A Hybrid Automatic Scoring System: Artificial Intelligence-Based Evaluation of Physics Concept Comprehension Essay Test

Wawan Kurniawan[1], Cicyn Riantoni[1,*], Neneng Lestari[1], and Doni Ropawandi[2]

[1]Department of Physics Education, Faculty of Teacher Training and Education, Universitas Jambi, Jambi, Indonesia
[2]Faculty of Education, National University of Malaysia, Malaysia
Email: kurniawan_wawan@unja.ac.id (W.K.); cicynriantoni@unja.ac.id (C.R.); nenenglestari@unja.ac.id (N.L.); p96161@siswa.ukm.edu.my (D.R.)
[*]Corresponding author

Manuscript received October 16, 2023; revised December 25, 2023; accepted January 26, 2024; published June 19, 2024

***Abstract*—Conceptual understanding is one of the keys in physics. The purpose of this research was to develop an artificial intelligence-based system used to assist instructors in evaluating students' conceptual understanding essay test results. This research used a method with Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model (). The research subjects were lecturers and undergraduate students of physics education at the University of Jambi. Data collection was done through interviews and tests. The instruments used in this research were interview sheets and essay questions. The results of the feasibility test analysis showed that the hybrid automatic scoring system was overall suitable with an average score of 4.40 (very suitable category) in the aspects of software engineering and visual communication. In addition, the analysis of students' conceptual understanding using the Hybrid Automatic Scoring System showed that only 21.9% of students had good conceptual understanding. The remaining 40.6% of students fell into the partial understanding category, 9.4% fell into the specific misconception category, and 28.2% had poor conceptual understanding. These results are not very different from the results of the manual analysis. The manual analysis showed that 21.9% of the students had good concept understanding, 40.6% had partial understanding, 12.5% had specific misconceptions, and 25% had poor concept understanding. These results indicate that the Automated Hybrid Scoring System can be used as an assessment tool for essay tests of conceptual understanding. The results of this research contribute in the form of a technology that can assist physics instructors in correcting, scoring, and providing feedback on physics test scores based on essay questions.**

***Keywords*—automatic scoring system, artificial intelligence, essay test, physics concept**

I. Introduction

The primary goal of physics education is to improve conceptual understanding [1–4]. It's about understanding physics concepts in depth and applying them in problem solving processes [5–7]. In addition, improving quantitative understanding of physics concepts is also a goal of physics education [8]. This is because comprehensive mastery of physics requires the ability to combine conceptual understanding with quantitative understanding of physics concepts.

Most of the literature on conceptual understanding research has focused primarily on identifying conceptual difficulties (i.e., misconceptions, alternative conceptions, and naive conceptions) [9, 10]. For example, research conducted by Scaife and Heckler [10] investigated students' difficulties in understanding the concepts of electric force and magnetic forces. The study by Dai *et al.* [11] assessed students' conceptual understanding of light interference. In addition, research conducted by Mesic *et al.* [12] measured students' conceptual understanding of wave optics using a modeling approach.

There are three theoretical frameworks that can be used to assess students' conceptual understanding, namely the theory of naive views, often referred to as the misconception theory, the theory of fragmented knowledge, often referred to as the resource theory, and the ontological perspective [2]. Among these three theories, the misconceptions theory has been more extensively developed and researched compared to the others [13, 14]. On the other hand, the resource theory is the least researched.

To assess students' conceptual understanding using these three theoretical frameworks, various types of test instruments can be used. These include multiple-choice tests, multiple-choice reasoning tests, two-tier tests, three-tier tests, four-tier tests, and essay tests. For example, a study conducted by Rahmawati *et al.* [15] aimed to collect information about students' understanding of rotational dynamics using multiple-choice questions. Sutopo [7] conducted research to assess students' understanding of mechanical waves using multiple-choice questions. Maison *et al.* [16] conducted a study to assess students' understanding of light using a four-tier instrument. Among the various types of tests that can be used to assess students' conceptual understanding, research that uses essay tests to assess students' conceptual understanding is quite rare.

Some studies indicate that instructors have difficulty evaluating essay responses, require a significant amount of time to evaluate students' essay responses, and encounter consistency problems both within the same evaluator and across evaluators [17]. When essay scoring is done by a single person, inconsistencies can occur if scores are done at different times. Furthermore, when scoring is done by different people, different results may be obtained for the same response. In addition, other research indicates that the problem of long grading times can lead to instructor fatigue [18, 19]. As a result, there may be errors in grading exams after an extended period of time.

Physics education plays a critical role in shaping students' understanding and knowledge of physics [20–22]. However, a significant challenge arises when it comes to accurately assessing students' understanding and knowledge through essay tests [21, 23]. The problem lies with certain physics

instructors who, for a variety of reasons, struggle to effectively evaluate essay tests. They may lack the necessary expertise in essay assessment techniques or have limited knowledge of the specific instructional strategies used in physics education research [24, 25]. In addition, these instructors may also be challenged by the subjective nature of essay assessment, making it difficult to ensure consistent and accurate grading [22]. Their limited expertise in essay assessment techniques can lead to inconsistent and inaccurate grading, making it difficult to determine the true level of student understanding and knowledge [24].

To address these issues, teachers are required to be innovative in seeking solutions to educational problems [26–28]. Some previous studies have attempted to find solutions to the challenges of analyzing students' conceptual understanding in essay tests. For example, an online-based essay assessment application was developed using an algorithm with cosine similarity method [29]. This research provides an automated essay assessment that goes through a series of steps, including text mining, question weighting, and cosine similarity, to obtain essay scores. The results of this study show that the automated essay scoring application can be used. The study also compared the learning outcomes between conventional (manual) grading and automated grading using the essay assessment application. However, the assessment results of the application were only presented in numerical form and were limited to testing specific subjects, including Bahasa Indonesia. It could not evaluate answers in the form of formulas or numbers.

To address these challenges, the researchers propose an automated assessment system that provides scores for students' conceptual understanding in numerical and graphical forms. This automated essay scoring system can also be used for questions and answers that involve formulas or numbers. In this context, the researchers use an Intelligent Teaching Assistant System (ITAS). ITAS is designed for both students and instructors and aims to intelligently support the education or training process by assisting instructors in their tasks and students in their learning [30]. ITAS can assist in accurate, systematic and faster learning diagnosis and assessment [31]. It can significantly reduce the burden on instructors, assist them with tedious or complex tasks, track student performance, report problems, while helping students practice at their own pace in a customized environment, and provide feedback and tailored exercises [30, 32, 33].

Some previous researchers have tried to develop artificial intelligence-based evaluation systems. For example, research using the K-Nearest Neighbor (KNN) method as an evaluation tool does not provide keywords, but only matches texts based on reference texts. Research using the support vector regression method also does not use keywords, but relies solely on document similarity [34]. Another example is research using the Nazief and Adriani stemming algorithm [29], which uses keywords as key terms. However, it has the disadvantage that the evaluation system does not consider words with the same meaning (synonyms). In this proposed research, the research team also uses Artificial Intelligence (AI). However, what distinguishes it from other studies and what makes the novelty in this research is that it lies in the developed Artificial Intelligence (AI) that is dedicated to evaluating the results of students' concept understanding essay tests in physics subjects. The results of this development are expected to help lecturers, teachers and physics instructors to easily assess students' concept understanding with essay tests.

Based on the problems and objectives of this research, this research team will develop and implement an artificial intelligence-based essay scoring system called A Hybrid Automatic Scoring System.

II. Literature Review

A. Artificial Intelligence (AI)

At present, with the rapid development of science and global technology, Artificial Intelligence (AI) technology has experienced significant growth. AI technology is continuously developing and being widely used in various fields [19, 35]. It is an undeniable fact that AI is increasingly finding its way into the educational environment and teaching processes in schools [36]. As it develops, more and more people are recognizing the importance of this technology in education. AI has been used extensively in education and has shown significant application benefits, significantly impacting the teaching process and classroom management [18].

One application of AI in education is in the learning assessment process. AI technology, such as AI-assisted assessment, provides convenience in evaluating learning outcomes [18]. Student assessment is an important aspect of the learning process. In traditional teaching, it takes a long time for teachers to complete assessment tasks, such as question preparation, grading, performance evaluation, and exam paper analysis. AI makes the assessment methods more diverse, the assessment process more scientific, and the assessment results more accurate [19, 35].

In addition to replacing teachers in the classroom, AI technology can automatically grade assignments and exams [37]. Grading student assignments and exams is a routine task for instructors. Lengthy grading times can lead to instructor fatigue, and errors may occur when grading exam papers after a long period of time. Image recognition technology helps instructors relieve themselves from the heavy workload of grading student assignments and exams [38], with a low error rate.

B. VOSviewer Data

To understand the importance of this research, the researcher conducted a literature review using vosviewer to identify gaps in novelty or urgency of upcoming research. The literature review was conducted on 2,000 articles published from 2010 to 2023, focusing on the topics of conceptual understanding, artificial intelligence, and forms of conceptual understanding tests. The results of the analysis of these articles using VOSviewer indicate relationships with previously researched topics related to conceptual understanding, artificial intelligence, and forms of conceptual understanding tests, as shown in Fig. 1.

Fig. 1 illustrates the overall structure of the relationships found in several previous research studies related to the topics of conceptual understanding, artificial intelligence, and forms of conceptual understanding tests. When analyzed more specifically, the data show that previous research has linked conceptual understanding to several other topics, such as

natural language processing, deep learning, prediction, conceptual understanding tests, and others. Most of the studies on conceptual understanding have linked it to natural language processing, deep learning, and prediction, with the subjects being primarily students. These results provide an opportunity for researchers to develop an AI-based essay assessment tool to assist instructors in assessing understanding of physics concepts, called a hybrid automatic scoring system.

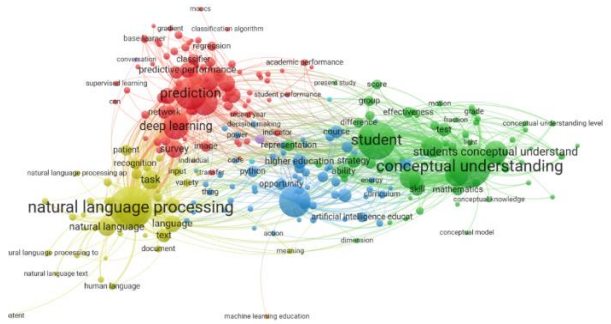

Fig. 1. The relationship of conceptual understanding and artificial intelligence with several research topics.

C. The ADDIE Research Design

The ADDIE research design is a systematic approach consisting of five main phases: Analysis, Design, Development, Implementation, and Evaluation [39]. This approach is used to design, develop, and evaluate learning or training programs. The ADDIE approach can be applied in a variety of contexts, including formal education, corporate training, and product development [40]. The main advantages of this approach are its flexibility and its ability to ensure that the design of learning or training is focused on identified goals and needs [41].

Many research topics in the field of education use the ADDIE approach to design, develop, and evaluate learning or training programs. Examples include the development of instructional materials based on the ADDIE model [42], the creation of training and development programs using the ADDIE method, and a comparative study of the ADDIE instructional design model in distance education [39].

III. Methods

This research is a type of research and development. The research design used is based on the ADDIE model (Analysis, Design, Development, Implementation, and Evaluation). Each procedure of the research design ADDIE model has its own methods, instruments, research subjects, and data analysis processes, which essentially form a unified whole with the aim of producing an artificial intelligence-based system called A Hybrid Automatic Scoring System, which is used to assist instructors in evaluating students' concept comprehension essay test results. The presentation of the research procedures using the ADDIE design model is explained in detail as follows.

A. Analysis

The purpose of the analysis is to gain an overview of the context of the issues that instructors face when assessing students' conceptual understanding. In addition, it aims to gather information related to the development of assessment formats, instruments, and media for conceptual understanding tests from several previous studies. Data collection in this phase of the analysis is based on the results of a literature review conducted using systematic literature review techniques. Data analysis from the systematic literature review is presented in a descriptive data presentation format. The analysis process consists of data condensation, data presentation, and data summarization.

B. Design

The purpose of this design phase is to create and organize the initial prototype of a hybrid automated scoring system that will be used to assist instructors in evaluating the results of students' conceptual understanding essay tests. The initial prototype is designed according to the steps of the universal design method developed by Lampos *et al.* [32], which include: (1) generating ideas; (2) reviewing and testing ideas; (3) design requirements; (4) design proposals; (5) product sketches; and (6) initial prototype. Each of these design steps has its own data collection techniques. In this design phase, the research team also considers input and suggestions from physics education instructors at the University of Jambi. The data obtained from each design step constitute the design of the initial prototype of the AI-based assessment system. The framework of the hybrid automated assessment system is shown in Fig. 2.

Fig. 2 explains the system of the A Hybrid Automatic Scoring System that will be developed.

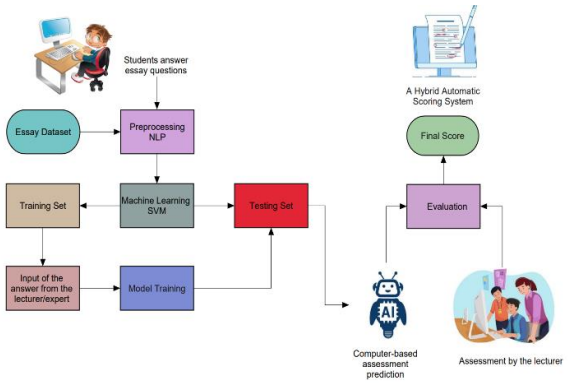

Fig 2. The framework of a hybrid automatic scoring system.

C. Develop

In this phase, the researcher transforms the designed prototype into a product called A Hybrid Automatic Scoring System that can be used to assist instructors in evaluating students' conceptual understanding essay test results. After the initial prototype is completed, the researcher collects information to obtain feedback from experts and users (instructors). The feedback and comments from experts and users (instructors) will serve as considerations for improving the prototype. The collection of feasibility data for the A Hybrid Automatic Scoring Program will be conducted through in-depth interviews using open-ended questions.

D. Implementation

In this phase, the researcher will collect information related to students' conceptual understanding based on an analysis conducted using the developed A Hybrid Automatic Scoring System and compare it with manually analyzed results. These results will provide information about the strengths and

weaknesses of the A Hybrid Automatic Scoring System in scoring the results of students' essay tests on conceptual understanding. The subjects of implementation are physics education students at the University of Jambi. The data collection instruments in this study consist of conceptual understanding assessment tools in the form of essay questions developed by the researcher, which have undergone validity and reliability testing. The assessment instruments will be entered into the A Hybrid Automatic Scoring System and used as a tool to provide information about the level of conceptual understanding of students.

IV. Result and Discussion

A. The Form of the Development Results of the A Hybrid Automatic Scoring System

In this development, the researcher created a prototype as a first step. This prototype will go through a series of testing stages aimed at achieving significant improvements in various aspects, such as concept, technical functionality, technical product operation, technology used, and desired value aspects. An example part of the prototype is shown in Fig. 3.

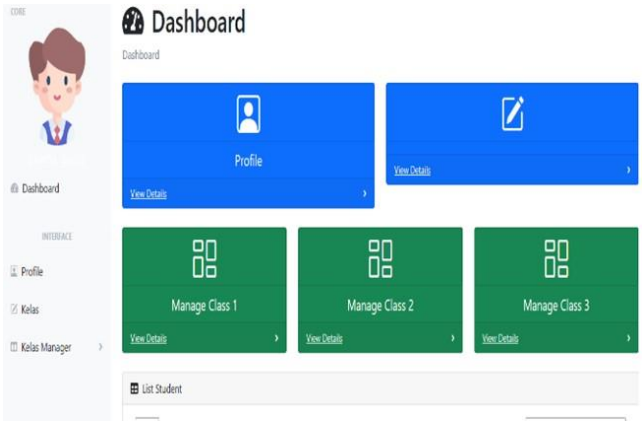

Fig. 3. Lecturer dashboard page.

This prototype will undergo a series of test phases aimed at achieving significant improvements in various aspects, such as the concept, the technical functionality, the technical operation of the product, the technology used and the desired value aspects.

B. The Feasibility of a Hybrid Automatic Scoring System in Assessing Students' Concept Comprehension Essay Test

Before the conceptual understanding assessment system is implemented, feasibility testing is required. At this stage, there are two types of tests, namely functionality test and expert validation test. The results of the functionality test of A Hybrid Automatic Scoring System show that each function in the system can function effectively. Therefore, the feasibility study process can proceed to the expert validation testing stage.

The expert validation stage is conducted to test the usability, functionality, and visual communication aspects. The results of expert validation are shown in Table 1.

Table 1 shows that the average scores obtained from the validators are 4.63 for usability, 4.47 for functionality, and 4.72 for visual communication aspects. The average percentage from this validation is 4.60. These results provide an overview that the developed system falls into the 'highly suitable' category for implementation as a tool to assist in the assessment of students' concept comprehension essay test scores.

Table 1. Analysis of expert validation results

Item	Indicator	Expert Validation		
		Mean	%	Category
1	Usability	4.63	92.50%	Highly Suitable
2	Functionallity	4.47	89.47%	Highly Suitable
3	Visual Communication	4.72	94.54%	Highly Suitable
	Mean	4.60	92.17%	Highly Suitable

C. Students' Understanding of Physics Concepts Based on the Assessment Results Using the A Hybrid Automatic Scoring System

Conceptual understanding refers to a person's ability to understand and internalize concepts or ideas related to a particular subject or discipline. It includes a deep understanding of the principles, theories, rules, and relationships within a field of knowledge or a specific subject.

There are two types of electrical circuit questions used to assess students' conceptual understanding. The first problem is presented in the context of the brightness of two lamps before and after the addition of a resistor in series. The second problem is presented in the context of changes in the brightness of lamps in series and in parallel after the switch in the circuit is closed. To solve these problems, students must have a basic understanding of 1) the concepts of current, voltage, and resistance in series and parallel circuits, and 2) the concepts of energy and electrical power.

The results of the analysis of the responses to one of the questions using A Hybrid Automatic Scoring System indicate several categories of students' conceptual understanding (Table 2).

The above analysis results indicate that only 21.9% of the students have a good understanding of the concept. The remaining 40.6% of students are categorized as having a partial understanding. In addition, 9.4% of the students are categorized as having specific misconceptions and 28.2% do not understand the concept. Nearly identical results were found based on manual analysis, as shown in Table 3.

The results in Table 3 show that the percentages of students who understand the concept well and those who have a partial understanding show consistent results with the analysis using a hybrid automatic scoring system. However, there is a slight difference in the results for the percentage of students with specific misconceptions and those who do not understand the concept. Approximately 12.5% of students are identified as having specific misconceptions, while 25% of students are identified as not understanding the concept.

Table 2. Categorization of students' conceptual understanding.

N	Category	Applied Concepts
3 (9.4%)	Specific Misconceptions	• The energy in the circuit will flow first to the A bulb. Even though the resistance is added, the brightness of the A bulb will remain the same. • At B bulb, the light will dim because it get a little energy. This because of the energy have used by resistance

N	Category	Applied Concepts
		The A bulb in circuits 1 and 2 are equally bright, because they are fed by the same current. However, the brightness of the B bulb in both circuits are different because a resistance has been added in circuit 2. This causes the current that flows through the B bulb on circuit 2 decrease
9 (28.2%)	Don't understand	unclear
13 (40.6%)	Partial Understanding	• The addition of resistances increases total resistance • The greater of total resistance make the current smaller, so that the light dims.
7 (21.9%)	Good Understanding	• The brightness of A and B bulb are caused by flowing currents and voltages associated with electrical power. • In both circuits arranged in series, the current for each bulb is the same, meanwhile the voltage is divided. • In circuit 2, there are additional resistance that cause the decrease of current, so that the power becomes smaller.

Table 3. Categorization of students' conceptual understanding.

N	Category	Applied Concepts
4 (12.5%)	Specific Misconceptions	• The energy in the circuit will flow first to the A bulb. Even though the resistance is added, the brightness of the A bulb will remain the same. • At B bulb, the light will dim because it get a little energy. This because of the energy have used by resistance
		The A bulb in circuits 1 and 2 are equally bright, because they are fed by the same current. However, the brightness of the B bulb in both circuits are different because a resistance has been added in circuit 2. This causes the current that flows through the B bulb on circuit 2 decrease
8 (25%)	Don't understand	unclear
13 (40.6%)	Partial Understanding	• The addition of resistances increases total resistance • The greater of total resistance make the current smaller, so that the light dims.
7 (21.9%)	Good Understanding	• The brightness of A and B bulb are caused by flowing currents and voltages associated with electrical power. • In both circuits arranged in series, the current for each bulb is the same, meanwhile the voltage is divided. • In circuit 2, there are additional resistance that cause the decrease of current, so that the power becomes smaller.

V. Discussion

One of the various applications of AI in education is in the learning assessment process. AI technology, for example, provides convenience for assessing learning outcomes [18, 43]. Student assessment is an important aspect of the learning process. In traditional teaching, it takes a long time for teachers to complete assessment tasks, such as preparing questions, grading, performance evaluation, and analyzing exam papers. AI introduces more diverse teaching evaluation methods, a more scientific evaluation process, and more accurate evaluation results [19, 35].

AI technology not only serves as a substitute for instructors in teaching, but can also automatically grade assignments and exams [44, 45]. Grading student assignments and exams is a routine task for instructors. The lengthy grading process often leads to instructor fatigue, which increases the likelihood of errors after a long period of time. Image recognition technology can help relieve instructors from the heavy workload of grading student assignments and exam results [38] with a low error rate.

The results of the comparison between students' conceptual understanding tests using A Hybrid Automatic Scoring System and manual scoring reveal several advantages of A Hybrid Automatic Scoring System. The scoring of the essay test can be done automatically, reducing the time needed to manually score students' responses. This can increase the efficiency of the assessment process. In addition, the A Hybrid Automatic Scoring System tends to provide consistent scores that are unaffected by emotional factors or fatigue. This helps eliminate the variability in scoring that can occur with manual methods. Using the A Hybrid Automatic Scoring System, essay test scoring can be automated for large numbers of students in a relatively short period of time. This helps address scalability challenges that can be difficult to achieve with manual methods.

An interesting finding from the implementation of A Hybrid Automatic Scoring System is a slight difference in scores for specific misconceptions and lack of understanding of concepts. This finding indicates that there are still weaknesses in the development of the system, namely that the developed A Hybrid Automatic Scoring System cannot provide a more in-depth analysis of students' responses to specific misconceptions and lack of understanding of concepts. This is because there are patterns that the system cannot identify in specific misconception indicators.

Similar issues were also identified in previous research, such as research using the stemming algorithm by Nazief and Adriani [29] with a high accuracy rate of 90.66% using keywords. However, it has a drawback where the assessment system does not consider words with the same meaning (synonyms), and it does not consider the provision of dynamic alternative answer keys as a solution to improve the system's performance in providing assessments. Research using the K-Nearest Neighbor method with an accuracy rate of 86.51% [46]. This study only considers the similarity of 44 documents and does not display keywords as search terms for weighting.

VI. Conclusion

This study provides information that the A Hybrid Automatic Scoring System is suitable for use as an assessment tool for essay responses related to conceptual understanding. This is supported by the results of validation and small-scale testing. The expert validation results indicate that the A Hybrid Automatic Scoring System is suitable for overall use, with an average score of 4.40 (highly suitable category) for software engineering and visual communication aspects. Furthermore, the analysis of the students' conceptual understanding after being assessed with the A Hybrid Automatic Scoring System shows that only 21.9% of the students have a good understanding of the concept. The majority, 40.6% of the students, are classified as having a partial understanding. In addition, 9.4% of students are categorized as having specific misconceptions and 28.2% do not understand the concept. The A Hybrid Automatic Scoring

System can be an alternative, more effective, and efficient tool for scoring essay response.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Wawan Kurniawan and Cicyn Riantoni developed the A Hybrid Automatic Scoring System, while Neneng Lestari conducted the implementation process within the classroom. Data analysis was carried out collectively by Doni Ropawandi. All authors had approved the final version.

Funding

This research was funded by a grant from the Department for Research and Community Service, Universitas Jambi.

Acknowledgment

The researchers would like to express their gratitude to the Department for Research and Community Service for funding this research. They also extend thanks to the lecturers and students of the Physics Education Program who have contributed to the success of this research.

References

[1] J. L. Docktor, N. E. Strand, J. P. Mestre, and B. H. Ross, "Conceptual problem solving in high school physics," *Phys. Rev. Spec. Top.—Phys. Educ. Res.*, vol. 11, no. 2, pp. 1–13, 2015. doi: 10.1103/PhysRevSTPER.11.020106

[2] J. L. Docktor and J. P. Mestre, "Synthesis of discipline-based education research in physics," *Phys. Rev. Spec. Top.—Phys. Educ. Res.*, vol. 10, no. 2, pp. 1–58, 2014. doi: 10.1103/PhysRevSTPER.10.020119

[3] L. M. Goodhew, A. D. Robertson, P. R. L. Heron, and R. E. Scherr, "Student conceptual resources for understanding mechanical wave propagation," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 2, 20127, 2019. doi: 10.1103/PhysRevPhysEducRes.15.020127

[4] M. G. Nugraha, I. Kaniawati, D. Rusdiana, and K. H. Kirana, "Combination of inquiry learning model and computer simulation to improve mastery concept and the correlation with critical thinking skills (CTS)," *AIP Conf. Proc.*, vol. 1708, no. 2016, 2016. doi: 10.1063/1.4941181

[5] N. Nehru, W. Kurniawan, and C. Riantoni, "Exploration of students' problem-solving skills in physics-based on expert and novice categories," no. 4, 2020. doi: 10.4108/eai.12-10-2019.2296469

[6] C. Riantoni, L. Yuliati, N. Mufti, and N. Nehru, "Problem solving approach in electrical energy and power on students as physics teacher candidates," *J. Pendidik. IPA Indones.*, vol. 6, no. 1, pp. 55–62, 2017. doi: 10.15294/jpii.v6i1.8293

[7] sutopo, "Students' understanding of fundamental concepts of mechanical wave," *J. Pendidik. Fis. Indones.*, vol. 12, no. 1, pp. 41–53, 2016. doi: 10.15294/jpfi.v12i1.3804

[8] E. Etkina and G. Planinsic, "Defining and developing 'critical thinking' through devising and testing multiple explanations of the same phenomenon," *Phys. Teach.*, vol. 53, no. 7, pp. 432–437, 2015. doi: 10.1119/1.4931014

[9] E. Campos, E. Hernandez, P. Barniol, and G. Zavala, "Phenomenographic analysis and comparison of students' conceptual understanding of electric and magnetic fields and the principle of superposition," *Phys. Rev. Phys. Educ. Res.*, vol. 17, no. 2, 20117, 2021. doi: 10.1103/PhysRevPhysEducRes.17.020117

[10] T. M. Scaife and A. F. Heckler, "Interference between electric and magnetic concepts in introductory physics," *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 7, no. 1, pp. 1–11, 2011. doi: 10.1103/PhysRevSTPER.7.010104

[11] R. Dai, J. C. Fritchman, Q. Liu, Y. Xiao, H. Yu, and L. Bao, "Assessment of student understanding on light interference," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 2, 20134, 2019. doi: 10.1103/PhysRevPhysEducRes.15.020134

[12] V. Mesic, K. Neumann, I. Aviani, *et al.*, "Measuring students' conceptual understanding of wave optics: A Rasch modeling approach," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 1, pp. 1–20, 2019. doi: 10.1103/PhysRevPhysEducRes.15.010115

[13] Nehru, C. Riantoni, D. P. Rasmi, W. Kurniawan, and Iskandar, "'Knowledge in pieces' view: Conceptual understanding analysis of pre-service physics teachers on direct current resistive electrical circuits," *J. Educ. Gift. Young Sci.*, vol. 8, no. 2, pp. 723–730, 2020. doi: 10.17478/jegys.695853

[14] L. Yuliati, C. Riantoni, and N. Mufti, "Problem solving skills on direct current electricity through inquiry-based learning with PhET simulations," *Int. J. Instr.*, vol. 11, no. 4, pp. 123–138, 2018. doi: 10.12973/iji.2018.1149a

[15] I. Rahmawati, S. Sutopo, and S. Zulaikah, "Analysis of students' difficulties about rotational dynamics based on resource theory," *J. Pendidik. IPA Indones.*, vol. 6, no. 1, pp. 95–102, 2017. doi: 10.15294/jpii.v6i1.9514

[16] Maison, Darmaji, Astalini, D. A. Kurniawan, Sumaryanti, and R. Perdana, "Supporting assessment in education: E-assessment interest in physics," *Univers. J. Educ. Res.*, vol. 8, no. 1, pp. 89–97, 2020. doi: 10.13189/ujer.2020.080110

[17] J. N. Chernega and A. K. Osgood, "Bringing students into a discipline: reflections on a travel study," *Coll. Teach.*, vol. 60, no. 3, pp. 85–86, 2012. doi: 10.1080/87567555.2011.652235

[18] M. Chassignol, A. Khoroshavin, A. Klimova, and A. Bilyatdinova, "Artificial Intelligence trends in education: A narrative overview," *Procedia Comput. Sci.*, vol. 136, pp. 16–24, 2018. doi: 10.1016/j.procs.2018.08.233

[19] X. Zhai, X. Chu, C. S. Chai *et al.*, "A review of artificial intelligence (AI) in education from 2010 to 2020," *Complexity*, vol. 2021, 2021. doi: 10.1155/2021/8812542

[20] Z. Arifin, S. Sukarmin, S. Sarwanto, and D. M. Sani, "Analysis of the need to development an assessment integrated with STEM literacy," *J. Phys. Conf. Ser.*, vol. 2098, no. 1, 2021. doi: 10.1088/1742-6596/2098/1/012036

[21] A. Madsen, S. B. McKagan, E. C. Sayre, and C. A. Paul, "Resource Letter RBAI-2: Research-based assessment instruments: Beyond physics topics," *Am. J. Phys.*, vol. 87, no. 5, pp. 350–369, 2019. doi: 10.1119/1.5094139

[22] D. E. Harahap, Festiyed, and D. Djamas, "Preliminary study on development of assessment performance instruments on physics learning to improve students' critical thinking ability," *J. Phys. Conf. Ser.*, vol. 1185, no. 1, 2019. doi: 10.1088/1742-6596/1185/1/012128

[23] E. W. Burkholder, J. K. Miles, T. J. Layden, K. D. Wang, A. V Fritz, and C. E. Wieman, "Template for teaching and assessment of problem solving in introductory physics," *Phys. Rev. Phys. Educ. Res.*, vol. 16, no. 1, 2020. doi: 10.1103/PHYSREVPHYSEDUCRES.16.010123

[24] R. Efendi, N. Y. Rustaman, and I. Kaniawati, "Self-perceived assessment skill of prospective physics teachers," *J. Phys. Conf. Ser.*, vol. 1013, no. 1, pp. 2–6, 2018. doi: 10.1088/1742-6596/1013/1/012047

[25] Maison *et al.*, "Science process skill in science program higher education," *Univers. J. Educ. Res.*, vol. 8, no. 2, pp. 652–661, 2020. doi: 10.13189/ujer.2020.080238

[26] C. Guerra, A. Moreira, and R. M. Vieira, "A design framework for science teachers' technological pedagogical content knowledge development," *Adv. Intell. Syst. Comput.*, vol. 621, pp. 193–203, 2018. doi: 10.1007/978-3-319-61121-1.17

[27] Y.-S. Hsu, T.-L. Lai, and W.-H. Hsu, "A design model of distributed scaffolding for inquiry-based learning," *Res. Sci. Educ.*, vol. 45, no. 2, pp. 241–273, Apr. 2015. doi: 10.1007/s11165-014-9421-2

[28] D. Persano Adorno, N. Pizzolato, and C. Fazio, "Long term stability of learning outcomes in undergraduates after an open-inquiry instruction on thermal science," *Phys. Rev. Phys. Educ. Res.*, vol. 14, no. 1, 2018. doi: 10.1103/PhysRevPhysEducRes.14.010108

[29] R. Ahmad and R. R. O. Sasue, "Automatic essay grading system using Nazief and Adriani's stemming algorithm," *J. Teknol. Transp. dan*, vol. 1, no. 2, pp. 101–108, 2020. https://jurnal.poltradabali.ac.id/jttl/article/view/15

[30] K. Yacef, "Intelligent teaching assistant systems," in *Proc. International Conference on Computers in Education, ICCE 2002*, 2002, pp. 136–140. doi: 10.1109/CIE.2002.1185885

[31] J. Mcniff and A. Taket, "Book reviews," January 2015, pp. 37–41, 2011. doi: 10.1080/09650790200200179

[32] V. Lampos, J. Mintz, and X. Qu, "An artificial intelligence approach for selecting effective teacher communication strategies in autism education," *npj Sci. Learn.*, vol. 6, no. 1, 2021. doi: 10.1038/s41539-021-00102-x

[33] H. D. Pfeiffer, D. I. Ignatov, J. Poelmans, and G. Nagarjuna, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 7735 LNCS. 2013. doi: 10.1007/978-3-642-35786-2

[34] L. Bin, L. Jun, Y. Jian-Min, and Z. Qiao-Ming, "Automated essay scoring using the KNN algorithm," in *Proc. Int. Conf. Comput. Sci. Softw. Eng. CSSE 2008*, vol. 1, 2008, pp. 735–738. doi: 10.1109/CSSE.2008.623

[35] T. Hamid, M. Chhabra, K. Ravulakollu, P. Singh, S. Dalal, and R. Dewan, "A review on artificial intelligence in orthopaedics," in *Proc. 2022 9th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2022*, 2022, pp. 365–369. doi: 10.23919/INDIACom54597.2022.9763178

[36] C. P. Rose and O. Ferschke, "Technology support for discussion based learning : from computer supported collaborative learning to the future of massive open online courses," *Int. J. Artif. Intell. Educ.*, pp. 660–678, 2016. doi: 10.1007/s40593-016-0107-y

[37] H. S. Lee and J. Lee, "Applying artificial intelligence in physical education and future perspectives," *Sustain.*, vol. 13, no. 1, pp. 1–16, 2021. doi: 10.3390/su13010351

[38] Y.-J. An and L. Cao, "Examining the effects of metacognitive scaffolding on students' design problem solving and metacognitive skills in an online environment," *J. Online Learn. Teach.*, vol. 10, no. 4, pp. 552–568, 2014.

[39] A. G. Spatioti, I. Kazanidis, and J. Pange, "A comparative study of the addie instructional design model in distance education," *Inf.*, vol. 13, no. 9, pp. 1–20, 2022. doi: 10.3390/info13090402

[40] R. M. Branch, *Approach, Instructional Design: The ADDIE*, vol. 53, no. 9. 2009.

[41] A. M. Almelhi, "Effectiveness of the ADDIE Model within an e-learning environment in developing creative writing in efl students," *English Lang. Teach.*, vol. 14, no. 2, 2021. doi: 10.5539/elt.v14n2p20

[42] N. Chidayati, I. W. Distrik, and A. Abdurrahman, "Improving students' higher order thinking skill with stem-oriented E-module," *Indones. J. Sci. Math. Educ.*, vol. 4, no. 3, pp. 274–286, 2021. doi: 10.24042/ijsme.v4i3.9930

[43] I. T. Sanusi, S. S. Oyelere, H. Vartiainen, J. Suhonen, and M. Tukiainen, "A systematic review of teaching and learning machine learning in K-12 education," *no. Ml. Springer US*, 2022. doi: 10.1007/s10639-022-11416-7

[44] Z. Kanetaki, C. Stergiou, G. Bekas, C. Troussas, and C. Sgouropoulou, "A hybrid machine learning model for grade prediction in online engineering education," *Int. J. Eng. Pedagog.*, vol. 12, no. 3, pp. 4–23, 2022. doi: 10.3991/IJEP.V12I3.23873

[45] H. Horanai, Y. Maejima, and L. Ding, "An education tool at supports junior learners in studying machine learning," *Front. Artif. Intell. Appl.*, vol. 360, pp. 111–116, 2022. doi: 10.3233/FAIA220432

[46] T. Thongyoo, S. Saelee, and S. Krootjohn, *SNRU Journal of Science and Technology*, pp. 87–95, 2018.