

Development and Initial Psychometric Evaluation of a Web-Based Rubric for Assessing Creativity in Mathematics Learning Media Design

Wiwin Sri Hidayati ^{1*}, Jauhara Dian Nurul Iffah ¹ and Moh.Anshori Aris Widya ²

¹ Department of Mathematics Education, University of PGRI Jombang, East Java, Indonesia;

² Department of Computer Science, KH.A.Wahab Hasbullah University, East Java, Indonesia.

* **Corresponding author:** hidayati25.upjb@gmail.com.

ABSTRACT: This study aimed to develop and conduct an initial psychometric evaluation of GoSkill, a web-based assessment instrument specifically designed to measure student creativity within Mathematics Learning Media courses. This innovation addresses the growing necessity for technology-enhanced tools that support systematic and objective evaluation of creative performance in higher education. Employing a Research and Development (R&D) framework, the research encompassed stages of needs analysis, instrument design, product development, expert validation, limited field testing and reliability analysis. The study involved 84 mathematics education students enrolled in the Mathematics Learning Media course at four universities in East Java, Indonesia: University of PGRI Jombang, KH. A. Wahab Hasbullah University, Islamic University of Majapahit, and Wahidiyah University. Content validity, assessed by four experts in instructional technology, yielded a Scale-level Content Validity Index (S-CVI/Ave) of 0.89 indicating good content validity. Construct validity was further examined through Exploratory Factor Analysis (EFA), which confirmed a unidimensional structure with a single dominant factor explaining 53.45% of the total variance (KMO = 0.76). Item factor loadings ranged from 0.62 to 0.81, while reliability analysis demonstrated high internal consistency (Cronbach's α = 0.865; McDonald's ω = 0.87). These findings suggest that GoSkill was a feasible instrument for the formative assessment of student creativity in designing mathematics learning media. However, as this study represents an initial validation, further research utilizing advanced analytical techniques and broader, more diverse samples is required to establish more comprehensive psychometric evidence for the instrument.

Keywords: rubric-based assessment; formative assessment; pre-service teacher education; higher education Indonesia, instrument development.

I. INTRODUCTION

The Quality Authority specifies that graduate competencies must be achieved through the National Standards for Higher Education, which include outcome, process, and input standards. Learning is expected to be engaging, inclusive, collaborative, creative, and impactful. Student learning outcomes are evaluated using various methods such as written and oral examinations, project assessments, assignments, competency tests, and other relevant forms of evaluation [1]. One learning model that aligns with these standards is project-based learning, which supports the development of both academic abilities (hard skills) and students' social and creative capacities (soft skills).

In the mathematics education curriculum, the Mathematics Learning Media course plays an important role in fostering soft skills, particularly creativity. Emphasize that creativity is one of the essential competencies needed in the 21st century, enabling future teachers to design innovative learning experiences [2]. According to [3], creativity involves fluency, flexibility, originality, and elaboration. In mathematics education, creativity involves not only generating new ideas but also applying them to develop instructional media that are engaging, effective, and aligned with students' learning needs [4]. Study [5] highlights creativity as a key component of divergent thinking,

enabling individuals to generate multiple solutions to a problem. This perspective aligns with [6], who explains that mathematical creativity includes posing problems, making conjectures, and producing non-standard ideas during learning. Therefore, instructors in this course need assignments and assessment tools that capture both the final product and the creative thinking processes involved in developing learning media.

The rapid development of digital technology presents opportunities to use web-based tools not only to support learning but also to improve assessment practices, as evidenced by studies on online assessment implementation and trends in digital assessment instruments [7–10]. Previous studies introduced spreadsheet-based models to assess creativity in mathematics learning media produced by students. These models have since evolved into web-based applications designed to enhance efficiency, usability, and the objectivity of feedback. This instrument aligns with the OECD PISA 2022 Creative Thinking framework, which emphasizes creativity as a productive competence to generate, evaluate, and improve ideas through authentic and contextual tasks [11, 12]. This approach prioritizes performance-based measurement, distinguishing it from traditional instruments that focus solely on dispositions or historical achievements. Creativity is the core of 21st-century soft skills, serving as the foundation for innovation, problem-solving, meaningful communication, and productive collaboration. Creativity does not function in isolation but is manifested through interaction with other skills such as critical thinking, collaboration, and social-emotional competencies [12, 13]. Therefore, focusing the assessment on creativity does not overlook other competencies; rather, it reflects the integrative ability to activate and coordinate supporting skills in instructional practice.

The evolution of creativity assessment in education has undergone a significant transformation alongside advancements in digital technology and artificial intelligence (AI). Traditional approaches based on rubrics and manual evaluation are increasingly being supplemented by automated scoring systems that leverage machine learning and large language models to enhance assessment efficiency, objectivity, and scalability [14, 15]. Platforms such as the Creativity Assessment Platform (CAP) demonstrate that creativity assessment can be conducted automatically and in multiple languages while maintaining high predictive validity ($r > 0.800$), thereby offering novel prospects for measuring 21st-century skills in a more comprehensive and adaptive manner [16]. Furthermore, computational approaches to measuring divergent thinking continue to advance, utilizing semantic analysis, natural language processing (NLP), and machine learning algorithms to evaluate the originality, flexibility, and fluency of ideas [17]. Nevertheless, rubric-based assessments remain vital, particularly within educational contexts, as they provide transparent, contextualized evaluations aligned with learning objectives [18]. Consequently, the integration of traditional and digital approaches represents a relevant trajectory for the development of creativity assessments within the context of mathematics education.

In this regard, the present study positions itself as an effort toward the initial development and validation of a web-based creativity assessment instrument contextualized for mathematics instructional media design, thereby bridging practical field requirements with the ongoing evolution of digital assessment. Previous research has examined creativity assessment in education through standardized instruments such as the Torrance Tests of Creative Thinking (TTCT) [3], the Creative Achievement Questionnaire (CAQ) [19], and Creative Problem Solving (CPS) [20]. However, these studies generally focus on general creativity measurement and have not adequately accommodated the specific context of mathematics learning, particularly in the development of instructional media. Moreover, digital technological advancements have not been fully accompanied by the development of web-based assessment instruments that measure creativity systematically, objectively, and contextually [21, 22].

Several studies integrating technology into assessment remain limited to using digital platforms as data collection tools, rather than developing structured and psychometrically validated instruments [23, 24]. While rubric-based assessment in higher education has shown potential in supporting formative evaluation, its implementation in digital environments to measure creativity requires further development [18]. Consequently, a research gap persists regarding the development of web-based creativity assessment instruments specifically designed to evaluate mathematics instructional media products. This gap serves as a crucial foundation for the present study to develop and perform an initial evaluation of a web-based creativity assessment instrument relevant to the needs of higher education.

This research contributes by developing a web-based evaluation model for assessing creativity in mathematical learning media produced by students. In this study, creativity is defined as a product quality that reflects the creative thinking process [25]. The creativity criteria include: (1) originality of ideas, (2) alignment with mathematical concepts, (3) educational function, (4) design and aesthetic integration, (5) application of material, and (6) practical value. Each criterion is accompanied by detailed descriptors, including conceptual accuracy, support for understanding, aesthetic presentation, and effective use of materials, while also considering classroom relevance.

The purpose of this study is to develop an assessment tool that is valid, reliable, and relevant for supporting undergraduate mathematics learning. Each criterion uses a four-level rubric (1 = poor, 2 = fair, 3 = good, 4 = excellent). As emphasized by [18], rubrics are essential for clarifying performance expectations, aligning instruction with assessment, and helping students understand success criteria to guide their learning. Based on the objectives outlined above, the research questions formulated for this study are as follows: (1) What is the level of content validity of the developed assessment instrument? (2) How reliable is the instrument in measuring student creativity within the Mathematics Instructional Media course?"

II. RELATED WORK

1. VALIDITY AND RELIABILITY OF EVALUATION INSTRUMENTS

An evaluation instrument is vital for determining whether students have achieved their learning objectives, how well they did, and how well the entire teaching process worked. For these test results to be valuable, the tools must possess two key characteristics: validity and reliability [18, 26]. Validity concerns how well an instrument measures what it is designed to measure [27]. An instrument is considered valid when its items or indicators accurately reflect the construct under investigation [28]. Establishing validity ensures that the gathered data accurately reflect the desired learning outcomes and provides a reliable foundation for educational decision-making and instructional enhancement [29].

Reliability, on the other hand, is the consistency of measurement results. A dependable instrument produces consistent results when repeatedly administered under similar conditions [30, 31]. This feature is invaluable for ensuring data validity, preventing bias, and minimizing random errors and the assessor's subjective assessment. In the academic realm, this reliability ensures that changes in student performance reflect genuine learning gains rather than instability in the instrument [32]. Validity and reliability are mutually dependent. A valid yet unreliable test yields inconsistent results, whereas a reliable but invalid instrument assesses the incorrect construct, rendering it meaningless [18]. Consequently, both facets must be meticulously scrutinized during the development and implementation of assessment instruments for research and educational contexts.

Several international studies underscore the significance of validating and testing reliability in educational measurement [33], for example, developed the Generative AI Literacy Assessment Test (GLAT) and demonstrated its construct validity, examining and reliability under CTT to ensure a consistent and reliable instrument ($\alpha = 0.80$). [34] investigated inter-rater reliability of large language models for student essay scoring. They contended that the high agreement between the grade estimates was used as a measure of fair decision-making. Study [35] stated that good content validity and internal reliability were needed before the development of robust assessment instruments. Taken together, such findings further validate the fact that testing validity and reliability is not purely technical and formal (an end in itself), but an inherent scientific undertaking that guarantees accuracy and consistency, thereby sustaining the credibility of evaluation results within global education.

2. SOFT SKILLS ASSESSMENT

The demand in 21st-century education is shifting away from cognitive performance and toward evaluating non-cognitive performance (soft skills). These are skills that include the ability to communicate, work well with others, think critically, be creative and adaptable, empathize with others, and be socially responsible. As noted by [36], soft skills are essential indicators of how prepared someone is to navigate the shifting dynamics of the global workforce and society. As a result, measuring such skills has become a significant component of state-of-the-art education systems. Soft skills assessment aims to reflect an overall picture of social and emotional competencies in students that the traditional academic exams generally do not measure [37–39]. These tests let teachers see how well students use their interpersonal and intrapersonal skills across different learning situations. Study [40] state that standardized soft skills tests with good psychometric properties can make evaluation results more valid and help make character-based education stronger. Study [41] found that effective assessments of soft skills can do more than just measure how well a program is working; they can also provide feedback to students and teachers. In addition, these assessments enable schools to adjust their curricula to better align with the needs of the workforce, which emphasizes collaboration, innovation, and leadership [42]. In this connection, the assessment of soft skills serves not only to measure them but also to foster their professional and personal development.

Soft skills may be important, but they can be challenging to measure, as people report their own biases; it can be hard to stay objective, and social behaviors might mean different things across cultures [43]. Thus, it is important to make sure the tools you use are valid, reliable, and sensitive to the situation, so the results accurately reflect what students can do [44]. However, the field still requires further research and development, because until robust

instruments are available, education cannot confidently claim to contribute to character formation and vocational guidance. Finally, soft skills assessment is a “new” burning issue in the field of education. It is both diagnostic and developmental, helping increase students’ social and emotional competence. It is here that evidence-based, defensible assessment practice comes into play; it provides educators with a sturdy foundation for decision-making and promotes the creation of graduates who are intellectually competent, socially responsible, and emotionally balanced.

3. CREATIVITY IN DESIGNING MATHEMATICS LEARNING MEDIA

Course Description: Students enrolled in the Mathematics Learning Media course will create original instructional systems that are creative, engaging, and effective in their content to aid learners' understanding of abstract mathematical ideas. Not only does this help develop their technical skills as media producers, but it also promotes the value of creativity as a characteristic of future mathematics teachers [45]. For pre-service teachers, creativity is an important skill that helps them turn complicated math concepts into visual, interactive, and contextually relevant representations that help students understand. Creativity is the ability to come up with new and valuable ideas, strategies, or products that can help you solve problems [46]. In mathematics education, it involves divergent and flexible thinking to design learning media that align with students’ needs and classroom realities [47]. As [48] there, creativity in the learning process helps students think more critically to draw new links among mathematics ideas and come up with different ways of solving problems.

Researching creativity in the design of mathematics learning media is highly relevant, as it will add value both professionally and pedagogically to pre-service teachers. Media design expert students are more likely to do this. Their choices (of approach, tool, and strategy) match learners' characteristics and the kind of technology as well [49–51]. Second, creativity is a hallmark of teachers' reflective and adaptive thinking, traits central to creating meaningful, student-centered learning environments [52]. Thus, the orientation of this study is important for math teacher education. Promoting creativity through media creation not only makes mathematics more engaging but also encourages interactive, enjoyable experiences for students.

Creativity vs. TPACK Framework "Creativity in mathematics learning media development requires a dedicated assessment instrument because its construct characteristics differ fundamentally from the TPACK framework. TPACK is designed as a conceptual framework to map teachers' professional knowledge in integrating technology, pedagogy, and content, rather than a psychometric tool for measuring specific creative thinking processes and outcomes [45]. Meanwhile, creativity is a multidimensional construct encompassing fluency, flexibility, originality, and elaboration of ideas, which are not explicitly measured in TPACK indicators [3, 53]. In mathematics, creativity is also domain-specific as it relates to the ability to produce innovative and mathematically meaningful representations, strategies, or learning media [54]. Therefore, while TPACK is relevant for evaluating the quality of technology integration, it is insufficient for assessing the level of creativity manifested in the product or process of mathematics learning media development.

Soft skills have been explored in various educational contexts, including communication skills [55] and interactive learning media development [56]. Technology-assisted soft skills assessments have also been explored, from multi-platform e-rubrics [57] to AI-based evaluation systems [58]. However, creativity assessment still largely relies on traditional instruments such as the Torrance Tests of Creative Thinking (TTCT) (3), the Creative Achievement Questionnaire (CAQ) [19], and the Creative Personality Scale (CPS) [59]. To date, limited research has focused on developing web-based soft skills assessment models specifically for evaluating creativity in the Mathematics Learning Media course. Positioning Against Existing Instruments (TTCT, CAQ, and CPS). Unlike the TTCT, which focuses on measuring divergent thinking potential through abstract and non-contextual tasks, the instrument developed in this study offers contextual advantages by measuring creativity directly within mathematics learning media development activities. Thus, it assesses not only general creative potential but also 'applied creativity' within the mathematical pedagogical domain. While the CAQ focuses on self-reported creative achievements reflecting long-term tracks, this instrument assesses actual creative performance in structured academic tasks. Furthermore, while the CPS maps creative personality traits, this instrument positions creativity as a tangible performance embodied in a product. Its primary contribution is not to replace the TTCT, CAQ, or CPS, but to enrich creativity assessment with a tool sensitive to pedagogical contexts and the specific needs of mathematics teacher education.

[46] identified four essential dimensions of creativity: fluency, flexibility, originality, and elaboration. When creating math learning materials, creativity also means combining pedagogical, content, and technological knowledge, as outlined in the Technological Pedagogical Content Knowledge (TPACK) framework by [45]. Study [60] stated that creativity indicators include originality, elaboration, and the ability to effectively utilize available resources. Study [48] also highlighted that creativity in education should be seen as the ability to generate new ideas

that are relevant to the learning context and can be practically implemented in classroom settings. Based on these theories, the aspects and indicators of creativity used in this study are presented in Table 1 below.

Table 1. Creativity assessment aspects and indicators.

Aspect Assessed	Indicator
1. Originality of Ideas	Learning media ideas are unique and not commonly used.
2. Alignment with Mathematical Concepts	The learning media accurately support concept learning.
3. Educational Function	The learning media help students understand the material visually/concretely.
4. Integration of Design and Aesthetics	The learning media are neat and visually appealing.
5. Innovation in Material Utilization	Use of creative and environmentally friendly materials.
6. Usability	The learning media are easy to implement in real classrooms.

III. MATERIAL AND METHOD

1. DATA COLLECTION

The study was carried out through several stages, as outlined below:

1.1 Instrument Development

The instrument was developed based on four core components of creativity: fluency, flexibility, originality, and elaboration. These components were further refined through observations in the Mathematics Learning Media course and interviews with course instructors. Through this process, six final aspects were established: originality of ideas, alignment with mathematical concepts, educational function, integration of design and aesthetics, innovative use of materials, and usability. Clear, measurable indicators were then formulated for each aspect and converted into rubric items, which were adapted for the web-based format to operationalize the assessment. The validation of the web-based creativity assessment instrument was conducted through content validity, construct validity, and reliability analysis. These stages were designed to ensure that the developed instrument possesses adequate levels of validity and reliability in measuring student competencies.

1.2 Content Validation

The content validity and clarity of the instrument were verified by experts. They also considered the relevance of each aspect, indicator, and rubric item, as well as the accuracy of phrasing and overall readability. To assess content validity, the validation results were analyzed using the Aggregate Validity Index. Content validity results were analyzed using the Content Validity Index (CVI) based on expert assessments.

1.3 Instrument Trial, Construct Validity, and Reliability

A pilot test was conducted with students from four universities participating in the Mathematics Learning Media course. The data collected from this trial were utilized to evaluate the construct validity and reliability of the web-based soft skills assessment tool. Construct validity was evaluated using Exploratory Factor Analysis (EFA). EFA was selected as it is more suitable for small sample sizes ($N = 84$) compared to Confirmatory Factor Analysis (CFA), which typically requires a larger sample size to achieve stable results. Furthermore, reliability testing was conducted to determine the instrument's internal consistency using the Cronbach's Alpha coefficient.

1.4 Instrument Revision

Before the instrument was widely used, it was revised based on expert feedback and trial results to ensure it met validity and reliability standards. Data were collected at multiple points:

- During expert validation, structured feedback on each rubric item, indicator, and aspect was gathered.
- During limited trials, students completed the web-based instrument, and their responses were automatically recorded for analysis.
- During reliability testing, the trial data were analyzed statistically to evaluate the instrument's consistency and accuracy in measuring creativity.

2. RESEARCH DESIGN

This research employed a Research and Development (R&D) approach consisting of six stages: (1) needs analysis, conducted to identify problems in current creativity assessment practices and determine the requirements for a web-based instrument; (2) design, where the assessment framework, indicators, and rubric structure were formulated; (3) product development, which involved constructing the web-based instrument and preparing initial components; (4) expert validation, carried out by specialists in instructional technology and educational assessment to evaluate content accuracy and appropriateness; (5) limited trials, implemented with students to test usability and practicality; and (6) reliability testing, performed to examine the consistency of the instrument through statistical analysis. These stages were used to develop, validate, and evaluate a web-based soft skills assessment tool designed to measure students' creativity in the Mathematics Learning Media course

A needs analysis was conducted to identify challenges in creativity assessment within the Mathematics Learning Media course. Subjects included lecturers and Mathematics Education students selected through purposive sampling. Data were collected through classroom observations and semi-structured interviews with lecturers, then analyzed using a descriptive-qualitative approach to formulate aspects, indicators, and instrument characteristics suitable for the course context.

The research subjects were 84 students from the Mathematics Education Study Program enrolled in the Mathematics Learning Media course at four universities in Indonesia: University of PGRI Jombang, KH.A.Wahab Hasbullah University, Islamic University of Majapahit, and Wahidiyah University, during the even semester of the 2024/2025 Academic. All participants were recruited voluntarily and provided written informed consent after receiving information regarding the study's objectives, procedures, risks, and benefits. To avoid power dynamic pressures, participation did not affect grades, and data collection was handled by non-faculty research assistants. Identities were anonymized, and data were stored securely. The researchers declare no conflicts of interest; government funding did not influence the research design or reporting.

Construct validity analysis was performed using Exploratory Factor Analysis (EFA) with the Principal Component Analysis method and Varimax rotation. Data adequacy testing revealed a Kaiser-Meyer-Olkin (KMO) value of 0.76, indicating that the sample size was meritorious and sufficient for factor analysis. Furthermore, Bartlett's Test of Sphericity yielded a significant result ($\chi^2 = 132.45$; $p < 0.001$), confirming that the data were suitable for further analysis.

Factor extraction results identified a single primary factor with an eigenvalue of 3.21 (> 1), which accounted for 53.45% of the total variance. This suggests that the instrument possesses a unidimensional structure in measuring the creativity construct. The factor loading values for each item ranged from 0.62 to 0.81, indicating that all items contributed significantly to the identified factor. The details of the factor loadings are presented in Table 2 below:

Table 2. Factor loading details.

No	Creativity Aspect	Factor Loading
1	Originality of Ideas	0.78
2	Alignment with Mathematical Concepts	0.74
3	Educational Function	0.69
4	Integration of Design and Aesthetics	0.81
5	Innovation in Material Utilization	0.76
6	Usability	0.62

These results demonstrate that all instrument items met the factor loading criterion of ≥ 0.50 , thus establishing construct validity at the exploratory stage. Consequently, the instrument exhibits a clear factor structure and is statistically robust for measuring student creativity in the development of mathematics instructional media. Construct validity is limited to evidence of internal consistency and logical links between aspects supported by theoretical foundations and expert validation. Reliability was analyzed using Cronbach's Alpha, where coefficients ≥ 0.70 indicate acceptable internal consistency for the initial development stage. In addition, four expert validators participated in the instrument validation process, including one expert in mathematical creativity, one in learning assessment, and two in information technology.

The instrument is based on four components of creativity: fluency, flexibility, originality, and elaboration, contextualized for mathematics learning media. Six assessment aspects were formulated: originality of ideas, alignment with mathematical concepts, educational function, integration of design and aesthetics, innovative use of materials, and usability. The instrument is used by lecturers/experts and in student peer assessment in

constructing mathematics learning media. Content validity was revised using the Content Validity Index (CVI) and validated by four experts. Analysis included Item-level CVI (I-CVI ≥ 0.78) and Scale-level CVI (S-CVI ≥ 0.80), with qualitative feedback used for formative revisions

The content validity procedure was revised using the Content Validity Index (CVI) approach and validated by four experts (mathematics creativity, learning evaluation, and information technology). Validity was analyzed at the item level (I-CVI) and scale level (S-CVI), with eligibility criteria of ≥ 0.78 and ≥ 0.80 , respectively. Qualitative feedback from experts was analyzed and used as the basis for formative revisions of the rubric indicators and descriptors. Data analysis consisted of two phases: (i) Qualitative Data, expert comments and suggestions were examined descriptively to identify areas for improvement. (ii) Quantitative Data, the aggregate validity index was applied to determine the instrument's level of content validity. All expert scores were converted into proportions relative to the maximum score to calculate the aggregate validity index:

$$\text{aggregate validity index} = \frac{\text{average Expert Score}}{\text{maximum score}} \quad (1)$$

The interpretation of the index was based on the following criteria:

Table 3. The interpretation of the index.

Index	Description
≥ 0.90	Excellent
0.80 – 0.89	Good
< 0.80	Needs revision/improvement

Exploratory Factor Analysis (EFA) was used to test construct validity. Cronbach's Alpha was used to measure internal reliability, with a coefficient ≥ 0.70 considered reliable.

IV. RESULTS AND DISCUSSION

1. RESULTS

Observations during students' project presentations revealed that most lecturers prepared their own assessment rubrics for the Mathematics Learning Media course. Three out of four lecturers did not participate in assessing student tasks. Several rubrics were developed without direct links to established theories or criteria, raising concerns about their validity and reliability. One lecturer reported using a digital rubric platform, indicating that the adoption of e-assessment in PjBL contexts was still at an early stage. Taken together, these findings highlight the need for:

- Developing consistent guidelines for rubric construction in PjBL contexts,
- Encouraging student participation in assessment design to enhance authenticity, and
- Lecturer capacity building in using and validating digital instruments or e-rubrics to improve project assessment quality.

Accordingly, the research team developed a web-based soft skills assessment model to measure students' creativity in the Mathematics Learning Media course.

2. INSTRUMENT DEVELOPMENT

The developed instrument successfully aligned with the established creativity aspects: Originality of Ideas, Alignment with Mathematical Concepts, Educational Function, Integration of Design and Aesthetics, Innovation in Material Utilization, and Usability. Each aspect was operationalized into relevant indicators and assessment items, enabling comprehensive measurement of students' creativity in the course. The web-based format enabled practical, interactive data entry and facilitated the evaluation process.

The following figure presents the design of the web-based soft skills assessment instrument.

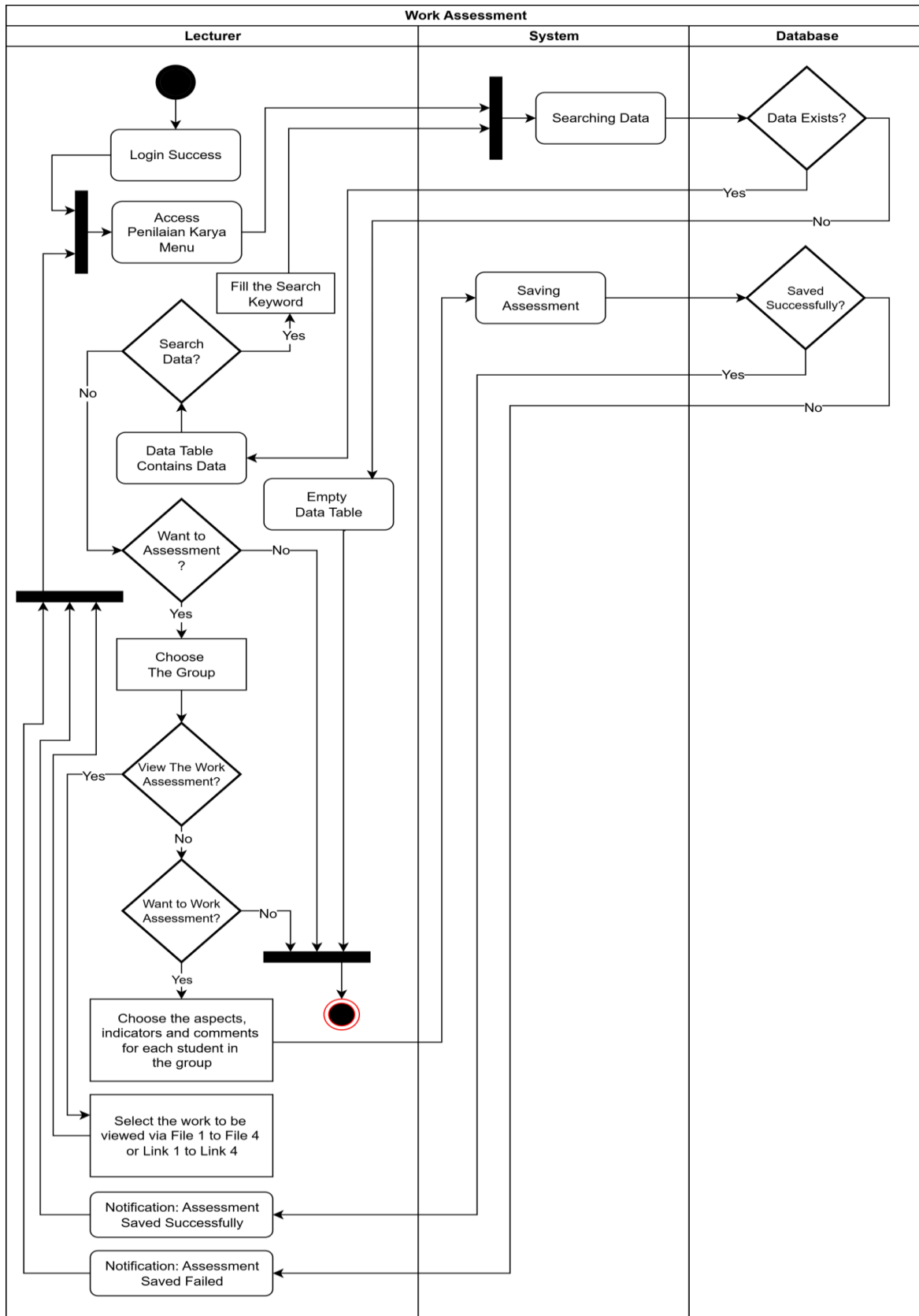


FIGURE 1. Design of the web-based soft skills assessment instrument.

Figure 1 illustrates the design of the web-based soft skills assessment instrument, depicting the workflow between lecturers, the system, and the database. Lecturers begin by logging in and accessing the work assessment menu, then search for student data and select the group to be assessed. After determining the aspects, indicators, and comments for each student, the lecturer selects the file or link of the work to be evaluated. The system processes the data search, saves the assessment results, and provides notifications of successful or failed storage, while the database verifies data existence and stores the assessment results. This diagram emphasizes the structured interaction between users, systems, and databases to ensure a valid and transparent assessment process. The following figures present the development of the web-based soft skills assessment instrument. This menu consists of Assessment Management, Student Management, Project Assessment, Final Score Generation, Student Project Scores, and Student Score Reports.

Table 4. Web-based soft skills assessment instrument.

Aspect Assessed	Indicator	Score 1 (Poor)	Score 2 (Fair)	Score 3 (Good)	Score 4 (Excellent)
1. Originality of Ideas	Learning media ideas are unique and not commonly used.	Replicates existing learning media	Simple modification of existing learning media	Modification with a different concept	Produces entirely new learning media
2. Alignment with Mathematical Concepts	The learning media accurately support concept learning.	Does not support concept learning accurately	Supports only part of the concept, learning accurately	Supports concept learning accurately	Strongly supports concept learning accurately
3. Educational Function	The learning media help students understand the material visually/concretely.	Does not help understanding	Helps in understanding part of the material	Helps in understanding all the material	Helps and facilitates a complete understanding of all material
4. Integration of Design and Aesthetics	The learning media are neat and visually appealing.	Presentation is untidy and unattractive	Presentation is somewhat neat and fairly appealing	Presentation is neat and appealing	Presentation is very neat and highly appealing
5. Innovation in Material Utilization	Use of creative and environmentally friendly materials.	Materials are irrelevant or wasteful	Materials are common and ordinary	Materials are creative and cost-effective	Materials are unique, innovative, and educational
6. Usability	The learning media are easy to implement in real classrooms.	Difficult to implement in real classrooms	Can be implemented with limitations in real classrooms	Easy to implement in real classrooms	Very easy to implement in real classrooms

Figure 2, part of the student work assessment, emphasizes the originality of ideas, adherence to mathematical concepts, educational function, and the aesthetic appeal of the media design. Each student is evaluated using specific indicators marked with a checklist. Assessment is focused on creativity, academic compliance, educational function, and aesthetics. In this section, the lecturer must input the data of students enrolled in the Mathematics Learning Media course. The lecturer must create student accounts with usernames and passwords. The lecturer also needs to assign a group leader, who will upload the files to be assessed. The uploaded files include a video link of the implementation of the developed mathematics learning media, a PowerPoint file, and the final implementation report at the partner school. Each group member is required to objectively assess their peers.



FIGURE 2. Details of the assessment of the work.

3. CONTENT VALIDATION RESULTS

The evaluation results of two information technology experts on the soft skills assessment instrument are presented in Table 5.

Table 5. Evaluation results of information technology experts on the soft skills assessment instrument.

No	Question	Evaluation Results	
		IT Expert 1	IT Expert 2
System Function			
1	I can easily log in to the GoSkill system	5	5
2	The system facilitates lecturers in grouping students appropriately	4	4
3	The system provides peer-assessment features among students within a group	5	4
4	Lecturers can easily assess students through the system	4	4
5	The system clearly displays the final results of the students' creativity assessment	4	5
Usability			
6	The GoSkill system interface is easy to understand	4	4
7	The system menu navigation is simple and not confusing	4	4
8	The assessment input process is fast	4	3
9	The system is user-friendly and does not require long or complex training to operate	4	4
10	Its visual layout is engaging and avoids repetitive or monotonous design elements	4	4
Assessment Process			
11	The creativity assessment instrument embedded in the system corresponds well with established creativity indicators.	5	4
12	Peer evaluations among group members are carried out in a fair and balanced manner.	4	3
13	Lecturers can track each student's progress and level of participation directly in the system.	4	4
14	The distribution of assessment weights between students and lecturers reflects appropriate proportions and expectations.	3	3
15	The system provides concise result summaries that help evaluate student performance.	4	4
Results and Reports			
16	My assessment data is stored securely and treated confidentially.	5	5
17	The system operates reliably with minimal technical errors or disruptions	4	5
18	Access speed remains stable and responsive, with no frequent lag.	4	4
19	The platform functions smoothly across various devices, such as laptops and smartphones.	4	5
20	Even when accessed simultaneously by multiple users, the system maintains good performance and responsiveness.	4	4
Security & Reliability			
21	My assessment data is secure and confidential	4	4
22	The system rarely experiences errors or interruptions during use	3	4
23	The system has good access speed without frequent delays	4	4
24	The system can be accessed smoothly using different devices (laptop/smartphone)	4	4
25	The system functions well even when accessed by many users simultaneously	4	4
Satisfaction & Development			
26	The GoSkill system helps to assess creativity more objectively	4	4
27	The GoSkill system facilitates lecturers in assessing students	5	4
28	The system improves the transparency of peer assessments within groups	4	5

29	I find the GoSkill system useful in enhancing student creativity	4	4
30	Overall, I am satisfied with using the GoSkill system	4	4
Total Score		123	123
Average		4.1	

Table 5 represents the IT expert validation results, showing that the GoSkill system possesses excellent technical quality, with an average score of 4.1 from both experts. High scores in the system function aspect indicate that key features such as login, student grouping, lecturer assessment, and the presentation of creativity assessment results are functioning well and meeting user needs. Regarding usability, the interface appearance, menu navigation, and ease of use were also rated well, although the speed of assessment input can still be optimized. In the assessment process aspect, experts assessed that the creativity instrument integrated into the system is aligned with standard creativity indicators, and the assessment process by lecturers and peers can be conducted fairly and monitored. The lowest score was recorded for the weighting between students and lecturers, indicating a need for review to achieve a more ideal proportion. In terms of results, security, and reliability, the system is considered stable, secure, and accessible via various devices without significant issues. Scores in satisfaction and development also indicate that GoSkill is perceived as beneficial, objective, and feasible for use as a student creativity assessment system in learning.

$$\text{aggregate validity index} = \frac{4.1}{5} = 0.82 \tag{2}$$

The mean expert rating score was 4.1 (out of 5), and the total content validity index was 0.82. According to the interpretation grades, this value is classified as “Good” (the desired use for the instrument). The instrument’s content validity was analyzed using the Content Validity Index (CVI) approach at both the item and scale levels. The analysis results show that the Item-level Content Validity Index (I-CVI) values range from 0.85 to 1.00, with most items obtaining a value of ≥ 0.78 , indicating an adequate level of relevance based on expert judgment. The instrument achieved a Scale-level Content Validity Index (S-CVI/Ave) value of 0.89, indicating overall content suitability. In addition to quantitative assessments, experts provided qualitative feedback regarding the clarity of indicator phrasing, consistency of rubric descriptors, and the alignment of assessment aspects with creativity measurement objectives. This input served as the basis for the formative revision of the instrument prior to limited testing. Table 6 presents calculations comparing a specialist in mathematical creativity with a learning assessment expert on the soft skills evaluation instrument

Table 6. Evaluation results by the mathematics creativity expert and the learning assessment expert on the soft skills assessment instrument.

Assessed Aspect	Indicator	Mathematics Creativity Expert	Learning Assessment Expert
1. Original and Innovative Thinking	The learning media demonstrates originality and avoids repetitive or conventional designs.	4	4
2. Appropriateness to Mathematical Concepts	It aligns accurately with the intended mathematical concepts.	4	4
3. Educational Function	The media supports comprehension by offering visual and tangible representations.	4	4
4. Design and Aesthetic Integration	Its overall presentation is tidy, engaging, and practical for classroom use.	4	4
5. Innovation in Material Utilization	The choice of materials shows innovation and ecological consideration.	4	4

6. Potential for Application	The product is feasible for replication and implementation in authentic learning environments	4	4
Total Score		24	24
Average		4	4

Table 6 shows the validation results from the Mathematics Creativity Expert and the Learning Evaluation Expert, indicating that the developed learning media possesses very high quality, with an average score of 4.0 across all aspects. This score indicates that the media meets eligibility criteria in terms of both creativity and pedagogical function. The originality and innovation aspects received high ratings, showing that the media is non-conventional and possesses elements of novelty relevant to mathematics learning. Adherence to mathematical concepts and educational function were also rated very highly, signifying that the media accurately represents mathematical concepts and assists students in understanding material through visual and concrete displays. In terms of design and aesthetics, the media is considered neat, attractive, and practical for classroom use. High scores in innovation in material use and potential application indicate that the media is not only creative and eco-friendly but also feasible for replication and implementation in real learning situations. Overall, these results confirm that the developed learning media is valid and ready for use in learning activities and further research.

The instrument received an average expert rating of 4.0 out of 4, indicating an overall validity index of 1.00. According to the criteria that have been set, this score is in the "Very Good" range, which means that the instrument is suitable for practical use.

4. CORRELATION ANALYSIS RESULTS

The analysis showed that all dimensions were significantly correlated with the total score at the 0.01 level of significance. The highest was, therefore, the correlation for Originality of Ideas ($r = 0.572$; $p < 0.01$) with the dependent variable following then Design Coherence and Aesthetics ($r = 0.552$; $p < 0.01$), Innovation in Material Utilization ($r = .531$; $p < .01$) and Correlation with Mathematical Concepts ($r = .520$, $p < 0.01$). Conversely, Educational Function ($r = 0.353$; $p < 0.01$) and Usability Potential ($r = 0.374$; $p < 0.01$) showed similarly weak but still statistically significant correlations in the bivariate analysis. Quote Table 7 shows that each contributes positively to the instrument's overall quality, with originality and design coherence having the most significant direct influence. The relationship pattern also shows that the Originality of Ideas and Design Coherence & Aesthetics criteria have a stronger relationship with the total quality index than the other criteria.

Table 7. Pearson correlation of student creativity aspects.

Original and Innovative Thinking	Pearson Correlation	.572
	Sig. (2-tailed)	.000
	N	84
Appropriateness to Mathematical Concepts	Pearson Correlation	.525
	Sig. (2-tailed)	.000
	N	84
Educational Function	Pearson Correlation	.353
	Sig. (2-tailed)	.001
	N	84
Design and Aesthetic Integration	Pearson Correlation	.552
	Sig. (2-tailed)	.000
	N	84
Innovation in Material Utilization	Pearson Correlation	.531
	Sig. (2-tailed)	.000
	N	84
Potential for Application	Pearson Correlation	.374

	Sig. (2-tailed)	.000
	N	84

. Correlation is significant at the 0.01 level (2-tailed).
 . Correlation is significant at the 0.05 level (2-tailed).

Table 8. Reliability statistics.

Cronbach's Alpha	N of Items
.865	6

Internal consistency of the parallel version of the original version was assessed using Cronbach's Alpha = 0.865 for the six items developed here as shown in Table 8. The value is > 0.60, indicating that the instrument possesses good reliability and is efficient for use in early educational research. While the results show good internal consistency among items, further modification of item statements could raise the reliability coefficient above 0.70, thereby placing it in the solid scale category. McDonald's omega (ω) coefficient was also calculated to provide a more accurate reliability estimate, particularly in addressing the potential for correlated measurement errors. The resulting value of 0.87 further reinforces the instrument's reliability. These findings indicate that the instrument possesses a high level of reliability and is suitable for academic use.

5. REVISION OF THE WEB-BASED SOFT SKILLS ASSESSMENT INSTRUMENT

Based on the validity and reliability results, the web-based soft skills assessment instrument was revised as illustrated in the following figure.

1. Use HTTPS protocol instead of HTTP. This revision will be implemented once the Production Server has been determined. Currently, the system is hosted on a temporary server using the HTTP protocol
2. Revise the indicators: any indicator with a value of zero (0) will be removed from the Score Report.

Aspect	Indicator	Score
Originality of Ideas	Learning Media Ideas that Generate New Learning Media	3,00
	Learning Media Ideas by Modifying Different Concepts	0,75
Suitability with Mathematical Concepts	Learning Media that Strongly Supports Accurate Conceptual Learning	4,00
Educational Function	Learning Media that Helps and Facilitates Understanding of All Materials	2,00
	Learning Media That Helps Understanding of All Materials	1,50
Integration of Design and Aesthetics	Learning Media Display that is Very Neat and Very Attractive	2,00
	Learning Media Display that is Neat and Attractive	0,75
	Learning Media Display that is Fairly Neat and Fairly Attractive	0,50
Innovation in The Use of Materials	Use of Unique, Innovative and Educational Materials	4,00
Potential for Application	Learning Media that is Very Easy to Apply in Real Classroom Condition	1,00
	Learning Media that is Easy to Apply in Real Classroom Condition	2,25
Total Score 21,75 with The Predicate "Very Creative"		

FIGURE 3. Revise the indicators in the soft skills assessment instrument.

Figure 3 displays the revised indicators of the assessment instrument used to evaluate creativity in learning media development. The instrument is presented in a table consisting of three columns: Aspect, Indicator, and Score. The assessed aspects include Originality of Ideas, Alignment with Mathematical Concepts, Educational Function, Integration of Design and Aesthetics, Innovation in Material Use, and Application Potential. Each aspect is elaborated into several indicators describing the learning media's ability to generate new ideas, modify concepts,

support accurate conceptual understanding, assist material comprehension, feature neat and attractive designs, use innovative and educational materials, and ensure ease of implementation in real classroom conditions.

3. **Revise:** Add an electronic signature for validating the authenticity of the Score Report.

Figure 4 shows the revised Score Report with the addition of an electronic signature as a means of validating document authenticity. This electronic signature serves to ensure that the score report is official, valid, and has not been altered after issuance. The report display includes structured assessment information, including student identity, assessment aspects and indicators, and the scores obtained. The presence of an electronic signature indicates that the report has been verified by the authorized party, thereby increasing the level of security, trust, and accountability of the assessment document. With this revision, the Score Report functions not only as an evaluation result but also as a valid and authentic digital document ready for academic and administrative purposes.



FIGURE 4. Score report with electronic signature for authenticity validation.

4. **Vision:** When conducting an assessment, only one indicator may be selected.

Figure 5 illustrates the assessment system interface designed to implement the 'Vision' principle, which ensures that during the assessment process, only one indicator can be selected for each aspect. The interface displays group data and a list of students, along with several assessment aspects such as Originality of Ideas, Compliance with Mathematical Concepts, Educational Function, and Integration of Design and Aesthetics. For each aspect, indicators are presented in a horizontally structured checkbox format. The system visually restricts the rater to selecting only the most appropriate indicator for each assessment aspect per student. This restriction aims to maintain consistency, clarity, and objectivity in assessment, as well as to prevent dual scoring on the same aspect. With this interface design, the assessment process becomes more directed, user-friendly, and aligned with systematic and standardized assessment principles.

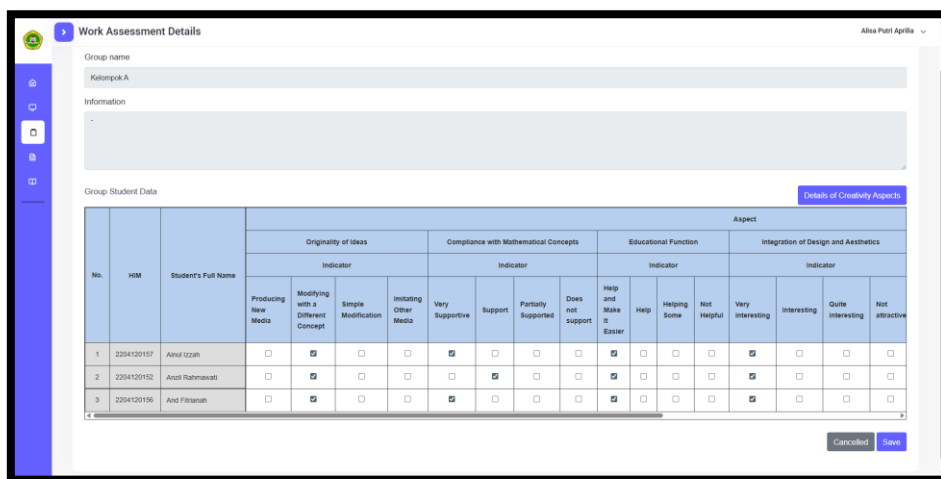
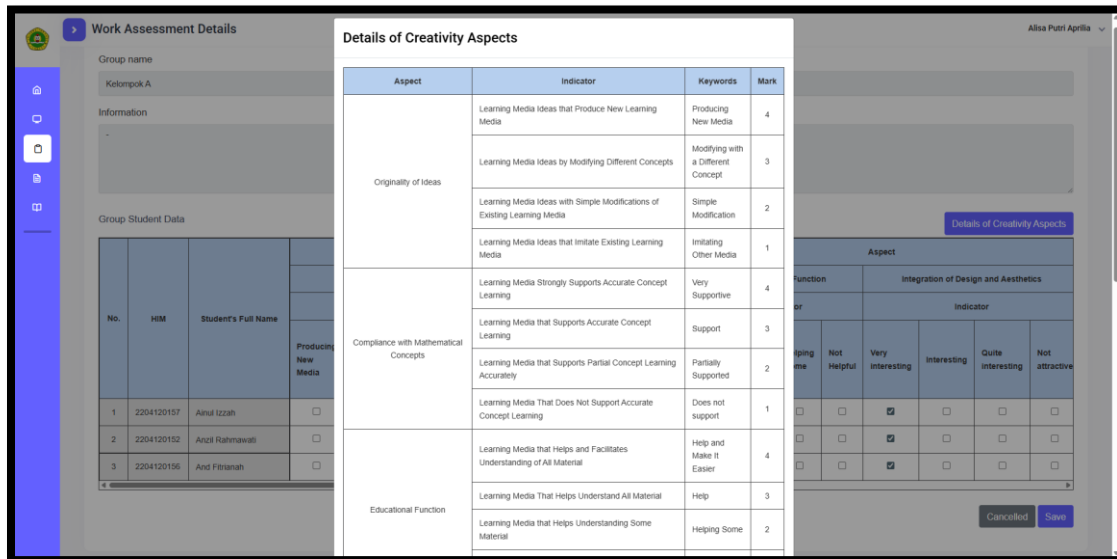


FIGURE 5. Assessment interface showing selection of a single indicator.

5. **Revise** the value scale guide and its trigger button.

Figure 6 shows the revised value scale guide featuring a trigger button on the assessment interface. In this view, the 'Details of Creativity Aspects' button is used as a trigger to display the value scale guide separately in a popup

window. The value scale guide provides more detailed information regarding assessment aspects, indicators, keywords, and the marks (weights) for each assessment level. This presentation helps raters clearly understand the meaning of each indicator before determining a score. With this revision, raters can access the value scale guide quickly and contextually without leaving the main assessment page. This enhances clarity, consistency, and accuracy in the scoring process while making the system interface more informative and easier to use.



URE 6. Revised value scale guide with trigger button.

FIG-

6. Revision using Predicates in the Score Report

Figure 7 displays the revised Score Report with the addition of assessment predicates as a form of qualitative interpretation of the scores obtained. The score report is presented in a table containing assessment aspects, indicators, and scores for each evaluated creativity component. At the end of the report, a total score of 21.75 is displayed, which is automatically accompanied by the predicate 'Very Creative'. The addition of this predicate aims to provide a more easily understood overview of student achievement, not only in numerical form but also as a qualitative category. This revision improves the clarity, readability, and meaning of the assessment results, making the Score Report more informative and communicative for both students and raters in evaluating the level of creativity in learning media development.

The design of the web-based soft skills assessment model for measuring students' creativity in the Mathematics Learning Media course in East Java was developed based on (1) classroom observations to identify learning practices, assessment strategies, and actual conditions; (2) interviews with course instructors to explore assessment methods, rubric usage, and challenges in evaluating students' creativity; and (3) findings regarding aspects and indicators of students' creativity in developing mathematics learning media, such as originality, flexibility, elaboration, and novelty of ideas. The model is built with three user roles: admin, Mathematics Learning Media course instructors, and students whose information has been written to the course. The admin role is responsible for platform uptime and operations, user registration administration, and system-wide stability maintenance. There is teacher empowerment in developing/revision assessment criteria, rating students' submissions, giving feedback, and controlling the creativity. Students, in turn, create media artifacts and submit them for peer review. They can also go back to the feedback and grading provided by instructors during the course.

The online assessment system was developed based on the Project-Based Learning (PjBL) principle, which facilitates the development of students' creativity through real, product-oriented learning. Using digital assessment instruments supports the transparency, consistency, and traceability of evaluation [8]. The plan to create this web-based soft-skills assessment model is a step toward using technology in education while ensuring it remains academically sound and digitally valid.

Alisa Putri Aprilia (2204120155)
 With Creativity Development Details as Follow :

Aspect	Indicator	Score
Originality of Ideas	Learning Media Ideas that Generate New Learning Media	3,00
	Learning Media Ideas by Modifying Different Concepts	0,75
Suitability with Mathematical Concepts	Learning Media that Strongly Supports Accurate Conceptual Learning	4,00
Educational Function	Learning Media that Helps and Facilitates Understanding of All Materials	2,00
	Learning Media That Helps Understanding of All Materials	1,50
Integration of Design and Aesthetics	Learning Media Display that is Very Neat and Very Attractive	2,00
	Learning Media Display that is Neat and Attractive	0,75
	Learning Media Display that is Fairly Neat and Fairly Attractive	0,50
Innovation in The Use of Materials	Use of Unique, Innovative and Educational Materials	4,00
Potential for Application	Learning Media that is Very Easy to Apply in Real Classroom Condition	1,00
	Learning Media that is Easy to Apply in Real Classroom Condition	2,25
Total Score 21,75 with The Predicate "Very Creative"		

FIGURE 7. Score report display incorporating predicates.

V. DISCUSSION

The practice of developing rubrics in-house aligns with findings from Project-based Learning (PjBL) studies, which show that lecturers tend to adapt or create assessment tools according to local contexts and teaching materials rather than following standardized protocols or comprehensive validation procedures [61,62]. However, the limited involvement of students means that assessments are mostly conducted by the teacher, which contradicts the principles of participatory assessment that emphasize peer and co-assessment to promote fairness, engagement, and meaningful feedback [62–64].

Meta-analyses indicate that effective rubrics must be aligned with learning outcomes, piloted, and revised to accurately measure the intended competencies [64]. While online rubric systems are increasingly common, researchers stress that they are effective only when supported by clear technical and pedagogical guidance to ensure reliable formative assessment [65].

These findings are consistent with previous research. [66] emphasize that both newness and usefulness are critical in determining the quality of creative work, which aligns with the strong relationship observed for Originality of Ideas and Design Coherence & Aesthetics. Similarly, [67] highlights that innovative material usage enhances both visual appeal and utility, supporting the observed strong correlation for Innovation in Material Utilization. Moderate correlations between Alignment with Mathematical Concepts and Educational Function align with [68], who found that using design thinking in STEM education improves conceptual understanding and fosters creativity. The positive correlation of Usability Potential further reinforces [66] argument that practicality and user experience are key determinants of product acceptance. Overall, these results suggest that high-quality educational media depend not only on theoretical alignment but also on originality, innovation, and aesthetic integration.

This finding is consistent with [69], who note that alpha coefficients of 0.60–0.70 are acceptable when the constructs are exploratory or newly developed in research. Moreover, [70] emphasized that although higher reliability indicates stronger consistency, excessively high values (>0.90) may suggest redundancy among items. Hence, with a value of 0.865, the instrument in this study can be considered adequate, while still open to revision or the addition of items to raise reliability to a higher category (>0.70).

The development of the model was informed by prior research and context-specific findings [71,72]. By combining classroom observations, instructor insights, and recognized creativity indicators—such as originality, flexibility, elaboration, and novelty—the assessment model captures both theoretical and practical aspects of students' creative performance in mathematics learning media. This approach ensures that the assessment tool is contextually relevant while aligned with established measures of creativity in educational design.

The integration of Information and Communication Technology (ICT) has revolutionized various sectors, notably education, by redefining the learning environment and pedagogical methods [73]. Additionally, studies show that technology-enhanced assessment environments can improve learning by enabling more accurate evaluation and providing timely, meaningful feedback that supports student reflection and learning improvement (74,75). The use of formative assessment and digital feedback tools also helps students learn how to learn independently, encouraging them to keep reflecting on how to improve [76]. These findings are consistent with research on formative assessment, which shows that continuous feedback can enhance student learning [77–79].

This study acknowledges several limitations. First, inter-rater reliability was not assessed using the Intraclass Correlation Coefficient (ICC). Second, while initial construct validity was established, the distinctiveness of each dimension remains to be fully confirmed through more advanced modeling. Third, data collection was restricted to four universities within East Java, which may limit the generalizability of the findings to broader geographic regions or diverse populations. Fourth, reliability was evaluated based on a single-administration design; thus, future research should employ longitudinal or test-retest measures to assess the temporal stability of the scores. Finally, although the instrument demonstrated adequate initial reliability and validity—comparable to established measures (coefficients ranging from 0.70 to 0.90) further refinement or item addition could enhance its psychometric robustness.

VI. CONCLUSION

This research developed and evaluated a web-based instrument designed to assess students' soft skills, specifically creativity, within a Mathematics Instructional Media course. The assessment criteria encompass conceptual coherence, aesthetics, material innovation, and usability. In general, the instructional media developed by students were well-structured, characterized by concrete visuals, eco-friendly components, and alignment with human-centered design principles.

Expert evaluations for content validity categorized the instrument as "good" to "excellent." Reliability analysis yielded a Cronbach's alpha of 0.865 and a McDonald's omega (ω) of 0.87, indicating strong internal consistency across the six items. Exploratory Factor Analysis (EFA) using Principal Component Analysis and Varimax rotation confirmed the data's suitability, with a Kaiser-Meyer-Olkin (KMO) value of 0.76 and a significant Bartlett's Test of Sphericity ($\chi^2 = 132,45$; $p < 0,001$). Factor extraction identified a single primary factor (eigenvalue = 3.21 > 1) explaining 53.45% of the total variance, confirming a unidimensional structure. Factor loadings ranged from 0.62 to 0.81, demonstrating that all items contribute significantly to the underlying construct.

Overall, this web-based instrument exhibits adequate content and construct validity alongside high reliability, making it a viable tool for assessing student creativity in mathematics instructional media design. The limitations of this study, including the modest sample size and the preliminary nature of the pilot test, necessitate further investigation with larger and more diverse cohorts. Future research should explore the application of this instrument across various educational levels and integrate additional soft-skill dimensions to enhance its practical implementation and pedagogical impact.

Funding Statement

This study obtained ethical approval from the Ethics Committee of Universitas PGRI Jombang, Indonesia, on June 14, 2025 (Ref. 027/TKIA/KL/2025). All participants were recruited voluntarily. Prior to data collection, participants received a written explanation regarding the research objectives, the procedures involved, potential risks and benefits, and the right to refuse or withdraw at any time without any academic consequences. Written informed consent was obtained from all participants.

Author Contributions

All authors declare that this study is original, has not been published elsewhere, and represents their own work. Wiwin Sri Hidayati: Designed the research, developed the web-based instrument, wrote the methodology section and discussion, and wrote and revised the manuscript. Jauhara Dian Nurul Iffah: Conducted data collection, analyzed the results, edited, and revised the manuscript. Moh. Anshori Aris Widya: Reviewed the literature, developed the web-based instrument, validated the content, wrote and revised the manuscript. All authors contributed equally to the conception and design of the study, and all have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The dataset generated and analyzed in this study is available upon reasonable request to the corresponding author. The Web-Based Creativity Assessment Instrument can be accessed at <http://goskill.unwaha.ac.id/>

Acknowledgments

The research team expressed their gratitude to the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology for funding this research through the Regular Fundamental Research Grant in 2025 with Number of DIPA: SP DIPA-139.04.1.693320/2025. The researcher also expressed his gratitude to the Rector of Universitas PGRI Jombang who had given motivation to the Research team. The funding party was not involved in the research design, data collection, analysis, interpretation, or writing of the manuscript.

REFERENCES

1. Kemendikisaintek. (2025). *Peraturan Menteri Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia Nomor 39 Tahun 2025 tentang penjaminan mutu pendidikan tinggi*. Indonesia: Kemendikisaintek.
2. Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons.
3. Torrance, E. P. (1966). *Torrance tests of creative thinking*. Scholastic Testing Service.
4. Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM*, 29(3), 75–80.
5. Guilford, J. P. (1975). Creativity: A quarter century of progress. *Journal of Creative Behavior*, 9(2), 75–88.
6. Siswono, T. Y. E. (2016). Proses berpikir kreatif siswa dalam memecahkan dan mengajukan masalah matematika. *Jurnal Ilmu Pendidikan*, 15(1), 60–68.
7. Chang, C. C., Liang, C., Chou, P. N., & Lin, G. Y. (2017). Is game-based learning better in flow experience and various types of cognitive load than non-game-based learning? *Computers in Human Behavior*, 71, 218–227.
8. Hidayati, W., Iffah, J., Trisanti, L., & Nabilah, A. (2024). Digital assessment model to identify student creativity in constructing mathematics instructional media. *Journal of Human University Natural Sciences*, 51(9).
9. Shaban, A. A., & Ali, R. I. (2026). Toward Human-Centered Artificial Intelligence in Education: Adaptive Learning Models for Personalized and Equitable Academic Outcomes. *Qubahan Techno Journal*, 5(1), 1-16.
10. Andriani, R., Mellyzar, M., Lukman, I. R., Muttakin, M., Pasaribu, A. I., & Fadli, M. R. (2023). A review of digital assessment in education: Tools, features, and effectiveness. In *Proceedings of the Malikussaleh International Conference on Education, Social Humanities, and Innovation (MICESHI)*.
11. Barbot, B., Besançon, M., & Lubart, T. (2015). Creative potential in educational settings: Its nature, measure, and nurture. *Education*, 43(4), 371–381.
12. OECD. (2023). *PISA 2022 assessment and analytical framework*. OECD Publishing.
13. Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences. *Journal of Curriculum Studies*, 44(3), 299–321.
14. Huang, J., Saleh, S., & Liu, Y. (2021). A review on artificial intelligence in education. *Academic Journal of Interdisciplinary Studies*, 10(3).
15. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? Opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
16. Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity Research Applications*, 4(2), 238–275.
17. Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., et al. (2021). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings of the National Academy of Sciences*, 118(8), e2022346118.
18. Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
19. Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, 17(1), 37–50.
20. Osborn, A. F. (1953). *Applied imagination: Principles and procedures of creative problem solving*. Scribner.
21. Redecker, C., & Johannessen, Ø. (2013). Changing assessment—Towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79–96.
22. Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning. *Journal of Computer Assisted Learning*, 33(1), 1–19.
23. Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education*, 39(1), 370–407.

24. Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the roles of task model variables in assessment design. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 97–128). Routledge.
25. Besemer, S. P., & Treffinger, D. J. (1981). Analysis of creative products: Review and synthesis. *Journal of Creative Behavior*, 15(3), 158–178.
26. Mardapi, D. (2012). *Pengukuran penilaian dan evaluasi pendidikan*. Nuha Medika.
27. Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50(9), 741–749.
28. Arslan, A. (2020). Reliability and validity of instruments measuring English teachers' TPACK. *International Journal of Assessment Tools in Education*, 7(3), 343–360.
29. Petra, T. Z. H. T., & Aziz, A. M. J. (2020). Investigating reliability and validity of student performance assessment using Rasch model. *Journal of Physics: Conference Series*, 1529(4), 042088.
30. Azwar, S. (2022). *Reliabilitas dan validitas* (4th ed.). Pustaka Pelajar.
31. El-Hamamsy, L., Zapata-Cáceres, M., Barroso, E. M., Mondada, F., Zufferey, J. D., & Bruno, B. (2022). The competent computational thinking test. *Journal of Educational Computing Research*, 60(7), 1818–1866.
32. Akib, E., & Ghafar, M. N. A. (2015). The validity and reliability of assessment for learning. *Education Journal*, 4(2), 64–68.
33. Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. *Computers and Education: Artificial Intelligence*, 100436.
34. Gaggioli, A., Casaburi, G., Ercolani, L., Collovà, F., Torre, P., & Davide, F. (2025). Assessing reliability and validity of large language models for essay assessment. *arXiv preprint*.
35. Kulllan, S., Mansor, M., & Ishak, R. (2022). Validity and reliability of learning organization instruments. *International Journal of Evaluation and Research in Education*, 11(4), 1725–1733.
36. Taguma, M., Feron, E., & Lim, M. H. (2018). *Future of education and skills 2030*. OECD.
37. Cimatti, B. (2016). Definition and assessment of soft skills. *International Journal for Quality Research*, 10(1), 97.
38. Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451–464.
39. Succi, C., & Canovi, M. (2020). Soft skills and employability. *Studies in Higher Education*, 45(9), 1834–1847.
40. Escolà-Gascón, Á., & Gallifa, J. (2022). Measuring soft skills in education. *Studies in Educational Evaluation*, 74, 101155.
41. Colledani, D., Robusto, E., & Anselmi, P. (2024). Assessing key soft skills. *Frontiers in Psychology*, 15, 1405822.
42. Phuti, F., Koloi-Keaikitse, S., Tsheko, G. N., & Oppong, S. (2023). Soft skills assessment scale. *SAGE Open*, 13(4).
43. Al-Sa'di, A., Yamjal, P., Ahmad, E., Panjabi, R., Allott McPhee, C. A. M., & Guler, O. (2023). Assessing educators' soft skills. *Administrative Sciences*, 13(9), 208.
44. De Fruyt, F., & Scheirlinckx, J. (2024). Assessment of social-emotional skills. *ECNU Review of Education*.
45. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge. *Teachers College Record*, 108(6), 1017–1054.
46. Torrance, E. P. (1974). *Torrance tests of creative thinking*. Personnel Press.
47. Leikin, R. (2006). Mathematical connections and problem solving. *Aleh Mathematics Journal*, 36, 8–14.
48. Beghetto, R. A., & Kaufman, J. C. (2014). Classroom contexts for creativity. *High Ability Studies*, 25(1), 53–69.
49. Koehler, M., & Mishra, P. (2009). What is TPACK? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
50. Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013). Creative learning environments. *Thinking Skills and Creativity*, 8, 80–91.
51. Henriksen, D., Mishra, P., & Fisser, P. (2016). Creativity and technology in education. *Journal of Educational Technology & Society*, 19(3), 27–37.
52. Sawyer, R. K., & Henriksen, D. (2024). *Explaining creativity: The science of human innovation*. Oxford University Press.
53. Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important? *Educational Psychologist*, 39(2), 83–96.
54. Leikin, R., & Lev, M. (2013). Mathematical creativity differences. *ZDM*, 45(2), 183–197.
55. Wahyu, M. N., Sutiarso, S., & Bharata, H. (2020). Soft skills communication learning. *Jurnal Cendekia*, 4(1), 406–413.
56. Satriawan, A., Sutiarso, S., & Rosidin, U. (2020). Interactive learning media. *Jurnal Cendekia*, 4(2), 950–963.
57. Mustofa, G. G. (2021). *Soft skills assessment using e-rubric* (Thesis). Universitas Pendidikan Indonesia.
58. Husairi, H., & Sodikin, S. (2023). Soft skills and AI in education. *Al-Jadwa Journal*, 3(1), 111–138.
59. Gough, H. G. (1979). *The creative personality scale*. Consulting Psychologists Press.
60. Munandar, U. (2016). *Pengembangan kreativitas anak berbakat*. Rineka Cipta.
61. Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). Project-based learning review. *International Journal of Educational Research*, 102, 1–13.
62. Dampierre, M. R. de, & Gaya-López, P. J. L. B. (2024). Project-based learning in engineering. *Education Sciences*, 14.
63. Panadero, E., Ortube, A. F., Krebs, R., & Roelle, J. (2025). Online rubric platforms. *Assessment & Evaluation in Higher Education*, 50, 31–49.
64. Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics. *Educational Psychology Review*, 35.
65. Cheng, Y. P., & Huang, Y. M. (2025). LUPDA rubric model. *International Journal of STEM Education*, 45, 1–31.
66. Mulet, E., Royo, M., Chulvi, V., & Galán, J. (2017). Creativity and design outcomes. *Dyna*, 84(200), 38–45.
67. Han, J., Forbes, H., & Schaefer, D. (2021). Creativity and design relationships. *Research in Engineering Design*, 32(3), 289–307.

68. Muneer, S., Santhosh, M., Parangusan, H., & Bhadra, J. (2025). Design thinking meta-analysis. *International Journal of Technology and Design Education*, 1–31.
69. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*.
70. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
71. Hidayati, W. S., & Trisanti, L. B. (2023). Creativity profile in mathematics learning. *JTAM*, 7, 836–847.
72. Windyariani, S., & Setiono, S. (2024). Creative thinking rubric in STEM. *Journal of Innovative Science Education*, 13(50), 64–73.
73. Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of e-assessment. *International Journal of Information and Education Technology*, 8(1), 34–37.
74. Ibarra-Sáiz, M. S., Gómez-Ruiz, M. Á., Balderas, A., & Rodríguez-Gómez, G. (2025). Technology-enhanced assessment. *Technology, Knowledge and Learning*, 1–31.
75. Trujillo, B. P. S., Velarde-Camaqui, D., Gonzales Nuñez, C. A., Castillo Silva, E. V., & Gonzalez Said de la Oliva, M. D. P. (2025). Formative assessment review. *Frontiers in Education*, 10, 1509983.
76. Nicol, D., Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning. *Studies in Higher Education*, 31, 199–218.
77. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
78. Kingston, N., & Nash, B. (2011). Formative assessment meta-analysis. *Educational Measurement: Issues and Practice*, 30(4), 28–37.
79. Wiliam, D. (2018). Assessment for learning implementation. *Assessment in Education*, 25(6), 682–685.