two torques produced.

Reason for choosing the answer ...
a. The equation used to determine torque is $\tau = r \cdot \vec{F}$
b. The equation used to determine torque is $\tau = r + \vec{F}$
c. The equation used to determine torque is $\tau = r - \vec{F}$
d. The equation used to determine torque is $\tau = r : \vec{F}$

1.4. The comparison of torque produced by the mechanic when using a wrench and a combination of iron and wrench is ...
a. 2 compared to 5
b. 5 compared to 2
c. 3 compared to 6
d. 1 compared to 5

The reason for choosing the answer ...
a. The torque produced by the wrench is 5 Nm, while the torque produced by the combination of the wrench and iron is 2 Nm.
b. The torque produced by the wrench is 2 Nm, while the torque produced by the combination of the wrench and iron is 5 Nm.
c. The torque produced by the wrench is 3 Nm, while the torque produced by the combination of the wrench and iron is 6 Nm.
d. The torque produced by the wrench is 1 Nm, while the torque produced by the combination of the wrench and iron is 5 Nm.

Figure 6. Example item

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

Figure 6 presents an example of a two-tier diagnostic test item comprising four primary questions. Among these is question item 1.1, which pertains to the challenges encountered by a mechanic in loosening a car wheel bolt. Test participants are prompted to provide a visualization of the issue associated with the problem statement and to justify their choice of that particular visualization. Item 1.2 encompasses inquiries regarding the characterization of the problem related to the problem statement and the rationale for selecting that particular characterization. Item 1.3 poses an investigation regarding the formulation of solutions pertinent to the problem statement and the rationale for choosing that specific solution strategy. Item 1.4 encompasses inquiries regarding the outcomes of executing the solution relevant to the problem statement and the justification for selecting those outcomes.

2. Creating flashcards based on augmented reality technology.

This design was created using the Paint 3D application, supplemented by auxiliary software such as Assemblr EDU and Canva, which are essential for picture editing and modification as needed. The flashcard comprises a title, accompanying graphics, usage instructions, and a barcode. In augmented reality, it contains the primary 3D object and accompanying issues for each question item. Figure 7 illustrates a flashcard display utilizing a barcode for access, allowing the loading of a 3D augmented reality model relevant to the challenge posed in the question item. The 3D model can be used in multiple orientations and dimensions, based on the user's selection, to analyze the issues presented in the question item.

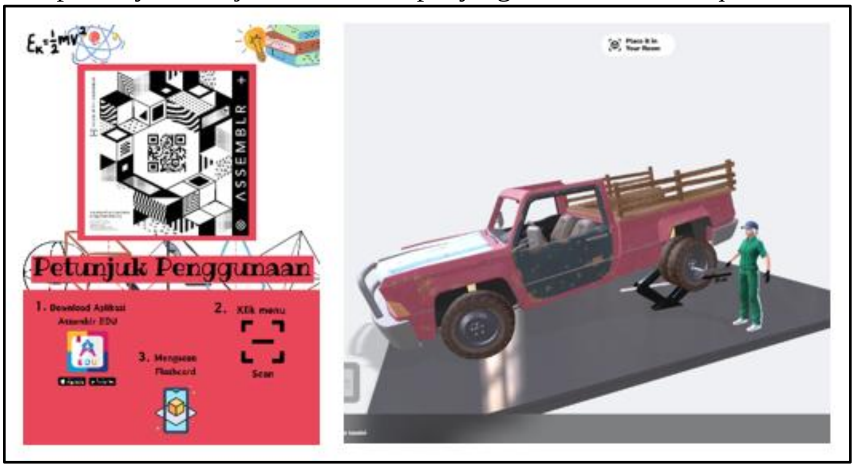


Figure 7. Flashcard and Augmented Reality Display

3. Designing a questionnaire

The questionnaire, designed using Google Forms, is divided into two parts: the validation questionnaire and the response questionnaire. The validation questionnaire consists of several components, such as title, identity, filling instructions, aspects, assessment instruments, scores, and notes. Meanwhile, the response questionnaire consists of components such as title, identity, filling instructions, statements, and assessments.

Development Stage

The development stage is the third phase of creating the two-tier diagnostic test instrument, supported by augmented reality flashcards. At this stage, content validation is performed, encompassing material validation, media validation, and language validation for the two-tier diagnostic test instrument facilitated by augmented reality flashcards. Content validation is an objective criterion that ensures the accuracy of the test instrument's content and aligns it with the overarching learning content and objectives. Consequently, an assessment that effectively gauges comprehension of the idea in question and possesses a substantial degree of content validity (Febriana, 2019). The content validation in this research was executed in two phases. The initial stage received validation from two physics education lecturers, while three high school physics instructors verified the subsequent stage. This procedure is used to acquire test enhancement recommendations and evaluations from the validators for analysis.

According to the validation results, the validator offered some recommendations for the testing. The physics lecturer's validation recommends incorporating cardinal directions into the response for question 1.1 to assist students in recognizing the object's movement direction and altering the colors of the conspicuous objects in augmented reality. Secondly, the validation from the high school physics instructor recommends substituting the term "Taxi" with "Taksi" in the initial statement and rewording the third statement for enhanced clarity and comprehension for the exam participants.

Alongside enhancement recommendations, the validators also delivered an evaluation of the content of the two-tier diagnostic test instrument enhanced by augmented reality flashcards. The scores from the validators were evaluated using the Aiken index, which considers content, medium, and language factors. The computations on the material aspect produced an average Aiken index value of 0.89, classified as high. The elevated category in the material element signifies that the evaluated content aligns with the learning outcomes and the progression of learning objectives, adheres to the question indicators, and meets the measurement objectives. The calculation results regarding the media component produced an average Aiken index value of 0.90. classified as high. The elevated classification in the media dimension indicates that the media is appropriate for testing purposes, as it corresponds with the content and is user-friendly. The calculating results for the linguistic feature produced an average Aiken index score of 0.91, categorizing it as high. The elevated category in the linguistic dimension signifies that the language utilized in the items adheres to proper grammatical conventions and does not generate ambiguous interpretations. The overall average Aiken index, based on these three aspects, is 0.90. classified as good.

In contrast, the material validation score was 0.89, the media validation score was 0.87, and the language validation score was 0.90. all of which were classified as high in the study by Muchtar & Koimah (2022), which created a web-based diagnostic test on redox concept material. Based on these three factors, the total average validation was 0.87, which is considered satisfactory. This suggests that, in comparison to the instrument used in the Muchtar & Koimah research, the instrument used in this investigation had a greater expert validity level. This enhancement results from the use of augmented reality (AR) technology, which not only improves visual and interactive elements but also helps pupils grasp concepts more thoroughly and with greater clarity. As a result, it has been demonstrated that including AR into the two-tier instrument greatly improves the evaluations' quality in terms of language, medium, and content.

This research includes both expert validation and early study validation. The objective is to elicit replies from test participants using a questionnaire. The examined reactions encompass elements of process, conceptual inquiries, questions, and stimuli. Participants' feedback on the test implementation suggests that the allotted 90 minutes were adequate to address all topics. Furthermore, the question instructions were unambiguous and did not perplex the test participants; each question was comprehensible, and the augmented reality objects in the flashcards were distinctly portrayed. Consequently, the two-tier diagnostic assessment enhanced by augmented reality flashcards is appropriate for implementation in extensive trials.

Implementation Stage

The implementation phase is the fourth stage of this research. A large-scale trial was undertaken, including 178 pupils (71 male and 107 female) from four distinct schools. The pupils were from the 11th and 12th grades, enrolled in physics. The students were allotted 90 minutes to complete the two-tier diagnostic assessment utilizing augmented reality flashcards.

The students' replies were evaluated utilizing the Rasch model via Winsteps software version 5.1.4. Characteristics identified through the Winsteps program encompass unidimensionality, dependability, item fit, Wright map, and item difficulty level.

Undimensionality signifies that the constructed assessment is capable of measuring solely a singular talent. Undimensionality signifies construct validity. A favorable indication of unidimensionality is a raw variance explained by a measure of over 20% (Sari & Mahmudi, 2024). The Winstep program yielded a raw variation explained by the measure of 20.2%, based on undimensionality data. This indicates that the augmented reality flashcard-assisted two-tier diagnostic test instrument is only valid for assessing problem-solving complexity. The challenge in problem-solving is in addressing Heller stage questions about rotational dynamics and rigid body equilibrium.

Reliability signifies that the devised test can yield consistent measurement outcomes (Febriana, 2019; Sari & Mahmudi, 2024). In the Rasch model, dependability is categorized into person reliability and item reliability. Cronbach's Alpha reliability measures the consistency of the relationship between test subjects and test items. Person reliability indicates the consistency of test-takers' responses, whereas item reliability reflects the consistency of the items being assessed. The Winstep program's reliability

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

statistics indicate a Cronbach's Alpha of 0.89, person reliability of 0.87, and item reliability of 0.88, all of which are categorized as good. This suggests that the reliability of test participants in responding to questions is commendable, and the quality of the test items in assessing the difficulty of problem-solving by the participants is classified as satisfactory.

Table 3. Rasch Analysis of the Result

Item	Total Correct Answer	Item Measure	Item FIT Outfit MNSQ	Outfit ZSTD	Pt Measure Corr
1	124	-0.32	0.95	-0.38	0.37
2	93	0.58	0.81	-2,42	0.55
3	131	-0.55	0.68	-2,36	0.56
4	119	-0.16	0.90	-0.88	0.43
5	138	-0.8	0.62	-2,46	0.57
6	142	-0.96	1.22	1,13	0.41
7	139	-0.84	0.79	-1,23	0.44
8	122	-0.25	1.16	1,27	0.39
9	87	0.74	1.10	1,20	0.37
10	140	-0.88	0.69	-1.84	0.52
11	125	-0.35	0.75	-2.05	0.56
12	90	0.66	1.42	4.48	0.15
13	111	0.08	0.97	-0.31	0.42
14	130	-0.51	0.97	-0.18	0.44
15	52	1,75	1.77	4,62	0.02
16	89	0.69	1.29	3,19	0.22
17	112	0.05	1.02	0.27	0.37
18	129	-0.48	1.08	0.63	0.36
19	91	0.64	0.92	-1.00	0.47
20	110	0.11	1.09	0.88	0.33
21	121	-0.22	1.00	0.00	0.36
22	102	0.33	0.86	-1.67	0.51
23	102	0.33	0.85	-1.72	0.52
24	100	0.39	1.11	1.24	0.33
25	103	0.31	0.84	-1.85	0.51
26	111	0.08	0.69	-3.44	0.63
27	121	-0.22	0.96	-0.29	0.41
28	106	0.22	0.99	-0.07	0.38
29	89	0.69	1.18	2.06	0.29
30	98	0.44	1.22	2.39	0.23
31	104	0.28	1.12	1.27	0.36
32	118	-0.12	1.34	2.75	0.14
33	131	-0.55	0.59	-3.23	0.66
34	125	-0.35	0.83	-1.30	0.50
35	122	-0.25	0.71	-2.63	0.60
36	119	-0.16	1.06	0.57	0.32
37	127	-0.41	0.92	-0.58	0.48
38	125	-0.35	0.74	-2,16	0.58
39	124	-0.32	0.98	-0.15	0.38
40	119	-0.16	0.96	-0.34	0.44
41	114	-0.01	0.85	-1.43	0.51
42	110	0.11	1.03	0.33	0.37
43	119	-0.16	0.90	-0.87	0.45
44	96	0.5	0.93	-0.77	0.44
45	105	0.25	0.99	-0.04	0.42
46	119	-0.16	1.20	1.65	0.24
47	112	0.05	1.10	0.94	0.32
48	105	0.25	1.25	2.58	0.22

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

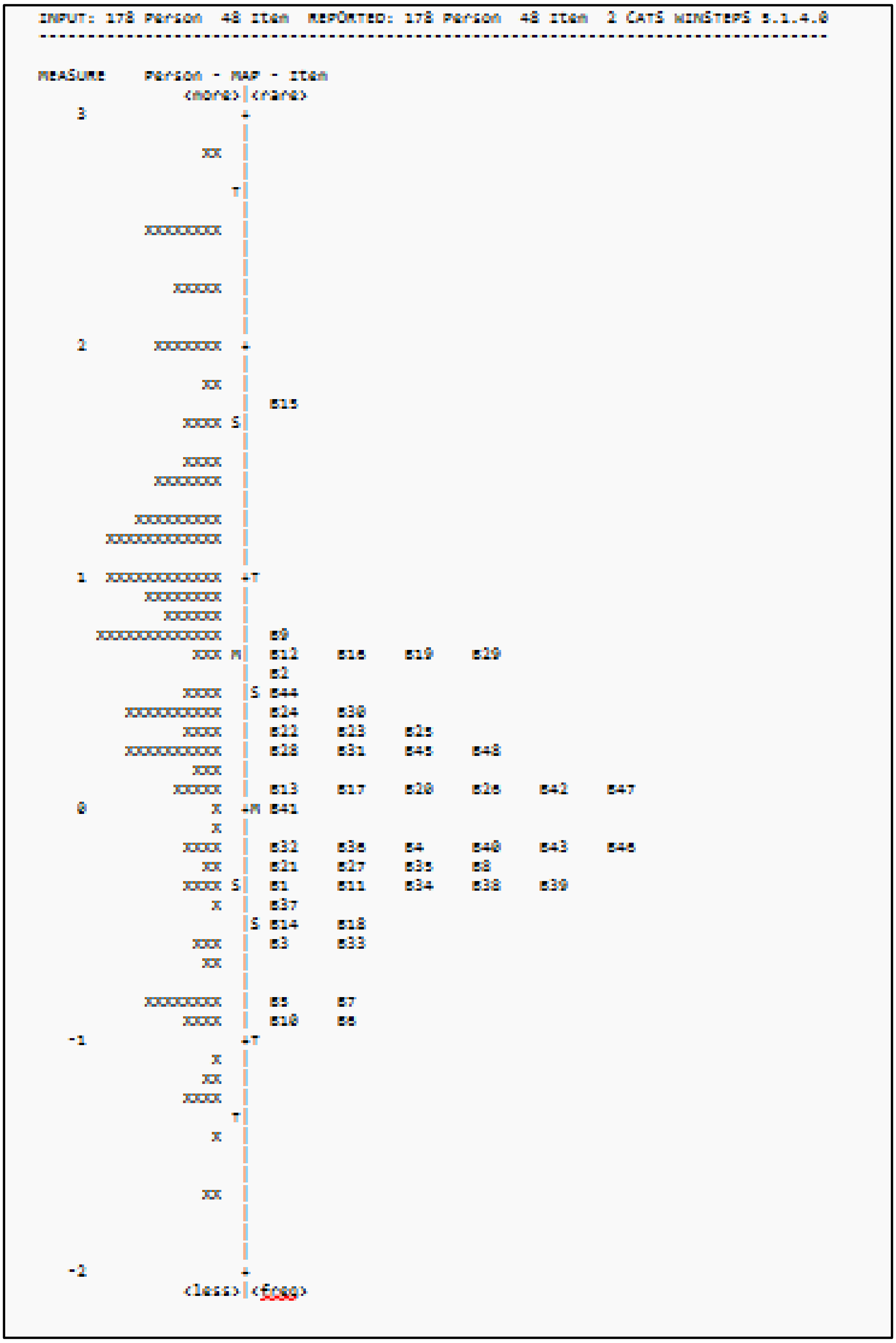


Figure 8. Wright Map

The degree of item fit signifies the item's efficacy in measuring (Sari & Mahmudi, 2024; Sumintono & Widhiarso, 2015). The fit of an item inside the Rasch model is assessed using the metrics of Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr (Istiyono & Tuanaya, 2023). The item-fit results in Table 3 indicate that the majority of items align with the Rasch model, except items 12, 15, 16, 29, 30, 32, and 48, which exhibit ZSTD values over +2 and Pt Measure values below 0.40. According to Table 3, items numbered 12, 15, 16, 29, 30, 32, and 48 exhibit Outfit ZSTD values that are deemed unreasonable, falling outside the acceptable range of -2 to +2, and Pt Measure Corr values that are also deemed unreasonable, exceeding the specified range of 0.4 to 0.85. Consequently, these items are classified as unsuitable for measurement and non-compliant with the Rasch model. Consequently, 41 items are considered suitable, while the remaining seven items have been amended.

The Wright map illustrates the distribution of student competencies alongside the distribution of item difficulty levels (Istiyono & Rifqi Kama, 2023; Planinic et al., 2019). The Wright map on the left illustrates the distribution of student skills relative to the test instrument, with the distribution derived from the logit value of the person measure. According to Figure 8, 80 sample students (44.94%) demonstrated enough comprehension of the test, as indicated by a logit measure value beyond 0.93, while 98 sample students (55.06%) exhibited insufficient understanding, with a logit measure value below 0.93. The "X" symbol denotes the level of a single student sample. The "XX" mark indicates the scores of two test participants with a logit value of 2.82, signifying their superior ability to answer nearly all questions correctly. Conversely, the two test participants marked with "XX" possess a logit value of -1.69, indicating their inferior ability, though they can answer some questions correctly.

The Wright map on the right illustrates the distribution of item difficulty levels, derived from the measured logit values of the items. The difficulty level of the test items reflects their ease or challenge based on the overall score of correct responses. The item identified as the most challenging for pupils, according to the map, is Item 15, which has a logit value of 1.75. This inquiry examines the outcomes of the implementation phase in addressing the issue of comparing the angular momentum of the front and rear bicycle gears. The primary reason most test takers answered this question poorly is their inadequate attention to the physical description provided and errors in mathematical computations. Table 3 indicates that merely 52 of the 178 students (29.2%) responded to these questions. The item most easily answered by test participants was Item 6, which had a logit value of -0.96. This article addresses the formulation of solution methods for comparing torque when loosening wheel nuts using wrenches of varying lengths. Item 11 can be answered by nearly all students, with 142 out of 178 students (79.8%).

The comprehensive analysis of augmented reality (AR)-based flashcard media suggests that this diagnostic tool enhances student engagement due to its visual appeal and ability to contextualize abstract concepts concretely. Moreover, the diagnostic feedback provided can help educators formulate more precise remedial instruction tailored to the specific challenges encountered by students. The presence of this instrument is both statistically valid and possesses significant pedagogical value in facilitating a more adaptive and data-driven learning approach.

Evaluation Stage

The evaluation phase constitutes the concluding stage of the ADDIE model. The evaluation outcomes from the product development process completed at each level of ADDIE are as follows: 1) The evaluation results during the Analysis stage indicated that the autonomous curriculum has been extensively applied. Nonetheless, Curriculum 2013 remains in effect. The subject deemed challenging is rotational dynamics. Furthermore, findings were obtained regarding awareness of augmented reality flashcard media and challenges in problem-solving, as per the Heller phases. Based on these results, the researcher correlated the developed product with the learning outcomes of rotational dynamics as defined in the independent curriculum and integrated it with augmented reality-based flashcard learning media to assess the problem-solving challenges encountered by students at the Heller stages. 2) During the design phase, the evaluation outcomes necessitated modifications to various narratives inside the statements and adjustments to punctuation. Moreover, various components were added, and several colors on the flashcards and 3D objects in augmented reality were modified. 3) During the development phase, the evaluation outcomes comprised the incorporation of multiple problem instructions regarding 3D objects in augmented reality, the modification of terminology in several statements, and the product's validity, which has been classified as satisfactory based on assessments by subject matter experts, media specialists, and language authorities. 4) During the implementation phase, the evaluation results revealed a constraint in the time allotted for the task, as certain test participants were unable to complete all questions in the two-tier diagnostic assessment utilizing augmented reality flashcards on rotational dynamics within the designated timeframe. The assessment during the implementation phase revealed varying degrees of student comprehension, a moderate average question difficulty, and satisfactory question reliability. The two-tier diagnostic test, enhanced by augmented reality flashcards, was deemed valid according to Rasch model data analysis conducted with Winsteps software version 5.1.4.

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

The study's results demonstrate that the two-tier diagnostic test, supplemented by augmented reality flashcards, effectively identifies student issues; however, it also has some drawbacks. The generalizability of the findings is constrained due to the sample being derived from a limited number of schools within a specific region. The restricted implementation timeframe precludes the possibility of a long-term assessment. Third, students possess varying degrees of experience with augmented reality-based media, which can influence both their comfort and the efficacy of media utilization. Contemplation of these constraints serves as a crucial foundation for subsequent research aimed at broadening the reach, enhancing the assessment of effects, and creating media that is better attuned to student characteristics..

CONCLUSION

This study demonstrates that augmented reality flashcard-assisted diagnostic examinations effectively identify pupils' problem-solving issues. The analysis, grounded in the ADDIE model stages, suggests a need for new media to enhance comprehension of problem-solving concepts. The augmented reality flashcard features a two-tier diagnostic assessment to investigate the underlying causes of challenges in problem-solving. The creation of augmented reality flashcard-assisted diagnostic examinations was conducted with expert validation and preliminary trials demonstrating feasibility, practicality, and user-friendliness. The executed implementation showed the efficacy of augmented reality flashcards in diagnosing problem-solving challenges. The final assessment demonstrated that augmented reality flashcards effectively detected pupils' difficulties in problem-solving. This study conceptually enhances the diagnostic evaluation approach by incorporating spatial and visual components that facilitate conceptual understanding. This tool provides an adaptive and engaging assessment alternative for students, yielding more precise data for teachers to inform learning interventions. Consequently, the findings of this study enhance evaluation methodologies in physics education by adopting a more contextual, individualized, and technology-driven approach.

REFERENCE

Anggraeni, Y., & Listiaji, P. (2024). An Application of Problem-Solving-Based Student Facilitating and Explaining Teaching Model to Improve Students' Metacognitive Ability. *Jurnal Inovasi Pendidikan IPA*, *10*(1), 47–61. https://doi.org/10.21831/jipi.v10i1.65617

Angraini, L. M., Susilawati, A., Noto, M. S., Wahyuni, R., & Andrian, D. (2023). Augmented Reality for Cultivating Computational Thinking Skills in Mathematics Completed with Literature Review, Bibliometrics, and Experiments for Students. *Indonesian Journal of Science and Technology*, *9*(1), 225–260. https://doi.org/10.17509/ijost.v9i1.67258

Anton, A., & Trisoni, R. (2022). Peran Guru dalam Mengembangkan Kompetensi Sosial pada Anak Usia Dini di TK Cempaka Balikpapan. *Edu Cendikia: Jurnal Ilmiah Kependidikan*, *2*(03), 528–535. https://doi.org/10.47709/educendikia.v2i03.1895

Arsyad, A. (2011). *Media Pembelajaran*. Jakarta : Rajawali Pers.

Fajriansyah, M. A., & Widodo, S. (2022). Augmented Reality Sebagai Media Pembelajaran di Rumah yang Efektif tetapi Menarik. *JURNAL TEKNIK INFORMATIKA UNIS*, *10*(2), 173–180. https://doi.org/10.33592/jutis.v10i2.2547

Febriana, R. (2019). *Evaluasi Pembelajaran*. Bumi Aksara, Jakarta.

Gagne, R. ., & Briggs, L. . (1979). *Principles of Instructional Design* (Second Edi).

Ginting, F. W. (2021). ANALISIS KEMAMPUAN PEMECAHAN MASALAH MAHASISWA CALON GURU FISIKA DALAM MENYELESAIKAN SOAL DINAMIKA ROTASI. *Jurnal Pendidikan Fisika*, *10*(1), 60–66. https://doi.org/10.22611/jpf.v10i1.25264

Habellia, R. C., Maria, H. T., & Hidayatullah, M. M. S. (2021). Pengembangan two tier diagnostik test untuk mendeteksi miskonsepsi siswa sma pada materi gerak. *Jurnal Pendidikan Informatika Dan Sains*, *10*(2), 195–201. https://doi.org/10.31571/saintek.v10i2.3222

Istiyono, E., & Rifqi Kama, A. (2023). Analysis of numeracy ability test item characteristics grade VIII students with mixed model item response theory (IRT) approach. *Materials of International Practical Internet Conference "Challenges of Science,"* 184–195. https://doi.org/10.31643/2023.22

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

Istiyono, E., & Tuanaya, R. (2023). Validity and Reliability Analysis of Character Instruments for High School Students in Coastal Areas of Maluku. *Materials of International Practical Internet Conference "Challenges of Science,"* 245–250. https://doi.org/10.31643/2023.31

Malik, A., Al-Hanafi, M. A., Chusni, M. M., & Nugraha, A. R. (2023). Development of CORE-based Electronic-modules to Improve Students' Problem-Solving Skills. *Jurnal Penelitian Pendidikan IPA*, *9*(8), 6415–6424. https://doi.org/10.29303/jppipa.v9i8.3172

Muliadi, R., Jatnika, R., Hanami, Y., & Adishesa, M. S. (2025). Evaluation of an environmentally friendly behavior scale among Indonesian university students using the Rasch model. *Acta Psychologica*, *257*, 105087. https://doi.org/10.1016/j.actpsy.2025.105087

Mutlu, A., & Sesen, B. A. (2015). Development of a Two-tier Diagnostic Test to Assess Undergraduates' Understanding of Some Chemistry Concepts. *Procedia - Social and Behavioral Sciences*, *174*, 629–635. https://doi.org/10.1016/j.sbspro.2015.01.593

Partono, P., Wardhani, H. N., Setyowati, N. I., Tsalitsa, A., & Putri, S. N. (2021). Strategi Meningkatkan Kompetensi 4C (Critical Thinking, Creativity, Communication, & Collaborative). *Jurnal Penelitian Ilmu Pendidikan*, *14*(1), 41–52. https://doi.org/10.21831/jpipfip.v14i1.35810

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 020111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Pujiastuti, H., & Haryadi, R. (2024). The Effectiveness of Using Augmented Reality on the Geometry Thinking Ability of Junior High School Students. *Procedia Computer Science*, *234*, 1738–1745. https://doi.org/10.1016/j.procs.2024.03.180

Rahmawati, Rustaman, N. Y., Hamidah, I., & Rusdiana, D. (2019). Pengembangan Instrumen Tes Keterampilan Pemecahan Masalah pada Materi Kelistrikan dan Kemagnetan. *Seminar Kontribusi Fisika*, 98.

Romero-Saldaña, M., Sánchez-Thevenet, P., Almodóvar-Fernández, I., González-Moret, R., Salas-Medina, P., Ranchal-Sánchez, A., Gutiérrez-Sánchez, D., & López-Leiva, I. (2024). Development and validation of a new satisfaction scale for objective structured clinical assessments (S-OSCA): A multicenter cross-sectional study. *Nurse Education Today*, *141*, 106308. https://doi.org/10.1016/j.nedt.2024.106308

Sari, E. D. K., & Mahmudi, I. (2024). *Analisis Pemodelan Rasch Assessment Pendidikan* (E. Safitry (ed.); Pertama). Jakarta : PT. Pena Persada Kerta Utama.

Siahaan, S. M., & Putri, R. E. (2024). The Augmented Reality Based Flashcards for Learning Heat and Thermodynamics in High School. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *10*(1), 99–108. https://doi.org/10.21009/1.10109

Silitonga, H. T. M., Oktavianty, E., & Habellia, R. C. (2024). Development Of Problem-Solving Measuring Instrument In Motion Topic Using the Testlet Model. *Jurnal Penelitian Pendidikan IPA*, *10*(11), 8755–8761. https://doi.org/10.29303/jppipa.v10i11.6327

Sirait, J., Ricky, P., & Hamdani, H. (2023). Development and Validation of the Force Diagrams Representation Test. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *9*(2), 183–196. https://doi.org/10.21009/1.09201

Subekti, A. O., Rahayu, R., & Trisnowati, E. (2025). The Application of PjBL with STEM Integration: Its Effect on Science Problem-Solving Abilities. *Jurnal Inovasi Pendidikan IPA*, *11*(1), 352–362. https://doi.org/10.21831/jipi.v11i1.77715

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi : Trim Komunikata.

Syahrial, A. H., Wulan Deliana, Vina Dwi Cahyani, & Ahmad Fakhri Husaini. (2022). Pembelajaran Fisika Materi Mekanika Benda Tegar: Review Media, Model, dan Metode. *Mitra Pilar: Jurnal Pendidikan, Inovasi, Dan Terapan Teknologi*, *1*(2), 119–140. https://doi.org/10.58797/pilar.0102.06

Utami, F., Rukiyah, R., & Andika, W. D. (2021). Pengembangan Media Flashcard Berbasis Augmented Reality pada Materi Mengenal Binatang Laut. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, *5*(2), 1718–1728. https://doi.org/10.31004/obsesi.v5i2.933

Utami, L., Festiyed, Ilahi, D. P., Ratih, A., Yenti, E., & Lazulva. (2024). ANALISIS INDEKS AIKEN UNTUK MENGETAHUI VALIDITAS ISI INSTRUMEN SCINETIFIC HABBITS OF MIND.

ISSN 2406-9205 (print), ISSN 2477-4820 (online)

Journal of Research and Education Chemistry, *6*(1), 59. https://doi.org/10.25299/jrec.2024.vol6(1).17430

Widyaningsih, R., & Aloysius, H. P. (2023). Identifikasi Miskonsepsi Termokimia Menggunakan Tes Diagnostik Pada Peserta Didik SMA Kelas XI MIPA. *Jurnal Riset Pembelajaran Kimia*, *8*(2), 58–66. https://doi.org/https://doi.org/10.21831/jrpk.v8i2.20627

ISSN 2406-9205 (print), ISSN 2477-4820 (online)