International Journal of Advanced and Applied Sciences

Journal homepage :http://www.ijaas.in

Developing the MathSci 21st app: Enhancing higher-order thinking skills assessment in mathematics and science education within an Islamic context

Zulfiani Zulfiani [1,*], Iwan Permana Suwarna [2], Abdul Muin [3], Tita Mulyati [4], R. Ahmad Zaky El Islami [5]

[1]*Department of Biology Education, Universitas Islam Negeri Syarif Hidayatullah Jakarta, South Tangerang, Indonesia*
[2]*Department of Physics Education, Universitas Islam Negeri Syarif Hidayatullah Jakarta, South Tangerang, Indonesia*
[3]*Department of Mathematics Education, Universitas Islam Negeri Syarif Hidayatullah Jakarta, South Tangerang, Indonesia*
[4]*Department of Primary School Teacher Education, Universitas Pendidikan Indonesia, Bandung, Indonesia*
[5]*Department of Science Education, Universitas Sultan Ageng Tirtayasa, Serang, Indonesia*

ARTICLE INFO

Article history:
Received 5 February 2023
Received in revised form
12 June 2023
Accepted 20 June 2023

Keywords:
Digital assessment
Higher-order thinking skills
MathSci 21st app
Islamic context
Competency assessment

ABSTRACT

The innovation of digital assessment holds profound potential for enhancing educational quality. To measure higher-order thinking skills, intrinsic to effective problem-solving in science and mathematics education, and to cultivate digital literacy, a specialized platform is imperative. This study delineates the developmental trajectory of the MathSci 21st app, designed to assess mathematics and science proficiency within the Islamic context. Emphasizing the pivotal role of higher-level thinking skills in the contemporary landscape, the research method unfolds across distinct phases: Akker (preliminary research), prototype stage (prototyping), summative evaluation, and systematic reflection and documentation. This article confines its focus to preliminary research and prototype stages. The validation of the application prototype engaged a panel of ten experts, while a controlled trial encompassed 32 high school students and one educator. Utilizing observation sheets, questionnaires, and tests as research tools, comprehensive data analysis was executed employing both quantitative and qualitative methods. Research outcomes affirm the feasibility of the Prototype MathSci 21st app, an Android-based competency assessment tool characterized by its integrated and contextual dimensions. Android-based applications not only heighten efficiency and efficacy but also exhibit environmental conscientiousness by reducing paper usage. Additionally, their user familiarity augments acceptability. Significantly, the MathSci 21st app expedites assessment, empowering educators to provide prompt feedback and expedite future learning analysis. This study pioneers a digital assessment paradigm tailored to intricate higher-order thinking skills, thereby addressing critical educational imperatives in mathematics and science within the Islamic milieu.

1. Introduction

Indonesia grapples with empirical realities surrounding contemporary educational challenges pertinent to the 21st century, with a particular focus on higher-order thinking. The curricula adopted by diverse nations incorporate visionary declarations aimed at fostering elevated cognitive capacities in students, constituting a prominent aspect of the science education reform agenda (Fensham and Bellocchi, 2013). However, low-level thinking demands such as remembering and dominance of knowledge transfer continue to dominate science education in schools (Budiarti and Tanta, 2021) and low literacy (Jufrida et al., 2019). There are two reasons why science education is failing in the twenty-first century: A lack of good scientific reasoning models and models for assessing high cognitive competence (Osborne, 2013).

The ability to make evidence-based decisions and to ask rational (reasoning) questions about scientific and social issues are important goals in science education (Beniermann et al., 2021). Higher-order thinking is required in science education to meet the

demands of today's society (Andriyatno et al., 2023; Pujiastuti and Haryadi, 2023; Haryadi and Pujiastuti, 2022; Al Aliyawinata et al., 2021). The OECD (2015) also confirmed that many of the challenges of the twenty-first century will necessitate innovative solutions based on scientific thinking and discovery. Higher-order thinking skills (HOTS) are domain-specific and rely on scientific content and concepts. Analytical, evaluative, and creative thinking are all required during the process (Anderson et al., 2001).

The success of education is largely determined by efforts to improve the learning process and assessments to determine competency achievement. The Minimum Competency Assessment (MCA), which is guarded by the Indonesian Ministry of Education and Culture and replaces the National Examination, was carried out in 2021, providing a new perspective on learning success. These changes include a process-oriented assessment concept that emphasizes teacher feedback after the teacher's assessment process. Initially, formative assessment relied more on informal activities, but recently, the principles of assessment for learning on formative assessment have been adopted into practice and policy around the world. Formative assessment can be viewed as an essential component of teaching and learning because it influences student learning, specifically the satisfaction of students' needs for autonomy, competence, and relatedness, and thus their autonomous motivation (Leenknecht et al., 2021; Weurlander et al., 2012). Formative assessments or assessments for learning have been adopted for educational assessment and evaluation policies or practices (Clark, 2011).

Assessment for learning (AFL) policies in school systems around the world have placed the principle of assessment for learning as a more strongly recommended lesson than assessment as learning as educational reform, including in Australia, Hong Kong, and New Zealand, as well as several other countries (Birenbaum et al., 2015). This type of assessment is better suited to improving learning outcomes for all students. Each country makes an equal effort to adopt a learning assessment approach that supports the national curriculum and achievement standards. Learning tools such as a syllabus, lessons, and assessments are among the integrated efforts made to make it happen. Syllabi and support materials in New South Wales (NSW) are designed to promote an integrated approach to teaching, learning, and assessment (Birenbaum et al., 2015). Apart from the MCA, the Indonesian Ministry of Education and Culture has undertaken the development of a computer-based module known as the CAIS. Furthermore, the Indonesian Ministry of Religion has initiated the trial phase of a Competency Assessment for Islamic school students (CAISS) utilizing a digital framework. Both CAIS and CAISS furnish comprehensive platforms for evaluating student competencies across various educational tiers, concurrently offering avenues for students to engage in the cultivation of information technology-driven higher-order cognitive proficiencies.

In addition, in 2022, CAISS has also been held concurrently for all Madrasa Ibtidaiah (same level as Elementary school) in Indonesia. CAISS is a comprehensive assessment module that identifies students' strengths and weaknesses in reading literacy, numeracy, and scientific literacy, as well as socio-cultural literacy and character surveys. CAISS, like CAIS, is a resource for improving classroom learning. A computer, laptop, or Android mobile phone can be used to access the technological infrastructure in these two computer modules. These two modules are tools for providing integrated problem-solving experiences related to a variety of skills and higher-order thinking skills.

According to the findings of interviews conducted at Senior High Schools in South Tangerang and South Jakarta, students are still not used to solving integrated problems involving multiple disciplines and HOTS dimensions. The questions taught at CAIS, CAISS, and even MCA are centered on scientific literacy, numeracy, social and cultural issues, and a variety of contextual and integrated issues. Providing experience with problems in an integrated scope of fields of knowledge, particularly at the Senior High School level, is a challenge; additionally, students at this level will face their future career paths. A simple computer module/computer program based on Android is required to provide students at the Senior High School level with simple and significant training.

To assess student performance in the digital era, good assessment must be supported by a wide range of assessment frameworks. In 2021, researchers created frameworks and high-level thinking instruments for the Senior High School level, with a novel integration of MathSci Mathematics and Science based on problem-solving, interdisciplinary with an Islamic context. The Mathematics and Natural Sciences assessment framework uses a ladder analogy to refer to Bloom's taxonomy, where assessment type 1 measures the cognitive processes of application in one discipline, assessment type 2 measures the dimensions of application and reasoning with two disciplines, and assessment type 3 measures the cognitive processes of integrated reasoning from three disciplines of mathematics and science. Type 1 assessment is a cross-check of understanding prior to being given the formula type 2 and type 3 with a more complex form of assessment, with type 3 becoming the peak assessment of higher-level thinking. Experts have declared the conceptual framework and MathSci instruments to be valid, and they have been tested on Senior High School students to produce a prototype instrument in the form of a paper and pencil test (Zulfiani et al., 2021)

MathSci uses the Islamic context to strengthen spiritual values, as well as a variety of stimulus questions that will provide information and connect science and Islam as a unified dimension of knowledge. Islamic mathematical integration is possible when the context and relevance of the concept are considered (Kurniawati et al., 2021). The

Islamic context that is integrated into mathematical literacy questions is questions related to verses of the Al-Qur'an, Al-Hadith, and events in everyday life that have the MathSci uses the Islamic context to strengthen spiritual values, as well as a variety of stimulus questions that will provide information and connect science and Islam as a unified dimension of knowledge. Islamic mathematical integration is possible when the context and relevance of the concept are considered (Kurniawati et al., 2021; Johar and Ahmad, 2018). The Islamic context that is integrated into mathematical literacy questions is questions related to verses of the Al-Qur'an, Al-Hadith, and events in everyday life which has the potential to build good character for students, especially teaching methods based on noble values (Putri and Aisyah, 2020; Nihayati et al., 2022). Similarly, research by Astuti et al. (2023) developed a mathematics e-module with a scientific approach integrated with Islamic values with very valid, very practical, and very effective criteria for improving students' mathematics learning outcomes.

Researchers emphasize the significance of digital learning as a tool for improving the authenticity of learning and assessment tasks; however, psychometrically valid instruments are still scarce (Darling-Aduana, 2020). Several researchers have reported the use of digital assessments using iPad technology contributes to effective student learning through formative assessment and introduces analytic approaches (Dalby and Swan, 2019); formative assessment through Quizziz (Rahmah et al., 2019); learning analytics methods to determine whether students' perceived level of digital literacy had an effect on their learning management system (LMS) navigation and overall academic achievement Le et al. (2022). The frequency of use of online assessments has been regularly used in the medium category by half of the teachers in the sample at public higher education institutions in Mexico City. Sullivan et al. (2021) found how teachers utilize computer-based formative assessment (CBFA) in the classroom to help develop targeted professional development to support teachers in using technology to formally assess students.

Consequently, there persists a compelling need to advance digital assessment methodologies, foster proficiency in digital literacy, and gauge the outcomes of learning endeavors. Subsequently, the researcher formulated a proficiency assessment mechanism, manifested as digital applications, underpinned by the aforementioned conceptual framework. This application is envisaged to emerge as an innovative conduit for nurturing digital literacy in the twenty-first century, leveraging the potential of vast data reservoirs and pedagogically-oriented analytical learning paradigms.

This tool serves as a conduit for delineating students' cognitive capacities, progressing through successive stages from intermediate to advanced tiers of thinking prowess. Notably, it elucidates the intricacies of higher-order thinking, a critical facet within science and mathematics education. The concept of HOTS is rooted in the realms of reasoning and problem-solving strategies, emblematic of twenty-first century competencies. Furthermore, the exploration of HOTS holds the potential to engender substantive insights into science education, transcending its status as both a scholarly aptitude and a realm of substantive knowledge.

The principal objectives of this inquiry encompass:

1. The formulation of a MathSci 21st application that is both valid and reliable, tailored for evaluating higher-level thinking proficiencies within the context of Islamic mathematics and science education.
2. The discernment of student and teacher responses, therein affording an appraisal of the application's efficacy and pragmatic utility.

2. Methodology

The research method used was Research and Development (R&D), which adhered to Akker's development model (Nieveen, 2007). This research was carried out at a Senior High School in South Tangerang, Indonesia. This study's sample included 32 class XI students for the 2020/2021 academic year. The sampling method used was non-probability sampling, also known as purposive sampling. Students were sampled based on their knowledge of the human circulatory system (Biology), hydrostatic pressure and fluid continuity (Physics), trigonometry, and minimum and maximum values (Mathematics). As a result, their competence could be measured by the application developed and having a compatible Android-based smartphone.

This study employed a multifaceted approach involving questionnaires from experts in material, media, and language domains, coupled with student questionnaires and both multiple-choice and true-false tests. The multiple-choice questions consisted of three items and were categorized as monodisciplinary questions as they exclusively addressed a single discipline (Biology, Physics, or Mathematics). Additionally, four true-false questions were administered; three fell under the umbrella of interdisciplinary 1 as they encompassed the fusion of two disciplines (such as Biology and Physics, or Physics and Mathematics), while one pertained to interdisciplinary 2 due to its amalgamation of three disciplines (Biology, Physics, and Mathematics). The preparatory framework for the test was aligned with the TIMSS framework in the cognitive domain, encompassing Levels 2 (application) and 3 (reasoning).

In accordance with Akker's framework, the research and development (R&D) endeavor transpired through four distinct stages: (1) preliminary research, (2) the prototyping stage, (3) the summative evaluation stage, and (4) systematic reflection and documentation. Nonetheless, this

study solely delves into two stages of developmental progression in line with the research objectives: the preliminary research stage and the prototyping stage.

The preliminary research phase encompassed a comprehensive literature review and intricate case studies targeting teacher challenges associated with the assessment of higher-order thinking within science and mathematics disciplines at the Senior High School level. During this phase, strategic action plans to ameliorate the identified issues were devised, one of which centered on the implementation of an appropriate assessment-oriented application, characterized by alignment with the intended objectives and demonstrable high reliability.

The subsequent stage, namely the prototype stage, was dedicated to the creation of the MathSci 21st prototype, involving the following sequence of steps: (1) delineation of program design (comprising conceptual design and program flowcharts), (2) optimization of prototype design (including software selection), and (3) formulation of procedures for crafting the MathSci 21st app media. The ensuing elucidation provides a comprehensive account of each of these developmental stages.

2.1. Program design

The MathSci application serves as a digital instrument for assessing Mathematics and Science within the context of Islamic education. The foundational structure of the MathSci assessment is distinguished by its thematic and interdisciplinary orientation. This assessment paradigm encompasses three distinct question variations, denoted as monodisciplinary questions, interdisciplinary questions of the first order (Interdisciplinary 1), and interdisciplinary questions of the second order (Interdisciplinary 2), all tailored to the Islamic educational milieu.

The monodisciplinary questions, categorized as Type 1, pertain to the individual scientific domains of biology, physics, and mathematics within the Islamic framework. These questions span cognitive Levels 2 (L2) and 3 (L3) and are strategically designed to gauge the comprehension of specific concepts within each scientific discipline.

Type 2 questions, classified as Interdisciplinary Question 1, revolve around the integration of two disciplines within the Islamic context, aligning with a cognitive level of L2. This category encompasses diverse permutations, including the fusion of biology and physics, biology and mathematics, as well as physics and mathematics, all contextualized within an Islamic framework.

Conversely, Type 3 questions, designated as Interdisciplinary Question 2 and grounded in cognitive Level L3, expound upon the integration of all three scientific disciplines—Biology, Physics, and Mathematics—within the backdrop of Islamic education. These nuanced question variations collectively articulate a comprehensive approach to assessing the interdisciplinary nexus of knowledge within an Islamic pedagogical context (Table 1).

Table 1: Characteristics of MathSci questions

MathSci question types	Information	Question form	Cognitive level
Type 1	Biology/physics/mathematics in Islamic context	Multiple choice with 5 options	L2 and L3
Type 2	Biology-physics in the Islamic context; biology-mathematics in the Islamic context; physics-mathematics in the Islamic context	True-false with 7 true-false statements; Total 3 questions with 21 statements	L2
Type 3	Biology, Physics, and Mathematics in the Islamic context	True-false with 7 true-false statements; Total 1 question with 7 true-false statements	L3

The MathSci concept material was created through document and material analysis using the 2013 Class XI Senior High School Curriculum Basic Competencies. The application design is based on the flowchart (Fig. 1) below, which includes menu options for either teacher or student users. Following the user's selection, the user will begin with email registration, fill out the bio, read the instructions for use, and continue working on the questions.

2.2. Prototype design optimization

Prototype optimization can be accomplished in a number of steps. The steps for optimizing the prototype include selecting the software or software chosen during the fabrication of the MathSci 21st prototype as follows (Table 2).

Table 2: MathSci 21st app prototype software

Software	Notes
Operating system	Windows 11
Android application development software	Android Studio
UI/UX development software	Figma
API development software	Visual studio code
Software for viewing API	Postman

Hardware, in addition to software, is required for the prototyping process. At this stage, the recommended hardware is a PC or notebook/laptop computer with a minimum specification of an i3 processor and 4GB RAM, or an Android smartphone.

2.3. MathSci 21st app media creation procedure

The procedure used as a guide for making prototypes is shown in Fig. 2.

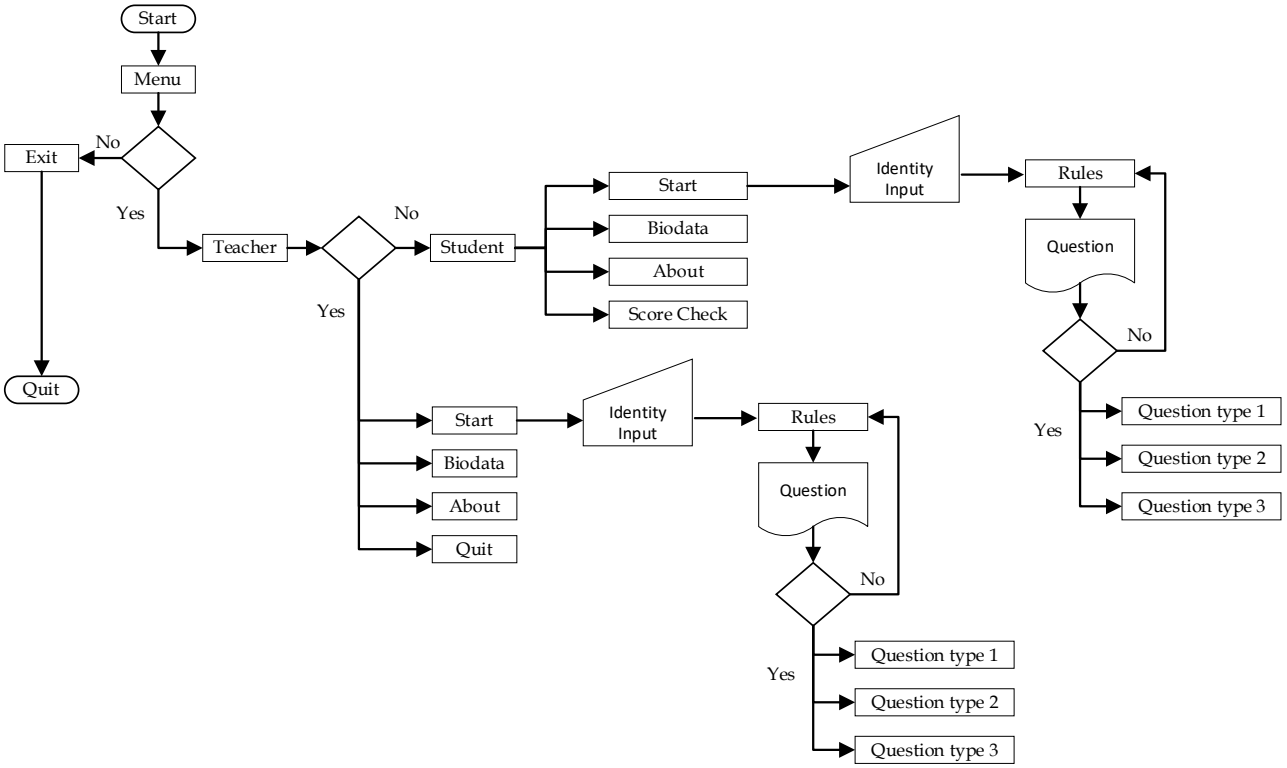

Fig. 1: MatSci 21st app flow chart

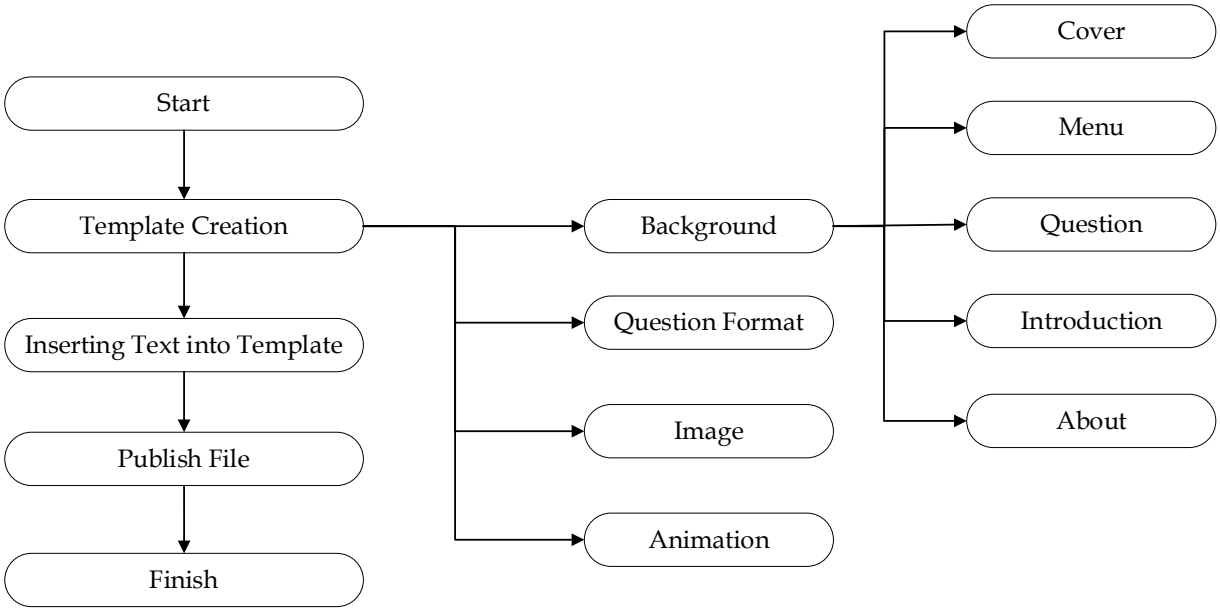

Fig. 2: MathSci 21st app development procedure and admin application

Furthermore, the system's design is carried out using the software. The MathSci application has the advantage of including an admin application (admin app). The admin application aims to store input user data with each individual's full identity, calculate test scores quickly, and the entire database can be downloaded in xls. (Microsoft Excel) format to facilitate data analysis. Furthermore, the application has been validated by experts, and limited trials have been conducted.

Quantitative data analysis was performed. Questions with cognitive level L2 received a 2, while questions with cognitive level L3 received a 3. Monodisciplinary questions were given a weight of 2, interdisciplinary questions were given a weight of 3, and interdisciplinary questions were given a weight of 4. SPSS software was used to analyze the test item data. Expert questionnaires were created using a Likert scale, with scores ranging from 1 (very poor), 2 (poor), 3 (enough), 4 (good), and 5 (excellent) (very good). The validity questionnaire of media experts was analyzed using the CVR (Content Validity Ratio) based on Lawshe (1975). Each question indicator on the questionnaire is labeled as inappropriate (score 0) or appropriate (score 1) based on the expert's response. The CVI (Content Validity Index) value is then calculated by dividing the total CVR value by the number of questions. The CVI calculation results are divided into three categories: 0.00-0.30 (inappropriate), 0.34-0.67

(suitable), and 0.68-1.00. (very suitable). Material and language expert questionnaires on the validity aspect were analyzed using Aiken's (1985) V index. The calculation is done by dividing the total score given by the expert minus the lowest score by the number of experts multiplied by the number of categories that can be chosen minus one. The calculated values are divided into five categories: 0.00-0.19 (very low); 0.20-0.39 (low); 0.40-0.59 (medium); 0.60-0.79 (high); 0.80-1 (very high).

3. Results

3.1. MathSci 21st app

The MathSci 21st app program display includes a home view, menu, application developer biodata, application description, data input, instructions for use, question menu display, and score tracking (Figs. 3 and 4).

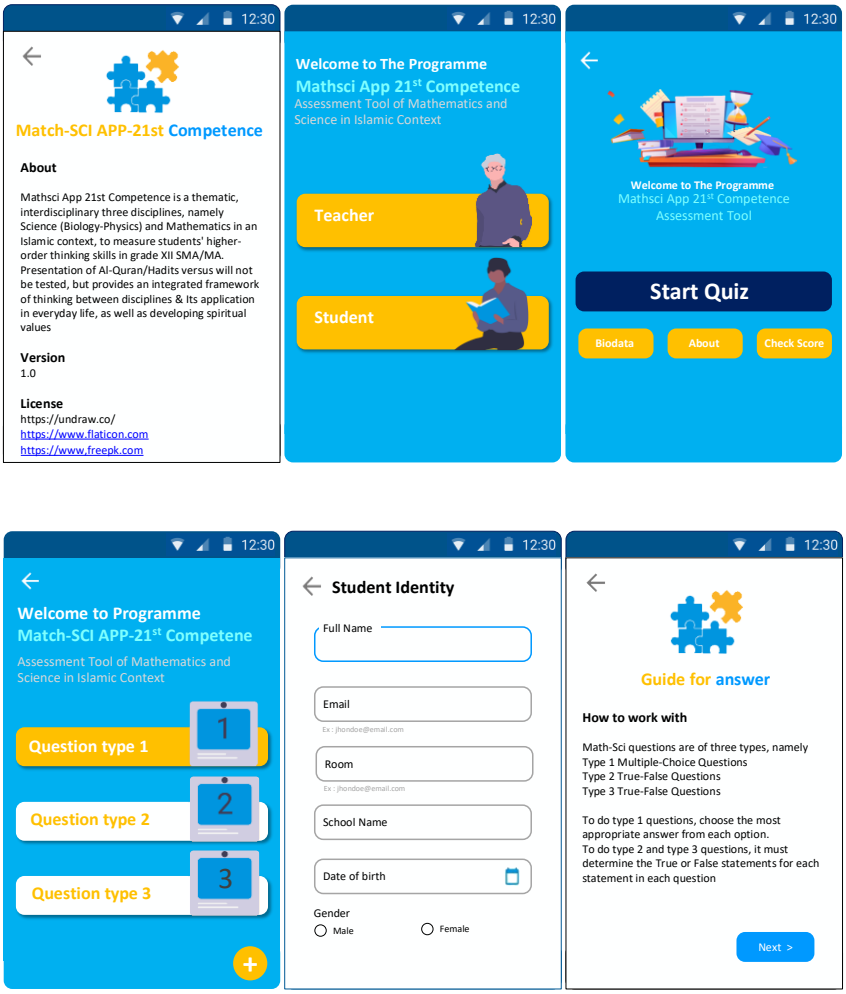

Fig. 3: MathSci 21st app

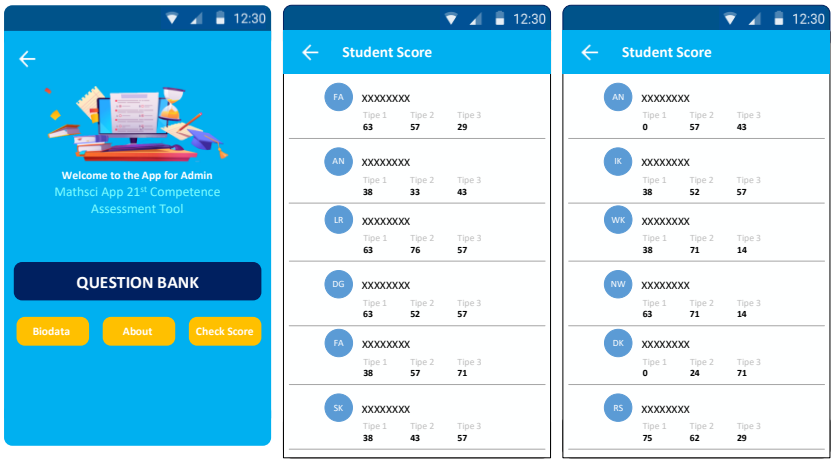

Fig. 4: Admin MathSci 21st app

3.2. Expert validation results

Expert validation was carried out by 10 validators, including 5 material experts, 3 media experts, and 2 language experts.

3.2.1. Material expert validation results

Material, construction, and language aspects are among the aspects examined in material validation. The material aspect is concerned with determining the suitability of the questions developed in conjunction with the Basic Competencies, indicators, and cognitive levels to be measured. The construction aspect seeks to determine the suitability of question construction in accordance with applicable rules, whereas the language aspect seeks to ensure that the language used is in accordance with good and correct Indonesian language rules. The results of the material expert validation were analyzed using the CVR (Content Validity Ratio) value stated by Lawshe (1975) (Table 3).

Table 3: Material expert validation results

MathSci question types	Aspects reviewed	Average CVR on each aspect	CVR value	Category
Type 1	A. Material	0.6	0.65	Appropriate
	B. Construction	0.62		
	C. Language	0.75		
Type 2	A. Material	0.4	0.74	Very appropriate
	B1. General construction	0.75		
	B2. Special construction	0.96		
	C. Language	0.87		
Type 3	A. Material	0.6	0.77	Very appropriate
	B1. General construction	0.62		
	B2. Special construction	1		
	C. Language	0.87		

Based on the CVR values in Table 2, the CVI (Content Validity Index) values are:

$$CVI = \frac{\Sigma\, CVR}{Number\ of\ questions}$$

$$CVI\ \frac{2.16}{3}$$

CVI=0.72

Thus, it can be concluded that the results of the material expert validation are categorized as very suitable with a value of 0.72.

3.2.2. Media expert validation results

Aspects of software engineering, visual communication, and operations are among those investigated in media validation. The software engineering aspect seeks to determine the accuracy and effectiveness of the software used to achieve the research objectives. The visual communication aspect assesses the appropriateness of the appearance, design layout, images, and application visualization. The operational aspect is concerned with determining the utility and ease of use in the field. The V index, proposed by Aiken (1985), was used to analyze the results of media expert validation (Tables 4 and 5). Based on index V in Table 4, the overall average of each aspect reviewed is shown in Table 5. The data in Table 5 shows that the overall average of all aspects examined in the media validation is categorized as Very High with a value of 0.83.

Table 4: Media expert validation results for each instrument item

No.	Aspects reviewed	Instrument item number	V (Rater agreement index)	Category
1.	Software engineering	1	0.75	High
		2	0.83	Very high
		3	0.75	High
		4	0.91	Very high
		5	0.75	High
2.	Visual communication	6	0.83	Very high
		7	0.75	High
		8	0.75	High
		9	0.75	High
		10	0.83	Very high
		11	0.83	Very high
		12	0.66	High
		13	0.83	Very high
3.	Operational	14	1	Very high
		15	1	Very high
		16	0.91	Very high
		17	0.83	Very high

Table 5: Average overall results of media expert validation

No.	Aspects reviewed	Average V (Rater agreement index)	Category
1.	Software engineering	0.80	Very high
2.	Visual communication	0.77	High
3.	Operational	0.93	Very high
	Overall average	0.83	Very high

3.2.3. Language expert validation results

Language validation examines straightforward, communicative, dialogic, and interactive aspects, conformity with student development, conformity with language rules, and the use of terms, symbols, or icons. The straightforward aspect seeks to assess the precision and effectiveness of the sentence structure employed. The communicative aspect seeks to ascertain the ease with which language or sentences are used. The dialogic and interactive aspects aim to assess the sentences' ability to motivate and encourage students to think. The aspect of suitability with student development aims to determine the accuracy of the sentences used with students' intellectual and emotional development. The aspect of conformity with language rules and the use of terms, symbols, or icons aims to determine the accuracy of grammar, terminology, and the use of symbols or icons with the structure of the questions. The V index proposed by Aiken (1985) was used to analyze the results of the validation of language experts (Tables 6 and 7). Based on index V in Table 6, the overall average of each aspect reviewed is shown in Table 7.

Table 6: Language expert validation results for each instrument item

No.	Aspects reviewed	Instrument item number	V (Rater agreement index)			Category		
			Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
1.	Straightforward	1	0.875	0.875	0.75	Very high	Very high	High
		2	0.875	0.875	0.75	Very high	Very high	High
		3	0.75	0.75	0.75	High	High	High
2.	Communicative	4	0.875	0.875	0.875	Very high	Very high	Very high
		5	0.875	0.875	0.875	Very high	Very high	Very high
3.	Dialogic and interactive	6	1	1	0.875	Very high	Very high	Very high
		7	1	1	1	Very high	Very high	Very high
4.	Compatibility with student	8	1	1	0.875	Very high	Very high	Very high
		9	0.875	0.75	0.875	Very high	High	Very high
5.	Development compatibility with language rules	10	0.625	0.75	0.875	High	High	Very high
		11	0.75	0.75	0.625	High	High	High
		12	0.75	0.875	0.75	High	Very high	High
6.	Use of terms, symbols, or icons	13	0.875	0.875	0.875	Very high	Very high	Very high
		14	1	1	1	Very high	Very high	Very high

Table 7: Overall average of language expert validation results

No.	Aspects reviewed	Average V (Rater agreement index)			Category		
		Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
1.	Straightforward	0.83	0.83	0.75	Very high	Very high	High
2.	Communicative	0.87	0.87	0.87	Very high	Very high	Very high
3.	Dialogic and interactive	1	1	0.93	Very high	Very high	Very high
4.	Compatibility with student development	0.93	0.87	0.87	Very high	Very high	Very high
5.	Compatibility with language rules	0.70	0.79	0.75	High	High	High
6.	Use of terms, symbols, or icons	0.93	0.93	0.93	Very high	Very high	Very high
	Overall average	0.88	0.88	0.85	Very high	Very high	Very high

The data in Table 7 shows that the overall average of all aspects examined in all types of questions is categorized as Very High. Type 1 and Type 2 question get a score of 0.88 while Type 3 questions get a score of 0.85.

3.2.4. One to one effectiveness trial results of MathSci 21st app–senior high school A Tangerang

The Pearson Correlation results for the three types of questions were classified as valid, and the questions' reliability was determined by Cronbach's Alpha of 0.646 (Strong Category). The following are the findings from the MathSci application trial at Senior High School A, including validity, reliability, discriminating power, and level of difficulty tests. Pearson Correlation was used to determine the validity of the questions for Question Type 1, Question Type 2/Interdisciplinary 1, and Question Type 2/Interdisciplinary 2 with N=32 (Tables 8–11). Other descriptive statistical studies describe the differentiability and level of difficulty of the 3 types of questions (Table 11).

Table 8: Validity of mono discipline question/ type 1

		Question 1	Question 2	Question 3	Total type 1	Interpretation
Question 1	Pearson correlation	1	.618**	.055	.793**	Valid
	Sig. (2-tailed)		.000	.764	.000	
Question 2	Pearson correlation	.618**	1	.049	.785**	Valid
	Sig. (2-tailed)	.000		.792	.000	
Question 3	Pearson correlation	.055	.049	1	.530**	Valid
	Sig. (2-tailed)	.764	.792		.002	
Total mono discipline	Pearson correlation	.793**	.785**	.530**	1	
	Sig. (2-tailed)	.000	.000	.002		

**: Correlation is significant at the 0.01 level (2-tailed)

Table 9: Validity of interdisciplinary questions 2/type 2

		Question 4	Question 5	Question 6	Total type 2	Interpretation
Question 4	Pearson correlation	1	.176	.066	.669**	Valid
	Sig. (2- tailed)		.336	.720	.000	
Question 5	Pearson correlation	.176	1	.162	.681**	Valid
	Sig. (2- tailed)	.336		.376	.000	
Question 6	Pearson correlation	.066	.162	1	.600**	Valid
	Sig. (2- tailed)	.720	.376		.000	
Total interdisciplinary 2	Pearson correlation	.669**	.681**	.600**	1	
	Sig. (2- tailed)	.000	.000	.000		

**: Correlation is significant at the 0.01 level (2-tailed)

Table 10: Validity of interdisciplinary questions 3/type 3

		Question 7	Total type 3	Interpretation
Question 7	Pearson correlation	1	.521**	Valid
	Sig. (2- tailed)		.002	
Total interdisciplinary 3	Pearson correlation	.521**	1	
	Sig. (2- tailed)	.002		

**: Correlation is significant at the 0.01 level (2-tailed)

Table 11: Discriminating power and difficulty level question of senior high school A

MathSci question types	Question number	Differentiating power score	Interpretation	Difficulty level score	Interpretation
Type 1 (monodiscipline)	1	0.75	Very good	0.56	Moderate
Type 1 (monodiscipline)	2	0.62	Very good	0.62	Moderate
Type 1 (monodiscipline)	3	0.31	Good	0.53	Moderate
Type 2 (interdiscipline 1)	4	0.20	Moderate	0.54	Moderate
Type 2 (interdiscipline 1)	5	0.16	Poor	0.44	Moderate
Type 2 (interdiscipline 1)	6	0.14	Poor	0.58	Moderate
Type 3 (interdiscipline 2)	7	0.28	Moderate	0.48	Moderate

The results showed that the differentiating power of monodisciplinary questions was in the good and very good categories. Interdisciplinary Question 1 category is moderate and less, while interdisciplinary questions are 2 moderate categories. Then, the difficulty level of the three types of questions is in the moderate category.

3.3. Questionnaire results for senior high school B students in south Tangerang

The results of the student questionnaire showed the "good" category in the aspects of the questions, the construction of the questions, and the implementation. Meanwhile, the technical quality aspect is in the "very good" category. The results of the questionnaire are shown in Table 12.

Table 12: Questionnaire results of senior high school B students

No.	Aspects reviewed	Mean	Interpretation
1.	Questions	3.76	Good
2.	Question construction design	3.96	Good
3.	Implementation	3.95	Good
4.	Technical quality	4.14	Very good
	Mean total	3.95	Good

3.4. Teacher questionnaire results

Aspects examined in the teacher's questionnaire include practical aspects (practicality) and aspects of effectiveness (effectiveness). The practical aspect aims to determine the ease and practicality of using the MathSci application in the field, whereas the effectiveness aspect aims to determine the accuracy and effectiveness of the MathSci application as an assessment application. According to the questionnaire results, teachers responded very well to both practicality and effectiveness (Table 13).

Table 13: Teacher questionnaire results

No.	Aspects reviewed	Average score
1.	Practicality	4.6
2.	Effectivity	4.42
	Mean	4.11
	interpretation	Very good

4. Discussion

According to the findings of the research, the MathSci 21 app program included the MathSci admin program as a user data store. The MathSci 21 App is a computer program designed to measure higher-order thinking skills as a competency assessment tool/tool for Mathematics and Natural Sciences (Mathematics and Natural Sciences) Islamic Context at the Senior High School level. A student or teacher menu is available in the "Home" view of the application. This separation is meant to make it easier to store user information. Students and teachers will be directed to sign up and provide an active email address. Before beginning to work on the questions, the next step is to complete the biodata. The user can then proceed by clicking "Start quiz" to begin working on the questions. There are three "Types of questions" displayed in order (Type 1-Type 2-Type 3).

The MathSci application includes three types of questions: Type I (monodisciplinary questions) with three questions (biology, physics, and mathematics) in the Islamic context, Type II (interdisciplinary questions two disciplines) with three questions (biology-physics integration, biology-mathematics, and mathematics-physics) in the Islamic context, and Type III (interdisciplinary 2) with one question (integration of biology-physics-mathematics) in the Islamic context. Monodisciplinary questions are multiple choice with 5 options, whereas interdisciplinary 1 and interdisciplinary 2 questions

are True-False with 7 true and false choice statements per question.

The material expert validation results show that the questions in the MathSci application are classified as very appropriate, with a score of 0.77. Material, construction, and language aspects are all examined in material validation. With a score of 0.83, the overall mean of all aspects examined in the media validation is classified as Very High. Media validation looks at aspects of software engineering, visual communication, and operations. The language validation results show that the overall average of all aspects examined in all types of questions is classified as Very High. Question types 1 and 2 received a score of 0.88, while Question type 3 received a score of 0.85.

Thus, the results of the material, media, and language expert validation show that the questions in the MathSci app application are very appropriate and very high. Pearson correlation results for the three types of questions are classified as valid and reliable (Cronbach Alpha 0.646). The dependability of the questions in the MathSci application demonstrates stability, consistency, and high measurement accuracy. Reliability demonstrates that the instrument can be relied on as a data collection tool (Heale and Twycross, 2015) and can capture actual information in the field Krüger et al. (2020) reported on an investigation into several aspects of the validity of test score interpretation of scientific reasoning competency tests, as well as aspects of reliability. The first phase of the project focused on the development of a paper-pencil assessment instrument, the KoWADiS competency test, for the longitudinal assessment of pre-service science teachers' scientific reasoning competencies during academic studies. Investigate the dependability of test scores and the validity of their interpretation in the second phase. The research employs a multi-method approach in order to address multiple sources of valid evidence. Overall, the results are coherent and provide adequate support for the validity assumptions.

The student questionnaire results indicated a "good" category in the aspects of the questions, the construction of the questions, and the implementation. Meanwhile, the technical quality is in the "very good" range. As a result, this application could be used and tested on a larger scale. The construction of MathSci questions is prepared by adjusting to the demands of basic competencies at the Senior High School/MA Curriculum 2013 level. According to the Islamic context, the development of the questions integrates two or three related disciplines (Mathematics and Science) (Fig. 5).

Interdisciplinary 1

1. Every Muslim is obligated to maintain good health. This is consistent with Rasulullah SAW's statement, "Indeed, your body has rights over you" (HR. Muslim). This order is always carried out by Nadya, a class X student. She monitors her heart health on a regular basis. Based on her medical records, the following is Nadya's blood pressure measurement data.

Name: Nadia Gender/Age: Female/16 years old	**Month**	**Blood**
	July	
	August	
	September	

According to the American Heart Association, the normal blood pressure range for 14–19-year-olds is 117/77 mmHg. Doctors have been observing and advising Nadya to exercise regularly and eat a healthy diet despite the fact that she has no symptoms since August. Do you think the following statements are true or false based on this information?

Statement	**True**	**False**
A. If not controlled, this condition can increase the risk of disease or complications such as stroke		
B. The blood pressure on this information will be equivalent to 118.420 Pascal/18.421 Pascal		
C. If the diameter of the blood vessel is 10 μm, then the magnitude on the pushing force from the flowing blood is 4.6×10^{-5} N to 2.96×10^{-4} N		

Fig. 5: The questions interdisciplinary 1

The stimulus dimension of the problem in the Islamic context gives students real-world experience with the application of knowledge in everyday life. A problem will lead to knowledge about a specific object, but it will also aid in the development of broad concepts with broad explanatory power, allowing new objects or events to be understood (Harlen, 2013). In this view, the most important learning outcomes are not only students' ability to construct new knowledge and direct their own learning but also the development of lifelong learners (Fullan and Langworthy, 2014).

A tested framework in the pencil-paper test format Zulfiani et al. (2021) and digital technology support in the form of application programs that are in line with the changing paradigm of 21st-century assessment and learning (Barak, 2017) in the form of a creative infusion of technology are used to

construct innovation in the MathSci app (Caena and Redecker, 2019). It is critical to help students take ownership of their learning through ongoing assessment and reflection on their progress. Learning Technology, driven by the use of technology, has the potential to develop creative and collaborative participants in a knowledge-based and interdependent world, as does such Learning Technology (NRC, 2012).

The MathSci application can quickly determine student competency achievement in stages. The framework is organized in the form of a ladder, with type 1 questions serving as a diagnostic test, type 2 questions serving as interdisciplinary from two disciplines, and type 3 questions serving as interdisciplinary from three disciplines. Type 2 assesses higher-order thinking skills of moderate complexity, while type 3 assesses higher-order thinking skills of high complexity. MathSci, an Android-based program, is more efficient and effective in supporting paperless, data-driven, and user-friendly learning. MathSci also accelerates the assessment of higher-order thinking skills, allowing teachers to provide feedback and analyze future learning more quickly. Feedback has emerged as a major focus of research and teaching practice (Wisniewski et al., 2020). The shift in education toward blended and online learning with problem-based and inquiry-based approaches prompted thought about technologies that can effectively support formative assessment and informative feedback for 21st-century learners (Spector et al., 2016).

The MathSci application is compatible with Android devices that have 6 MB of memory. This assessment application program also familiarizes digitally literate students with problem-solving. All data is managed in the MathSci admin application, which can record all student and teacher work results (based on big data). We live in the era of big data, where data is generated every second. The main challenges of today's organizations are data that is more varied and has a very complex structure, with problems of indexing, sorting, searching, analyzing, and visualizing (Zhang et al., 2021). The study conducted by Balaji et al. (2022) demonstrates the impact of promoting AI, IoT, and Social Business for higher education students by providing a platform for aspiring Engineers and Science students to think about and apply new ideas to solve societal problems. To train students in application areas, the authors used AI, IoT-based Social Business Models, and STEMSEL (Science Technology Engineering Mathematics Social Enterprise Learning) Programming techniques. The use of these tools to create prototypes and working models for applications in the teaching and learning process can lead to the development of rapid online learning methods (Huda et al., 2018). The use of MathSci received positive responses from both students and teachers. It is consistent with the results of Afriyanti et al. (2021) which found that the design of e-modules is very suitable for stimulating HOTS and can be carried out at the development stage of STEM-based interactive e-module to stimulate students' HOTS on static fluid material. The MathSci application facilitates assessment for learning, which has the potential to improve learning quality and is expected to provide feedback to teachers in order to redesign classroom learning.

5. Conclusions

The journey of the MathSci application through the various stages of research and development has unequivocally demonstrated its viability as an Android-based tool. Meticulously crafted as a sophisticated competency assessment instrument, this application boasts a distinctive characteristic—its integration of contextual dimensions—an aspect that accentuates its pertinence and relevance within educational settings. The strategic choice of an Android-based framework for its deployment bears notable advantages, most notably its reduction of paper dependency and its familiarity with a wide user base. This sophisticated technological intervention stands poised to revolutionize the assessment landscape, particularly in its potential to facilitate expedient feedback provision and the meticulous analysis of learning trajectories.

The advent of this application heralds a noteworthy innovation within the realm of learning methodologies, thereby significantly contributing to the realization of the principle of continuous improvement. By affording educators and learners alike a seamless and efficient assessment ecosystem, the MathSci application underscores the imperative of adopting progressive and adaptive educational approaches. As contemporary education increasingly gravitates towards adaptive and technologically-integrated paradigms, the MathSci application emerges as a pivotal exemplar of harnessing digital advancements to foster elevated pedagogical outcomes.

In its capacity to enhance not only the efficiency but also the efficacy of assessment practices, this innovation is poised to reverberate across educational contexts. The infusion of technology into the assessment process augments the precision and timeliness of feedback, which, in turn, serves as a catalyst for informed pedagogical adjustments. Consequently, the MathSci application represents a transformative milestone, aligning seamlessly with the ethos of continual enhancement that underpins modern educational paradigms. As education navigates a landscape characterized by rapid technological evolution, the MathSci application stands as an emblematic model of how digital innovations can harmoniously blend with pedagogical principles to drive the inexorable march toward educational excellence.

Acknowledgment

The research team would like to express their gratitude for the support of PUSLITPEN LP2M UIN

Syarif Hidayatullah Jakarta for the applied research grant for higher education collaboration No. UN 01/KPA/228/2022.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Afriyanti M, Suyatna A, and Viyanti (2021). Design of e-modules to stimulate HOTS on static fluid materials with the STEM approach. Journal of Physics: Conference Series, 1788: 012032. https://doi.org/10.1088/1742-6596/1788/1/012032

Aiken LR (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45(1): 131-142. https://doi.org/10.1177/0013164485451012

Al Aliyawinata TT, Utari E, and Mahrawi M (2021). The effect of discovery learning on students' higher-order thinking skills. International Journal of Biology Education Towards Sustainable Development, 1(1): 1-9. https://doi.org/10.53889/ijbetsd.v1i1.47

Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, and Wittrock MC (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, New York, USA.

Andriyatno I, Zulfiani Z, and Mardiati Y (2023). Higher order thinking skills: Student profile using two-tier multiple choice instrument. International Journal of STEM Education for Sustainability, 3(1): 111-124. https://doi.org/10.53889/ijses.v3i1.79

Astuti Y, Asmar A, Musdi E, and Yerizon (2023). Development of mathematics e-module using scientific approach integrated Islamic values for integrated Islamic junior high school. AIP Conference Proceedings, 2698(1): 060003. https://doi.org/10.1063/5.0122559

Balaji K, Selvam M, Rajeswari R (2022). Impact of artificial intelligence (AI), Internet of Things (IoT) and STEM social enterprise learning based applications in the teaching and learning process of engineering education. In: Kumar A, Senatore S, and Gunjan VK (Eds.), ICDSMLA 2020: Lecture notes in electrical engineering: 1217–1226. Volume 783, Springer, Singapore, Singapore. https://doi.org/10.1007/978-981-16-3690-5_116

Barak M (2017). Science teacher education in the twenty-first century: A pedagogical framework for technology-integrated social constructivism. Research in Science Education, 47: 283-303. https://doi.org/10.1007/s11165-015-9501-y

Beniermann A, Mecklenburg L, and zu Belzen AU (2021). Reasoning on controversial science issues in science education and science communication. Education Sciences, 11(9): 522. https://doi.org/10.3390/educsci11090522

Birenbaum M, DeLuca C, Earl L, Heritage M, Klenowski V, Looney A, and Wyatt-Smith C (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. Policy Futures in Education, 13(1): 117-140. https://doi.org/10.1177/1478210314566733

Budiarti IS and Tanta T (2021). Analysis on students' scientific literacy of Newton's law and motion system in living things. Jurnal Pendidikan Sains Indonesia [Indonesian Journal of Science Education], 9(1): 36-51. https://doi.org/10.24815/jpsi.v9i1.18470

Caena F and Redecker C (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (*DIGCOMPEDU*). European Journal of Education, 54(3): 356-369. https://doi.org/10.1111/ejed.12345

Clark I (2011). Formative assessment: Policy, perspectives and practice. Florida Journal of Educational Administration and Policy, 4(2): 158-180.

Dalby D and Swan M (2019). Using digital technology to enhance formative assessment in mathematics classrooms. British Journal of Educational Technology, 50(2): 832-845. https://doi.org/10.1111/bjet.12606

Darling-Aduana J (2020). High school student experiences and learning in online courses: Implications for educational equity and the future of learning. Ph.D. Dissertation, Vanderbilt University, Nashville, USA.

Fensham PJ and Bellocchi A (2013). Higher order thinking in chemistry curriculum and its assessment. Thinking Skills and Creativity, 10: 250-264. https://doi.org/10.1016/j.tsc.2013.06.003

Fullan M and Langworthy M (2014). A rich seam: How new pedagogies find deep learning. Technical Report, Pearson, London, UK.

Harlen W (2013). Assessment and inquiry-based science education: Issues in policy and practice. InterAcademy Partnership (IAP), Trieste, Italy.

Haryadi R and Pujiastuti H (2022). Enhancing pre-service physics teachers' higher-order thinking skills through STEM-PJBL model. International Journal of STEM Education for Sustainability, 2(2): 156-171. https://doi.org/10.53889/ijses.v2i2.38

Heale R and Twycross A (2015). Validity and reliability in quantitative studies. Evidence-Based Nursing, 18(3): 66-67. https://doi.org/10.1136/eb-2015-102129 **PMid:25979629**

Huda M, Maseleno A, Atmotiyoso P, Siregar M, Ahmad R, Jasmi K, and Muhamad N (2018). Big data emerging technology: Insights into innovative environment for online learning resources. International Journal of Emerging Technologies in Learning, 13(1): 23-36. https://doi.org/10.3991/ijet.v13i01.6990

Johar R and Ahmad A (2018). The quality of learning materials through mathematics reaslitic to improve students' mathematical communication ability in the elementary school. Journal of Physics: Conference Series, 1088(1): 012077. https://doi.org/10.1088/1742-6596/1088/1/012077

Jufrida J, Basuki FR, Kurniawan W, Pangestu MD, and Fitaloka O (2019). Scientific literacy and science learning achievement at junior high school. International Journal of Evaluation and Research in Education, 8(4): 630-636. https://doi.org/10.11591/ijere.v8i4.20312

Krüger D, Hartmann S, Nordmeier V, and Upmeier zu Belzen A (2020). Measuring scientific reasoning competencies: Multiple aspects of validity. In: Zlatkin-Troitschanskaia O, Pant HA, Toepper M, and Lautenbach C (Eds.), Student learning in German higher education: Innovative measurement approaches and research results: 261–280. Springer VS, Wiesbaden, Germany. https://doi.org/10.1007/978-3-658-27886-1_13

Kurniawati L, Miftah R, Kadir K, and Muin A (2021). Student mathematical literacy skill of madrasah in Indonesia with Islamic context. TARBIYA: Journal of Education in Muslim Society, 8(1): 108-118. https://doi.org/10.15408/tjems.v8i1.3184

Lawshe CH (1975). A quantitative approach to content validity. Personnel Psychology, 28(4): 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Le B, Lawrie GA, and Wang JT (2022). Student self-perception on digital literacy in STEM blended learning environments. Journal of Science Education and Technology, 31(3): 303-321.

https://doi.org/10.1007/s10956-022-09956-1
PMid:35132301 PMCid:PMC8809496

Leenknecht M, Wijnia L, Köhlen M, Fryer L, Rikers R, and Loyens S (2021). Formative assessment as practice: The role of students' motivation. Assessment and Evaluation in Higher Education, 46(2): 236-255. https://doi.org/10.1080/02602938.2020.1765228

Nieveen N (2007). Educational design research. In: Van den Akker J, Gravemeijer K, and McKenney S (Eds.), Educational design research. Volume 2, Routledge, London, UK.

Nihayati N, Khoiriyah S, Nurmitasari N, and Kayyis R (2022). Mathematics teaching materials of set integrated with Islamic values. International Journal of Trends in Mathematics Education Research, 5(2): 174-179. https://doi.org/10.33122/ijtmer.v5i2.152

NRC (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. National Research Council, National Academies Press, Washington, USA.

OECD (2015). PISA 2015: Draft science framework. Organisation for Economic Co-operation and Development, Luxembourg, Luxembourg.

Osborne J (2013). The 21st century challenge for science education: Assessing scientific reasoning. Thinking Skills and Creativity, 10: 265-279. https://doi.org/10.1016/j.tsc.2013.07.006

Pujiastuti H and Haryadi R (2023). Higher-order thinking skills profile of Islamic boarding school students on geometry through the STEM–based video approach. International Journal of STEM Education for Sustainability, 3(1): 156-174. https://doi.org/10.53889/ijses.v3i1.135

Putri RII and Aisyah N (2020). Learning integers with realistic mathematics education approach based on Islamic values. Journal on Mathematics Education, 11(3): 363-384. https://doi.org/10.22342/jme.11.3.11721.363-384

Rahmah N, Lestari A, Musa LAD, and Sugilar H (2019). Quizizz online digital system assessment tools. In the IEEE 5th International Conference on Wireless and Telematics, IEEE, Yogyakarta, Indonesia: 1-4. https://doi.org/10.1109/ICWT47785.2019.8978212

Spector JM, Ifenthaler D, Sampson D, Yang JL, Mukama E, Warusavitarana A, and Gibson DC (2016). Technology enhanced formative assessment for 21st century learning. Journal of Educational Technology and Society, 19(3): 58-71.

Sullivan P, McBrayer JS, Miller S, and Fallon K (2021). An examination of the use of computer-based formative assessments. Computers and Education, 173: 104274. https://doi.org/10.1016/j.compedu.2021.104274

Weurlander M, Söderberg M, Scheja M, Hult H, and Wernerson A (2012). Exploring formative assessment as a tool for learning: Students' experiences of different methods of formative assessment. Assessment and Evaluation in Higher Education, 37(6): 747-760. https://doi.org/10.1080/02602938.2011.572153

Wisniewski B, Zierer K, and Hattie J (2020). The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10: 3087. https://doi.org/10.3389/fpsyg.2019.03087
PMid:32038429 PMCid:PMC6987456

Zhang Y, Geng P, Sivaparthipan CB, and Muthu BA (2021). Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. Sustainable Energy Technologies and Assessments, 45: 100986. https://doi.org/10.1016/j.seta.2020.100986

Zulfiani Z, Suwarna IP, and Muin A (2021). Framework and prototype development of MathSci instruments for measuring 21st century skills in Islamic context. TARBIYA: Journal of Education in Muslim Society, 8(1): 96-107. https://doi.org/10.15408/tjems.v8i1.22120