An Indonesian Case Study on the Integration of Real-time Feedback into Science Practicum Performance Assessments for Measuring Critical Thinking Skills in Universities

Sudirman Sudirman[1,4], Agus Ramdani[1], Aris Doyan[2], and Yunita Arian Sani Anwar[3]

[1] Postgraduate Science Education Doctoral Program at Indonesia's University of Mataram
[2] Study Program for Physics Education, University of Mataram, Indonesia
[3] Study Program for Chemistry Education, University of Mataram, Indonesia
[4] Study Program for Science Education, University of Qamarul Huda Badaruddin Bagu, Indonesia
sudirmansiddiq@gmail.com

Abstract. This research aims to collect initial data for the creation of critical thinking instrument that incorporate real-time feedback on performance assessments in scientific practicums at two state universities and two private universities in Indonesia. The data were analyzed descriptively through a chi-square test, employing a qualitative case study methodology alongside cluster random sampling techniques. Data was collected through the utilization of questionnaires, in-depth interviews, and observations. The research sample comprised 376 university students and lecturers, alongside comprehensive interviews conducted with 4 associate deans, 4 program directors, 4 laboratory directors, and 4 teaching assistants from four distinct universities. The findings indicate that there was an absence of an evaluation regarding performance critical thinking due to the lack of an appropriate instrument. The assessment results indicate public and private universities do not differ significantly ($p > 0.05$) regarding the measurement of critical thinking incorporates immediate feedback. It appears that neither of the categories of institution has implemented critical thinking performance assessments that incorporate real-time feedback within the context of science practicum. Lecturers' and students' responses to the implementation of critical thinking performance assessments, which are incorporated with immediate feedback during the science practicum, there was no notable difference ($p > 0.05$). This suggests that university students share a similar experience with lecturers in evaluating critical thinking performance with feedback in the context of science practicum. This research proposes the development of assessment tools for critical thinking performance that offer immediate feedback during science practicum activities.

Keywords: Automatic Feedback, Critical Thinking, Real-Time, Performance Assessment, Science Practicum.

D. S. Elmiana et al. (eds.), *Proceedings of the 6th International Conference on Education and Social Sciences (ICESS 2024)*, Advances in Social Science, Education and Humanities Research 921,
https://doi.org/10.2991/978-2-38476-392-4_12

1 Introduction

The importance of possessing the capability to comprehend and effectively convey meaning through various textual formats, commonly referred to as critical thinking, is widely acknowledged in contemporary society. This proficiency is deemed essential for active engagement in the information-driven realm and for enhancing academic performance. Therefore, providing timely assistance to students experiencing difficulties in critical thinking acquisition is of paramount importance. While the agenda at hand may appear straightforward, it is essential to employ reliable and valid assessment methods in order to facilitate error recognition and improvement among students, as well as to provide lecturers with valuable information for appropriate interventions.

Performance assessment is a pedagogical approach utilized to evaluate student conduct, wherein data is systematically collected through observational means to inform decision-making regarding an individual [1,2]. The process under consideration is of prolonged duration and involves the utilization of diverse tools and methodologies, alongside a methodical approach of firsthand observation. Performance assessment encompasses several key components, including the incorporation of numerous criteria [3], the establishment of predetermined quality standards [4], and the inclusion of subjective (opinion-based) evaluation [5]. The procedure is characterized by its extensive duration and requirement of significant time investment, encompassing the utilization of progress charts, work sample exams, and portfolios [6]. Performance-based education systems necessitate the utilization of several forms of evidence, including but not limited to evidence gathering, real-life tasks, oral questions, written evaluations, projects, case studies, and field observation [7]. The process of conducting a performance review entails several key steps. Firstly, it is necessary to establish the objective of the evaluation. This involves clearly defining the purpose for which the review is being conducted [8][9]. Secondly, one must identify the specific concepts, knowledge, and abilities that will be assessed during the review. These elements serve as the criteria against which the individual's performance will be evaluated [10]. Thirdly, it is important to establish the desired level of achievement that the individual should strive for. This sets the benchmark against which their performance will be measured [11]. Finally, one must select the appropriate type of activity or assessment method to carry out the performance review. These steps collectively contribute to the effective execution of a performance review process [12].

A comprehensive understanding of performance assessment can be derived from the following definition: it refers to the systematic collection of activity-based data from an individual who is being evaluated, employing several procedures and tools to ascertain the individual's state [13]. The process of conducting a performance review is often characterized by its extensive duration and the significant investment of time it requires. Performance evaluation devices commonly utilized in educational settings encompass several tools such as progress charts [9,14], work sample tests [5], and portfolios [8]. Progress reports commonly consist of graphical representations that contain students' names, daily assignments, and test scores. The utilization of this graphical progress record does not permit a direct assessment of students' abilities. The aforementioned recording holds significant utility for educators in evaluating the progression of the learning experience [7].

A work sample test refers to an evaluative setting wherein the individual under scrutiny is required to do one or more authentic job-related tasks. A portfolio refers to an assemblage of student work that has been compiled over a specific period of time. The utilization of this portfolio proves to be highly advantageous for educators as it facilitates the systematic monitoring of students' aptitude progression on a consistent basis [1,15]. Competence-based education systems necessitate the inclusion of multiple forms of evidence, referred to as sources of evidence, to demonstrate that a student has achieved predetermined competence levels within a specific timeframe. As stated by the source cited [16], evidence within the context of competency-based education encompasses several forms, such as observation and testimony [15].

Practice of evaluating and providing timely feedback on performance or progress in a given context. This approach involves the use of immediate and ongoing assessment methods to gather data. Online assessments provide immediate feedback that is automated and prompt, thereby increasing students' comprehension and communication skills. According to [17], this feedback has the potential to assist students who have difficulties with course content or exhibit reticence or reluctance in reaching out to instructors directly. The provision of online feedback additionally enhances group comprehension and fosters a more dynamic and engaging learning experience. The learning outcomes of students are influenced by the quality and level of detail in the feedback they receive. Research studies have shown that when students are supplied with high-quality and detailed feedback, their learning is enhanced and consolidated [11,12].

The utilization of digital technology possesses the capacity to enhance the design of instructional materials, empower students, and facilitate transformative changes within the educational system. As stated by [10], the implementation of this technology has the capacity to enhance adaptive learning, provide recommendations for student-centered learning materials, and identify areas where learning deficiencies are present.. Moreover, the advancement of technology, the improvement of knowledge delivery platforms, and the evolution of the instructor-student relationship have the capacity to drive transformative changes in education [1,12,13,15]. Nevertheless, as stated by the source [4], Attaining a comprehensive examination of intricate social, educational, and environmental factors is necessary to ensure that scientific education is of the highest quality. The process of choosing and effectively integrating digital technology into science education to accommodate various factors is a significant obstacle. This detailed analysis explores the integration of digital technology across various domains of science education [1,9].

Several domains of higher education have made use of digital technology, such as instructional methodologies for mathematics, science, and tutoring. This approach has been implemented in various fields, such as feedback-driven programming education [3,6,7,18]. However, as stated by [1,15], few literature evaluations have addressed the integration of digital technology into science education. Digital technology facilitates the automation of student performance assessments and the creation of questions or assignments for educators. Such automated systems have the potential to make science learning more accessible and efficient for both teachers and students. Science teachers can save a lot of time and effort with the help of automated scoring, argument evaluation systems, and question generation technologies [2,13].

Digital technology integration in science education has been used to enhance learning outcomes and instructional strategy creation, implementation, and assessment. Its use in this field focuses on developing Android-based platforms, to support and enhance the teaching and learning of scientific concepts. Digital technology offers numerous benefits, including automated assessments and learning analytics in scientific education [11,12,19]. According to Walters et al. (2017), educational research indicates that there is significant potential for utilizing performance assessment to support teaching and learning including critical thinking [16]. Additionally, Hattie and Timperley (2007) emphasize the importance of providing instant feedback in the educational context, highlighting its potential for enhancing the learning process [20]. The utilization of technology-based assessment and automated text scoring approaches in the field of assessment has facilitated the provision of quick feedback, reducing the time required for evaluation and minimizing the influence of human bias.

Based on the above formulation, critical thinking assessment can be integrated with feedback automatedly among faculty members and colleges students. To evaluate this hypothesis, a study must be undertaken utilizing the following research questions:

1. How is the evaluation of science practicums at these four colleges currently?
2. Does the evaluation of critical thinking differ statistically significantly depending on the types of institutions (public and private) and respondents' responses (students and lecturers)??

2 Article Review

2.1 Performance Assessment and Feedback

Throughout history, assessment has served several purposes, such as collecting pertinent data on student performance or advancement, as well as evaluating student interests in order to make informed judgments about their educational journey. Teachers have the ability to modify their instructional methods based on student requirements by utilizing data obtained from assessments. The utilization of assessments in the educational setting can serve as a catalyst for motivating students to engage in the learning process. Furthermore, in recent times, governments worldwide have been utilizing assessment data obtained from national or large-scale international testing initiatives to assess the quality of education and enact reforms in their respective national educational frameworks [21].

In the conventional approach, summative assessment, also known as the evaluation of learning [22-24], is employed to gather crucial data and information to inform decision-making processes, in order to fulfill the aforementioned objectives. According to previous research conducted by [25], it has been found that this particular approach has a limited impact on students' learning experiences. Therefore, if the primary goal is to enhance learning outcomes, it is imperative to shift the purpose of assessment from evaluating learning to facilitating learning. This shift entails a focus on providing constructive feedback and fostering the student's ability to engage in self-assessment and reflection on their own learning process [26-28].

The use of summative assessment for enhancing learning is associated with several significant shortcomings. Firstly, there is a considerable time gap between the administration of the test and the provision of feedback, which hampers the timely identification of areas requiring improvement. Secondly, the utility of summative test results for planning interventions is limited. This is highlighted by [29]. Thirdly, the phenomenon of washback occurs, wherein teachers may adopt a teaching approach that specifically caters to test preparation, or learners may modify their learning strategies in response to the announced assessment. This observation is made by [30]. Lastly, the use of summative assessment can have adverse effects on the overall school climate and contribute to increased stress levels among teachers.

The utilization of regular and prompt feedback obtained from formative [31] and diagnostic assessment [32] is widely recognized as a very effective strategy for enhancing the learning process [20]. While formative assessment is widely acknowledged as an essential tool for gathering information to improve teaching and learning in educational research, a major challenge in its implementation stems from the substantial commitment it demands. This presents a particular challenge for classroom teachers, who are frequently overwhelmed with a multitude of responsibilities and may find it impractical to develop comprehensive formative assessment strategies [26]. Diagnostic evaluation is commonly perceived as a trade-off between the necessity for prompt feedback and the additional burden it places on teachers.

The distinction between formative and diagnostic assessments lacks clarity. Diagnostic assessment is commonly defined in the literature as an evaluative process that centers on identifying and analyzing problems, investigating potential challenges, evaluating students' readiness for a learning endeavor, and gauging their foundational knowledge [29]. Diagnostic assessment is a helpful tool for educators as it offers insights into students' existing knowledge and areas of deficiencies. This information is crucial for the development of targeted interventions aimed at supporting students who are facing challenges in their learning journey. Simultaneously, students are provided with feedback on their performance, enabling them to actively evaluate their own learning progress [16].

Prior research has indicated that the efficacy of feedback may be assessed based on various criteria, including the attributes of the feedback itself, the attributes of the learners, and the characteristics of the tasks [33]. Previous research has examined the timeliness and content of feedback in relation to its features. When considering the aspect of timing, feedback can be categorized as either immediate or delayed. Research suggests that quick feedback is more advantageous than delayed feedback [34], as it enables students to swiftly identify and comprehend conceptual and procedural errors [18,16].

The feedback may include two components: first, delivering clear information about the correctness or incorrectness of the answer, and second is providing additional insights into the student's overall performance on the test or in relation to a specific answer. Research indicates that the provision of expanded feedback in educational settings has been found to be more efficacious and can substantially enhance the process of learning [20]. Furthermore, the research conducted by [35] the effectiveness of detailed feedback in promoting learning is improved when it is customized to the unique context of each activity, commonly referred to as contextualized feedback. Multiple research studies have indicated that effective and valuable feedback should consist of three essential components. Firstly, It ought to fairly evaluate the student's actual performance, which is defined as their present-day knowledge or skills [13]. Secondly, the areas that require improvement for the learner should be highlighted, known as their desirable performance [36]. Finally, it should provide recommendations or guidance on how the student can attain the desired level of performance [16,17]. As a result of the information technology revolution in education, technology-based assessments, such as diagnostic and formative assessments, have been acknowledged as an important tool to supplement the regular instructional process. Students rely on these evaluations for valuable information and feedback, which helps them track their progress.

2.2 Automated Assessment

Historically, technology-driven evaluation has been regarded as a means of offering prompt feedback to students. However, prior research indicates that this assertion is not entirely accurate, as the evaluation of created response items remains a complex and time-intensive procedure, even when employing technological advancements [2]. Over the past decade, progress in machine learning has facilitated the development and integration of automated text scoring in technology-based assessments, enabling the evaluation of multi-sentence text responses [3].

One new possibility for technology-driven review is automated text scoring, which can score long text replies in real-time while also reducing the biases and mistakes that come with human grading, and streamline the allocation of time, financial resources, and effort for educators [18,37,38]. Automated text scoring is a technique used to assess the quality of a student's response to a predefined question or item. This evaluation involves assigning a score to the answer based on a predetermined scale. Following the allocation of a numerical score to a given response, immediate feedback is given to the student using the predetermined scoring criteria or framework. When evaluating the substance of answers, automatic text scoring is typically used [18,19].

Recent research in automated text scoring has demonstrated the efficacy of this strategy in delivering prompt formative feedback to students and bolstering their scientific reasoning [35,39]. Recent research indicates that the most effective model of a technology-based assessment (TBA) system, this system covers the entire duration of primary education, spanning from the initial stages of schooling to the culmination of the six-year program. The integrated assessment system described by [19,20] encompasses various stages of the assessment process, including item generation, test administration, automated scoring, data analysis, and feedback provision.

3 Methodology

3.1 Research Design

This study used the cluster random sampling technique and a case study method, which combines quantitative and qualitative approaches.

3.2 Population and Sample

Participants were chosen from four different colleges using cluster random selection. Two governmental universities, Mataram University and Mataram State Islamic University, and two private universities, Universitas Pendidikan Mandalika and Universitas Qamarul Huda Badaruddin Bagu, were among these educational establishments. Two hundred sixty-one college students and one hundred fifty-six faculty members make up the study population. There are 275 females and 101 males among them. The online survey was supplemented by a dataset that included in-depth interviews with four academic vice deans, four study program directors, four laboratory heads, and four assistant lecturers to confirm the data findings and round out the dataset.

3.3 The demographic characteristics of the respondents

According to the data presented in Table 1, the proportion of respondents who were university students was greater (69.4%) compared to lecturers (30.6%). It is worth noting that both the students and lecturers included in the study were affiliated with the Science study program. Additionally, the student respondents were specifically those who had previously participated in a science practicum. This phenomenon is observable within the demographic of individuals between the ages of 19 and 25, with a majority occurring at the age of 21. Specifically, 30% of the total population accounts for 69.4%, indicating that approximately half of the population falls within the age range above 21 years. It can be inferred that individuals of this age have completed four semesters, indicating their involvement in practicum activities, rendering them eligible for inclusion as participants in this study. Furthermore, among the demographic of individuals aged 19-25, it can be inferred that their proficiency in utilizing technology is greater compared to older age groups. Consequently, this enhanced technological aptitude facilitates a more seamless integration of technology in the process of educational evaluation. The proportion of females (73.1%) exceeded that of boys (26.9%), potentially attributable to a higher number of female students being exposed to internet material regarding the dissemination of this questionnaire. The University of Mataram (UNRAM) had the highest proportion of responses, accounting for 31.4% of the total. This indicates that UNRAM has the largest student population among the universities in West Nusa Tenggara, particularly in the field of science education. In addition, UNRAM is regarded as one of the preferred universities in the Eastern region of Indonesia.

Table 1. Demographic Characteristics of respondent (N=376)

Demographic Data	N	%	Mean	SD
Respondent			1.69	0.461
University Students	261	69.4		
Lecturers	115	30.6		
Sex			1.73	0.444
Male	101	26.9		
Female	275	73.1		
Age			26.03	8.350
19-25	261	69.4		
26-35	41	10.9		
36-45	68	18.1		
46-55	6	1.6		
College			2.41	1.190
Mataram University	118	31.4		
Mataram State Islamic University	87	23.1		
Mandalika University of Education	69	18.4		
Qamarul Huda Badaruddin Bagu University	102	27.1		
In-depth Interview				
Vice Dean of Academic	4	25.0		
Head of the study program	4	25.0		
Head of Laboratory	4	25.0		
Assistant Lecturer	4	25.0		

3.4 Research Instruments

The instrument comprises 3 components, specifically the initial component pertaining to the profiles of the participants. The subsequent section pertains to the provision of evidence for the assessment of critical thinking on science practicum, including a total of eighteen question items. The tools employs a Likert scale as well as a set of questions that require binary yes/no responses.

3.5 Validity and Reliability Instruments

The study instrument was tested on a sample of 207 individuals, which included university students and lecturers associated with the University of Qamarul Huda Badaruddin Bagu. The item validity study indicated that three items were classified as invalid according to the Pearson correlation test, which was conducted at a significance level of $p < 0.05$. As a result, these items were removed from further examination. Question items with a validity score exceeding 0.8 were categorized into the high validity group, signifying their appropriateness for use as research instruments. The results of the reliability analysis performed on all items substantiate this claim, as demonstrated by the application of Cronbach's Alpha. The mean value of the alpha test scale is greater than 0.9, which signifies a strong degree of internal consistency and reliability. As a result, utilizing this instrument for research purposes is highly feasible.

3.6 Data Collection

The data collection process involved the utilization of an online questionnaire survey administered through a Google form, complemented by in-depth interviews and direct practicum observation.

3.7 Data Analysis

The Kruskal-Wallis statistical analysis method was employed in this study because to its suitability for analyzing unpaired categorical data, specifically, it was applied to analyze the data on more than 2 categorical data of dependent variable.

4 Results

4.1 What is the current evaluation process for science practicums at these four colleges?

The results of the study indicated a notable resemblance in the assessment approaches utilized for practicum in both public and private colleges, as seen in Table 2.

Table 2. Data pertaining to the assessment of scientific practicum

No	Data description	N	%	Mean (M)	SD
1	The practicum evaluation is carried out manually.	203	54.0	2.56	2.098
2	There has been no real-time feedback on the practicum assessment.	281	74.7	0.25	0.435
3	Agreed, if the practicum assessment process is based on performance achievement	272	72.3	3.05	0.562
4	Agree critical thinking must be assessed on science practicum	261	69.4	3.19	0.557

5	Agree real-time performance assessment practicum using the application:			2.40	0.632
	a. Android and Web	179	47.6		
	b. Android	167	44.4		
	c. Web	30	8.0		
6	Agree real-time practicum assessment can be accessed through:			2.08	0.971
	a. All devices	193	51.3		
	b. Handphone	163	43.4		
	c. Computer/Laptop	20	5.3		

According to the data shown in Table 2. The majority of respondents currently choose manual assessment methods when it comes to the science practicum (54.0%, M = 2.56, SD = 2.098) indicating this approach. Furthermore, the assessment is not performed in real time and does not provide feedback, as indicated by a considerable percentage of participants (74.7%, M = 0.25, SD = 0.435). Additionally, there is a consensus among respondents (72.3%, M = 3.05, SD = 0.562) that the evaluation of the practicum should be based on the achievement of performance processes. Moreover, a majority of participants (69.4%, M = 3.19, SD = 0.557) expressed agreement with the notion that critical thinking assessment should be integrated into the science practice process. The results align with the data collected during the practice period and in-depth interviews (see Figure 1) conducted in four universities (two state universities and two private universities). Specifically, the findings indicate that there was no performance assessment during the practice, including a lack of critical thinking assessment. Additionally, assessments were not accompanied by feedback. The evaluation process comprised an initial test conducted before the practice began and a conclusive assessment carried out after the practice ended. The lack of a validated and tested standardized assessment instrument is the primary reason assessments are not performed during practice.

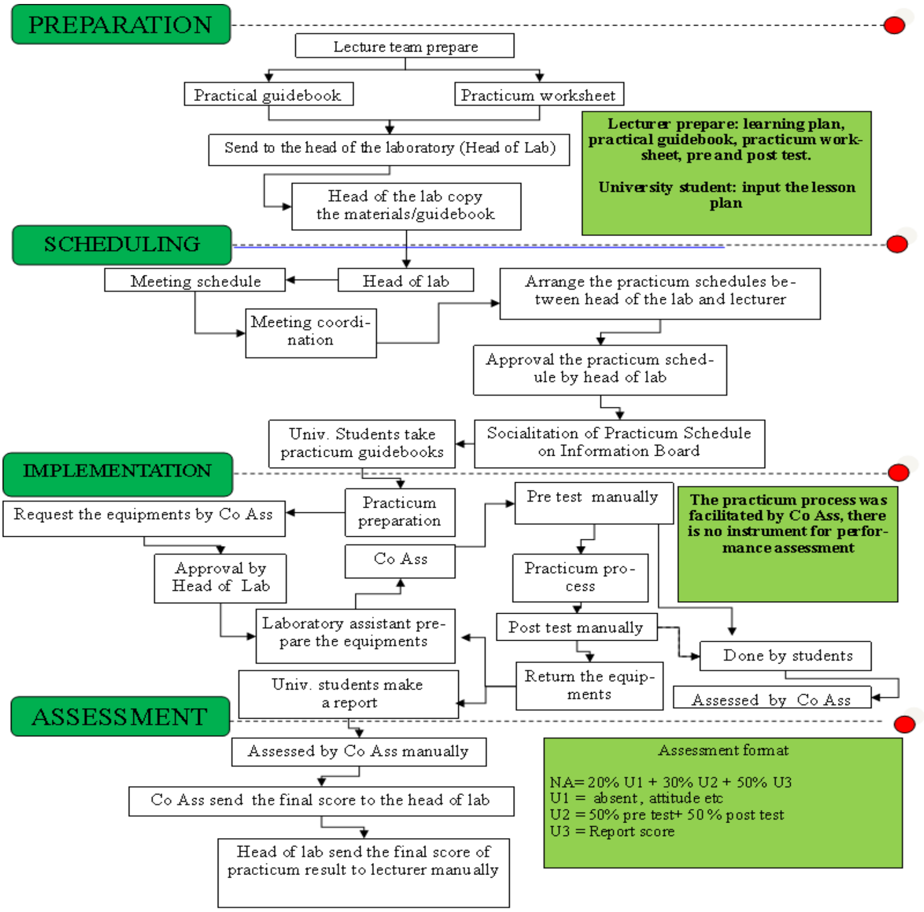

Fig. 1. The current protocols employed in the scientific practicum

4.2 Does the evaluation of critical thinking differ statistically significantly depending on the types of institutions (public and private) and respondents' responses (students and lecturers)?

In accordance with the data shown in Table 3. The findings indicate that a significant proportion of participants (N = 307, M = 1.70, SD = 0.459) reported not having conducted critical thinking assessments during science practicum. Similarly, a substantial number of respondents (n = 317, M = 5.46, SD = 0.449) indicated that they had not administered any critical thinking measurements during the implementation of practicum in both public and private universities.

Table 3. Critical thinking measurement based on respondent and university variables

Variables	Critical Thinking Measurement	N	Mean (M)	SD	Sig. Kruskal – Wallis
Respondent	Never	307	1.70	0.459	0.140
	Seldom	52	1.71	0.457	
	Often	17	1.53	0.514	
University	Never	317	5.46	0.449	0.383
	Seldom	48	5.42	0.457	
	Often	11	5.41	0.507	

Additionally, according to the results of the Kruskal-Wallis test (see Table 4), no statistically significant difference was found ($p > 0.05$) in the responses of lecturers and students towards the measurements of critical thinking. This indicates that both groups, shared a similar perception that the majority of critical thinking assessments had not been implemented during the science practicum (refer to Figure 2). This is due to the absence of appropriate and reliable instruments for measuring critical thinking throughout the time of the study. Confirmation of this conclusion came from in-depth interviews with the academic dean, heads of study programs, laboratory heads, and assistant lecturer, which showed that these people used only non-standardized instruments to evaluate the practicum.

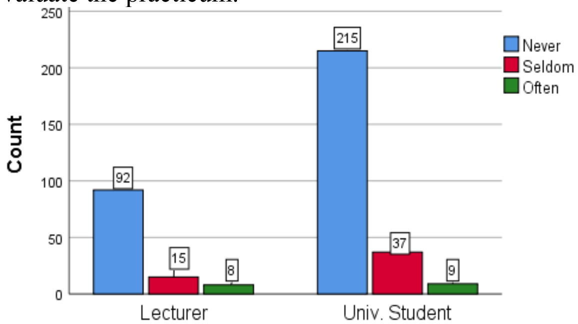

Fig. 2. Response from the Lecturer and Student on Critical Thinking Measurement

Table 4. Critical Thinking Assessment Using Kruskal-Wallis Statistical Analysis

Total N	376
Test Statistic	3.931
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.140

Table 3 presents evidence indicating that the four universities did not conduct critical thinking assessments during the implementation of the science practicum. The results of a statistical test (Table 5) further support this finding, indicating no statistically significant difference ($p > 0.05$) between the type of university and the measurement of critical thinking. Most (Figure 3), at public universities (N=165) and private institutions (N=142) reported that critical thinking measurements had not been undertaken at the time of the practice. The reliance on practicum observations is a result of the absence of assessment instruments during the practice period. Informants from four colleges participated in in-depth interviews to make up for this constraint. These interviews revealed the significance of critical thinking measurements; however, it was noted that there were insufficiently standardized and validated tools for validity and reliability. Consequently, the absence of such instruments poses challenges in ensuring accountability for the obtained results.

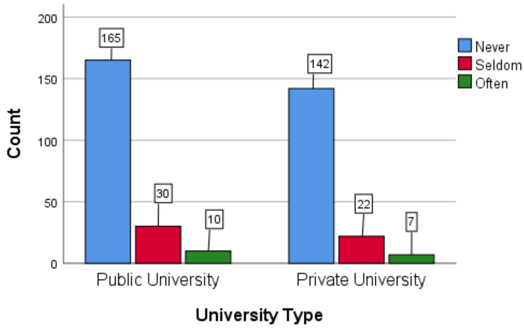

Fig. 3. Responding to a Critical Thinking Assessment in University

Table 5. A Kruskal-Wallis test of critical thinking using university-type variables

Total N	376
Test Statistic	1.921
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.383

The use of questionnaires, in-depth interviews, and observational methods all point to a significant constraint in the evaluation of the scientific practicum. This restriction relates to the absence of a thorough performance evaluation, which results from the lack of measurement tools that exhibit validity and reliability. Access should be restricted to authorized personnel only to oversee the practicum assessment. Figure 4 illustrates the schematic representation of the Integrated Realtime Performance Assessment (IRPA) model.

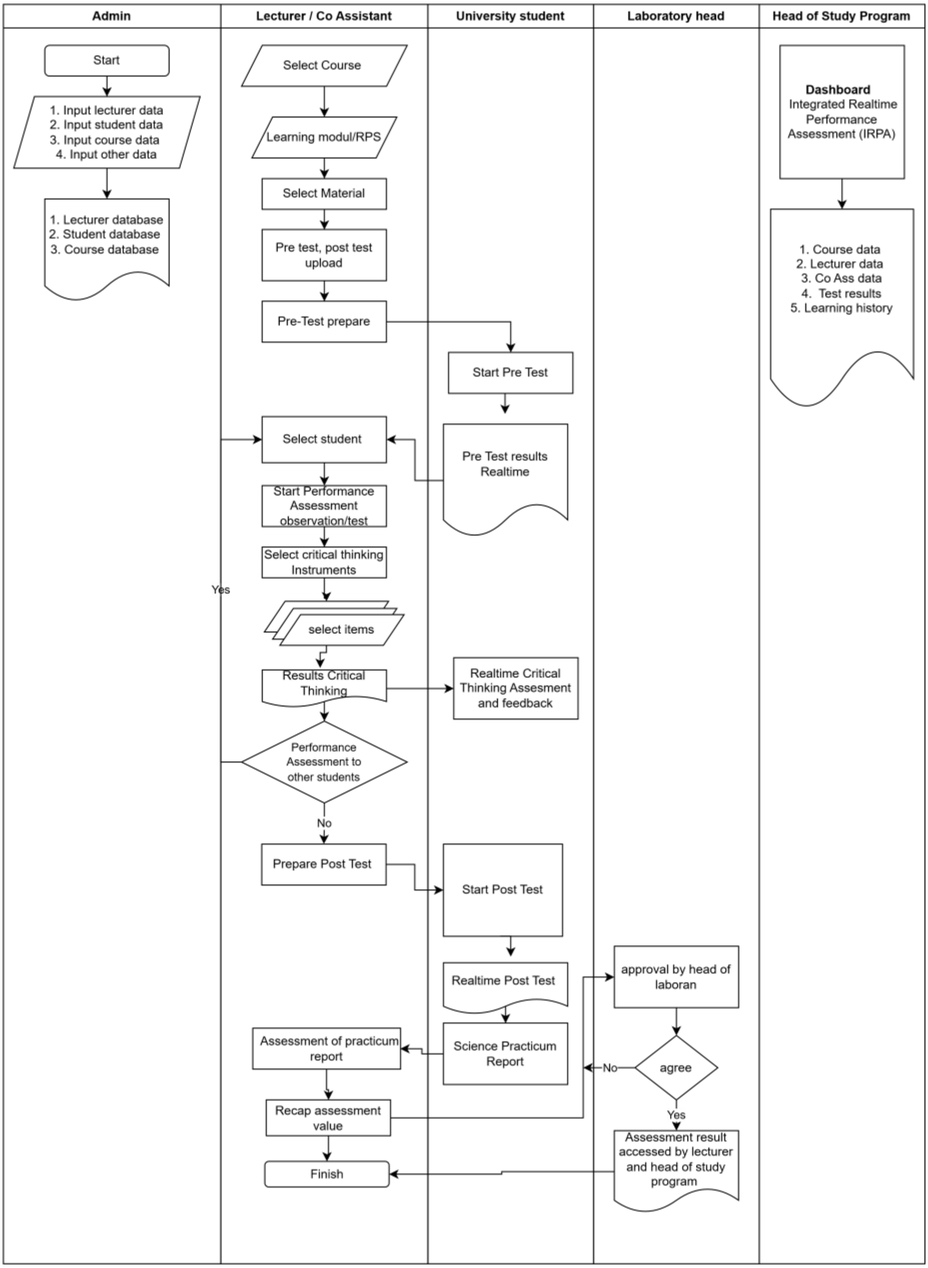

Fig. 4. Design for Automated Feedback of Realtime Critical Thinking Assessment on Science Practicum

5 Discussion

5.1 The Assessment of Current Science Practicum

According to Table 1, comprehensive interviews were carried out at 4 colleges involving management personnel. The aim of carrying out in-depth interviews is to validate findings, enhance data completeness, or gather supplementary information that has been acquired via a questionnaire. The findings from comprehensive interviews indicated that The 4 colleges practicum evaluation protocols, regardless of being public or private, exhibited minimal differences.

Teaching assistants assessed the practicum by analyzing the final report, the post-test, and the initial reaction (pre-test). The assessment findings are manually submitted to the course lecturer after being approved by the laboratory head. There is currently no performance assessment tool available, and observations indicate that the evaluation is still completed by hand. To evaluate the practicum implementation process, it is also essential to create relevant and trustworthy performance assessment tools. The absence of a performance evaluation indicates a lack of active engagement, and collaboration has occurred without the application of valid standardized instruments and reliability measures. The results align with the findings obtained from the comprehensive interviews conducted with faculty personnel.

The practicum implementation is less effective since it has not been backed by a performance assessment procedure [12-14]. The majority of assessments are conducted within a week following the practicum, in written reports because skills are a significant component of the activities that are not tracked or quantified. Therefore, in order to save time, effort, and money, we require performance evaluation tools that are digitally accessible and can be evaluated rapidly as valid, consistent, and trustworthy tools [11]. It is advised to develop a performance evaluation instrument during the practicum because lecture assistants frequently rely on their recollection while administering tests [10].

The assessment must be conducted throughout the entire practicum, beginning at the start and continuing until completion. This approach ensures a comprehensive evaluation at all stages, encompassing cognitive, psychomotor, and affective dimensions [13,32]. Furthermore, Kruit et al. (2018) additionally, performance evaluation is necessary for conducting assessments of students' attitudes and abilities in addition to their cognitive domain abilities [12]. In order to conduct this evaluation, a valid, trustworthy, and user-friendly performance assessment tool is required.

5.2 Realtime Critical Thinking Evaluation using technology

Based on the findings of the study, it is evident that both instructors and students have a preference for practitioners to participate in real-time practices that make use of digital technology, specifically through applications for Android and web browsing (47.6%) and mobile phone accessibility (43.4%). The mobile application is designed to enable real-time analysis of performance assessments.

Mobile technology provides consumers with many advantages, including the capacity to link people across time zones and places, the efficiency and responsiveness of operating systems and applications, the mobility of devices, and the improvement of social connections. A global trend in education is the extensive use of mobile applications and technologies [1]. For academics, especially those working in the field of education, mobile technology is quite appealing [2]. Mobile learning is the process of facilitating electronic learning through the use of portable devices, including computers, laptops, cellphones, audio players, and e-books [3]. Through internet-based platforms and technology breakthroughs, mobile learning efficiently removes the constraints of time and physical location while fostering collaborative learning among students [4]. Students can access course materials and engage in dynamic and captivating interactions with educational information thanks to the broad availability of mobile devices with constant connectivity.

Keeping students engaged and actively involved in effective learning through mobile devices poses a considerable challenge in the twenty-first century [14]. The significance of cultivating advanced cognitive abilities, including problem-solving and critical thinking, in students has been extensively acknowledged [15]. Mobile learning has surfaced as a groundbreaking method in the field of education [5]. Mobile learning demonstrates a beneficial influence on students and educators alike, facilitating comprehension of educational content and fostering a range of cognitive skills such as communication, problem-solving, creativity, and advanced thinking capabilities.

Technology is frequently characterized as an assemblage of instruments that offer diverse solutions to the obstacles encountered in the realm of education [17]. Education aims to cultivate individuals with extensive knowledge, creativity, proficiency in digital technologies, and the capacity to adapt to changing situations [5]. Moreover, the integration of information technology, such as the internet and multimedia systems, within educational environments seeks to enhance the quality of learning by facilitating students' access to vital resources and services [18].

Three main areas are covered in the research of higher-order thinking skills (HOTS): the capacity to apply ideas in diverse contexts, the categorization of HOTS as knowledge, and the capacity to draw connections in unexpected circumstances. The ability to comprehend logical complexities, think reflectively, and participate in arguments linked to reaching well-informed conclusions or judgments are further components of developing critical thinking abilities. Furthermore, the ability to solve problems, especially the ability to come up with original, non-traditional, and creative solutions, is a critical component of people's abilities [23].

5.3 Performance Assessment and feedback.

The statistical study revealed no statistically significant association between instructor evaluations and student performance. Similar findings were found in statistical analyses of academic achievement at both public and private colleges. The performance assessment was conducted without offering feedback. This finding was reinforced by in-depth interviews with management people and practicum assistants, as well as direct observations of the practicum itself. In order to enhance the assessment of performance and provide automated feedback that can be readily accessed by both students and teachers, employing a digital application is deemed most advantageous. According to [1,41] the automated program produces descriptive assessment reports for evaluating student performance and is utilized by researchers to assess student proficiency in critical Thinking. The researchers incorporated an extra variable into the standard analytics rubric of the application. Assessment denotes the evaluation of students' performance against defined objectives, whereas formative feedback entails providing comments and recommendations for enhancing the work to achieve the requisite criteria [14-17]. Feedback denotes the disparity between the actual values and the reference values of a system's parameters, which is subsequently utilized to modify and rectify the inconsistency in a particular manner.

Through formative assessment, the study aimed to help educators and students improve their teaching and learning. It effectively communicates performance expectations to students, provides valuable feedback that helps them understand their strengths and areas for improvement, ensures that assessments are conducted consistently and fairly, promoting equity among students, and has been shown to enhance student learning while encouraging self-evaluation [20]. Rubrics also help teachers (a) set clear goals for instruction and learning, (b) grade students according to predetermined standards and abilities, (c) communicate student achievement consistently, and (d) spot trends in the strengths and weaknesses of student work [23]. The reliability of performance appraisal ratings may be increased by using analytical rubrics that are appropriate for the topic area.

5.4 Critical Thinking Assessment on Science Practicum

The results obtained through questionnaires, in-dept interviews, and direct observations to examine the execution of practicum reveal a significant lack of evaluation of critical thinking in scientific practice. The conduction of critical thinking assessments has been impeded due to the lack of valid instruments, as indicated by the findings from interviews. Nevertheless, it is crucial to underscore the importance of carrying out these evaluations. The Critical Thinking and Science Practicum is a course specifically designed to foster the growth of critical thinking skills in the realm of scientific inquiry, encompassing the development of critical thinking abilities and process-oriented thinking skills. The development of thinking skills should be complemented by the cultivation of attitudes that arise from the process of learning [12,13]. In the context of natural sciences, particularly science education, attitudes towards scientific procedures play a crucial role in ensuring the integrity and progression of knowledge among students. These scientific attitudes can be fostered through the utilization of process skills, a scientific work procedure [14,15].

Critical thinking skills recognized as part of facilitate the development of higher-order thinking skills, inquiry [19], problem-based learning [20,21), and project-based learning [22,23] paradigms are commonly employed by researchers to enhance students' critical thinking abilities. In order to enhance their critical thinking abilities, it is imperative for students to acquire substantial experience in science teaching. The underperformance of students in science education can be attributed to the extensive content coverage and the emphasis on rote learning in schools, which leads to a limited implementation of science instruction, particularly in terms of scientific performance. The analysis suggests that the poor implementation of teaching strategies is a key factor contributing to the lack of critical thinking skills and scientific performance observed among students in science classes.

The science practicum model is a teaching approach that not only emphasizes science as a product but also as a process, particularly in enhancing students' scientific performance and critical skills. This practicum learning model is grounded in constructivist philosophy, which posits that knowledge is a product of cognitive construction through activities that develop students' scientific skills and attitudes, allowing them to meaningfully build their knowledge through real-life experiences. This teaching method focuses on concepts and engages students in problem-solving tasks, providing opportunities for them to work independently, construct their own learning, and ultimately produce valuable and realistic outcomes [7]. Practicum learning is more effective than traditional classroom instruction in enhancing learning achievement in science education [8]. This aligns with previous research, which demonstrates that practicum learning can help students' complete tasks, represent their concepts, and establish connections between different concepts [9]. In science education, the practical learning model with scientific evaluation is preferred due to its significant potential to enhance students' scientific performance, stimulate cognitive processes aimed at fostering critical thinking abilities, and cultivate a scientific mindset. This practicum learning model helps students build sophisticated and rich memory representations of their experiences, demonstrating strong connections between semantic and episodic events and actions. Moreover, it motivates students to be more active and creative in their learning, thereby improving their scientific achievements. Science practicum is an effective way to maximize student engagement in learning, enhance their scientific activity and performance, and support the development of their critical thinking skills and scientific attitudes.

6 Conclusions

The utilization of valid and trustworthy instruments for performance assessment has not been conducted. It is advisable to undertake the development of evaluation instruments that can effectively measure the process of critical thinking assessment. Both students and lecturers share a common observation that practicum performance assessments are currently conducted manually and lack integration with feedback mechanisms. There is a desire to digitize practicum assessments in the future, with the aim of integrating them with feedback mechanisms to enhance progress. This digitization process is intended to promote objectivity, transparency, and accountability in the assessment procedures. Neither public nor private universities have implemented automated feedback for critical thinking examinations. It is recommended by management to implement a critical thinking assessment that is accompanied by real-time feedback, since this approach is bolstered by sufficient facilities or infrastructure. The present study proposes the amalgamation of a critical thinking practicum with an automated feedback system within a single Android-based application, hence ensuring convenient accessibility through smartphones or laptops.

Acknowledgments. The financing for this work came from the Doctoral Dissertation Research program (Contract Number: 3176/UN18.L1/PP/2023) in the Ministry of Education and Culture, Research, and Technology. I am grateful to everyone who helped with the research and production of the manuscript.

References

1. D. Zhao, A. Chis, G.-M. Muntean, and C. H. Muntean, "a Large-Scale Pilot Study on Game-Based Learning and Blended Learning Methodologies in Undergraduate Programming Courses," *EDULEARN18 Proc.*, vol. 1, no. July, pp. 3716–3724, 2018, doi: 10.21125/edulearn.2018.0948.
2. R. N. Ronau, C. R. Rakes, and M. L. Niess, "Educational technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches," *Educ. Technol. Teach. Knowledge, Classr. Impact A Res. Handb. Fram. Approaches*, pp. 1–409, 2011, doi: 10.4018/978-1-60960-750-0.
3. M. Montenegro-rueda, A. Luque-de la Rosa, J. L. S. Sánchez-serrano, and J. Fernández-cerero, "Assessment in higher education during the covid-19 pandemic: A systematic review," *Sustain.*, vol. 13, no. 19, 2021, doi: 10.3390/su131910509.
4. M. Wiethe-Körprich and S. Bley, "Prospective educators as consumers of empirical research: An authentic assessment approach to make their competencies visible," *Empir. Res. Vocat. Educ. Train.*, vol. 9, no. 1, 2017, doi: 10.1186/s40461-017-0052-5.
5. S. Sudirman, "Pengembangan Aplikasi Untuk Pemantauan Dan Evaluasi Pendidikan Anak Usia Dini," *INFOTECH J. Inform. Teknol.*, vol. 1, no. 2, pp. 107–115, 2020, doi: 10.37373/infotech.v1i2.61.
6. M. Saleh and R. M. Salama, "Recommendations for building adaptive cognition-based e-Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 385–393, 2018, doi: 10.14569/ijacsa.2018.090850.
7. M. Saqr, U. Fors, and M. Tedre, "How learning analytics can early predict under-achieving students in a blended medical education course," *Med. Teach.*, vol. 39, no. 7, pp. 757–767, 2017, doi: 10.1080/0142159X.2017.1309376.
8. S. Sudirman *et al.*, "Praktik Penilaian Guru Pendidikan Sains antara Keyakinan atau Pengetahuan Guru? Perspektif Filsafat," *J. Ilm. Profesi Pendidik.*, vol. 7, no. 3c, pp. 2018–2025, 2022, doi: 10.29303/jipp.v7i3c.889.
9. E. M. Y. Yan, "Using Online Peer Assessment Activities to Enhance Team Collaboration in Two Undergraduate Courses," *Innov. Teach. Learn.*, vol. 4, no. 1, pp. 59–73, 2022, doi: 10.4208/itl.20220104.
10. M. A. Dhina, G. Hadisoebroto, S. R. Mubaroq, and I. Gustiana, "Implementation of digital performance assessment to measure pharmacy physics laboratory skills," *Momentum Phys. Educ. J.*, vol. 5, no. 1, pp. 65–72, 2021, doi: 10.21067/mpej.v5i1.5146.
11. J. J. Tseng, "Developing an instrument for assessing technological pedagogical content knowledge as perceived by EFL students," *Comput. Assist. Lang. Learn.*, vol. 29, no. 2, pp. 302–315, 2016, doi: 10.1080/09588221.2014.941369.
12. P. M. Kruit, R. J. Oostdam, E. van den Berg, and J. A. Schuitema, "Assessing students' ability in performing scientific inquiry: instruments for measuring science skills in primary

education," *Res. Sci. Technol. Educ.*, vol. 36, no. 4, pp. 413–439, 2018, doi: 10.1080/02635143.2017.1421530.
13. D. A. Bensley, C. M. Masciocchi, and K. A. Rowan, "A Comprehensive Assessment of Explicit Critical Thinking Instruction on Recognition of Thinking Errors and Psychological Misconceptions," *Scholarsh. Teach. Learn. Psychol.*, vol. 7, no. 2, pp. 107–122, 2021, doi: 10.1037/stl0000188.
14. N. L. R. Riantini, I. W. Suastra, and P. Budi Adnyana, "Development of science practicum performance assessment in junior high school," *SHS Web Conf.*, vol. 42, p. 00090, 2018, doi: 10.1051/shsconf/20184200090.
15. H. Zhao, Y. Li, X. Wang, Y. Kan, S. Xu, and H. Duan, "Inter-brain neural mechanism underlying turn-based interaction under acute stress in women: a hyperscanning study using functional near-infrared spectroscopy," *Soc. Cogn. Affect. Neurosci.*, vol. 17, no. 9, pp. 850–863, 2022, doi: 10.1093/scan/nsac005.
16. S. R. Walters, P. Silva, and J. Nikolai, "Teaching, learning, and assessment: Insights into students' motivation to learn," *Qual. Rep.*, vol. 22, no. 4, pp. 1151–1168, 2017, doi: 10.46743/2160-3715/2017.2777.
17. N. Chafiq, M. Talbi, and M. Ghazouani, "Design and implementation of a risk management tool: A case study of the moodle platform," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 458–461, 2018, doi: 10.14569/ijacsa.2018.090858.
18. S. Sudirman, "Online System on Monitoring and Feedback for Education," *JISA(Jurnal Inform. dan Sains)*, vol. 4, no. 1, pp. 73–79, 2021, doi: 10.31326/jisa.v4i1.900.
19. J. Zheng, W. Xing, and G. Zhu, "Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment," *Comput. Educ.*, vol. 136, no. August, pp. 34–48, 2019, doi: 10.1016/j.compedu.2019.03.005.
20. J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, 2007, doi: 10.3102/003465430298487.
21. W. J. Boone, "Rasch analysis for instrument development: Why,when,and how?," *CBE Life Sci. Educ.*, vol. 15, no. 4, 2016, doi: 10.1187/cbe.16-04-0148.
22. N. Popovic, T. Popovic, I. R. Dragovic, and O. Cmiljanic, "A Moodle-based blended learning solution for physiology education in Montenegro: A case study," *Adv. Physiol. Educ.*, vol. 42, no. 1, pp. 111–117, 2018, doi: 10.1152/ADVAN.00155.2017.
23. G. Valero and P. Cárdenas, "Formative and summative assessment in veterinary pathology and other courses at a mexican veterinary college," *J. Vet. Med. Educ.*, vol. 44, no. 2, pp. 331–337, 2017, doi: 10.3138/jvme.1015-169R.
24. A. Hume and R. K. Coll, "Assessment of learning, for learning, and as learning: New Zealand case studies," *Assess. Educ. Princ. Policy Pract.*, vol. 16, no. 3, pp. 269–290, 2009, doi: 10.1080/09695940903319661.
25. B. J. Yik *et al.*, "Evaluating the impact of malleable factors on percent time lecturing in gateway chemistry, mathematics, and physics courses," *Int. J. STEM Educ.*, vol. 9, no. 1, 2022, doi: 10.1186/s40594-022-00333-3.
26. B. Alharbi, "Saudi Teachers' Knowledge of Critical Thinking Skills and Their Attitudes Towards Improving Saudi Students' Critical Thinking Skills," *Probl. Educ. 21st Century*, vol. 80, no. 3, pp. 395–407, 2022, doi: 10.33225/pec/22.80.395.
27. M. Muhammad, "Integrating Concept of Entrepreneurship Into Science Education," *IJECA (International J. Educ. Curric. Appl.*, p. 153, 2018, doi: 10.31764/ijeca.v0i0.1982.
28. R. Phungsuk, C. Viriyavejakul, and T. Ratanaolarn, "Development of a problem-based learning model via a virtual learning environment," *Kasetsart J. Soc. Sci.*, vol. 38, no. 3, pp. 297–306, 2017, doi: 10.1016/j.kjss.2017.01.001.

29. B. Csapó and G. Molnár, "Online diagnostic assessment in support of personalized teaching and learning: The eDia system," *Front. Psychol.*, vol. 10, no. JULY, 2019, doi: 10.3389/fpsyg.2019.01522.
30. J. Leber, A. Renkl, M. Nückles, and K. Wäschle, "When the type of assessment counteracts teaching for understanding," *Learn. Res. Pract.*, vol. 4, no. 2, pp. 161–179, 2018, doi: 10.1080/23735082.2017.1285422.
31. K. Possin, "CAT scan: A critical review of the critical-thinking assessment test," *Informal Log.*, vol. 40, no. 3, pp. 489–508, 2020, doi: 10.22329/il.v40i30.6243.
32. Z. Yan, "Self-assessment in the process of self-regulated learning and its relationship with academic achievement," *Assess. Eval. High. Educ.*, vol. 45, no. 2, pp. 224–238, 2020, doi: 10.1080/02602938.2019.1629390.
33. V. J. Shute, "Focus on formative feedback," *Rev. Educ. Res.*, vol. 78, no. 1, pp. 153–189, 2008, doi: 10.3102/0034654307313795.
34. J. A. Kulik and J. D. Fletcher, "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review," *Rev. Educ. Res.*, vol. 86, no. 1, pp. 42–78, 2016, doi: 10.3102/0034654315581420.
35. H. Liu, Z. Liu, Y. Lai, and L. Li, "Factors Influencing Collaborative Innovation Project Performance: The Case of China," *Sustainability*, vol. 13, no. 13, p. 7380, 2021, doi: 10.3390/su13137380.
36. C. Harsch, S. Seyferth, and S. V. Larenas, "Evaluating a collaborative and responsive project to develop language assessment literacy," *Lang. Learn. High. Educ.*, vol. 11, no. 2, pp. 311–342, 2021, doi: 10.1515/cercles-2021-2020.
37. A. Poce, F. Amenduni, M. R. Re, and C. De Medio, "Automatic Assessment of University Teachers' Critical Thinking Levels," *Int. J. Adv. Corp. Learn.*, vol. 12, no. 3, p. 46, 2019, doi: 10.3991/ijac.v12i3.11259.
38. S. Sudirman, M. Sarjan, J. Rokhmat, and I. Fauzi, "Penilaian Pendidikan IPA secara Realtime dan Terintegrasi dengan Artificial Intelligence : Perspektif Filsafat," *J. Ilm. Profesi Pendidik.*, vol. 7, no. 4b, pp. 2658–2668, 2022.
39. D. Sun, F. Ouyang, Y. Li, and C. Zhu, "Comparing learners' knowledge, behaviors, and attitudes between two instructional modes of computer programming in secondary education," *Int. J. STEM Educ.*, vol. 8, no. 1, pp. 1–15, 2021, doi: 10.1186/s40594-021-00311-1.
40. J. Tai, R. Ajjawi, D. Boud, P. Dawson, and E. Panadero, "Developing evaluative judgement : enabling students to make decisions about the quality of work Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved .," pp. 467–481, 2018.
41. L. Zhang and P. Li, "Problem-Based mHealth Literacy Scale (PB-mHLS): Development and Validation," *JMIR mHealth uHealth*, vol. 10, no. 4, pp. 1–15, 2022, doi: 10.2196/31459.