Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Usability Testing Hybrid Method for Indonesian Automatic Question Generator

Handaru Jati[1], Yuniar Indrihapsari[1,*], Stanislaus Jiwandana Pinasthika[1], Susilo Romadhon[1], Danang Wijaya[2], Dhista Dwi Nur Ardiansyah[1], Satya Adhiyaksa Ardy[1], Widya Ardianto[1], Zainal Ma'ruf Abidin[1], Damar Albaribin Syamsu[1], Muhammad Naufal Khoirul Imamilhaq Alhifdi[1]

[1] Department of Electronics and Informatics Engineering Education, Faculty of Engineering, Universitas Negeri Yogyakarta, Sleman, Yogyakarta 55281, Indonesia
[2] International Graduate Program in Artificial Intelligence, College of Electrical Engineering and Computer Science, National Central University, Taoyuan City, Taiwan

ARTICLE INFO

Article history:
Received 7 February 2025
Received in revised form 10 March 2025
Accepted 30 June 2025
Available online 20 July 2025

Keywords:
Question bank; TAM; ISO 25023; learning assessment; user centred tool

ABSTRACT

As educational technology becomes increasingly essential, usability testing is crucial to ensure IT products meet user needs, particularly in the context of teaching and learning. The Indonesian Automatic Question Generator (IAQG) was developed to assist educators in creating question banks tailored to their curricula. This study evaluates the IAQG usability and user acceptance from the perspective of Indonesian vocational high school teachers, using a hybrid methodology that combines the Technology Acceptance Model (TAM) and ISO 25023 standards. The results indicate that while the IAQG is generally well-received, with high marks for user interface design and operational consistency, improvements are needed in areas such as print functionality and language support. The ANOVA analysis of TAM indicates that perceived behaviour control and subjective norm influence teachers' attitude toward use and influence their behavioural intention to use the IAQG. The study concludes that the IAQG has significant potential to alleviate teachers' workloads and improve the quality of learning outcome assessments. However, the respondents raised several suggestions toward development to meet their needs and expectations during the IAQG usage. This research contributes to develop more effective, user-centred educational tools that align with the specific needs of Indonesian educators.

1. Introduction

Present-day teachers face more complex challenges and one of the factors is the massive implementation of information and communication technology (ICT) [1]. Teachers were demanded to be creative to involve ICT in their teaching method while updating their knowledge as young students nowadays might have understood world knowledge of ICT's impact of ICT [4]. Instead of challenging and fun, it burdens Indonesian teachers who must deal with the demands of education stakeholders, administrative work, poor work conditions and shifting government policies on

* *Corresponding author*
E-mail address: yuniar@uny.ac.id

https://doi.org/10.37934/ard.138.1.137150

national education [2]. All these loads might cause a decrease in teacher quality, even giving them no time to generate their test questions. In contrast, comprehensive learning evaluation is the most important way to measure learning outcomes.

Teachers usually use AI-based tools like ChatGPT to overcome the problem and generate exam questions. However, the results sometimes need to meet the expectations as they bring no context and do not meet the learning outcomes. Most question generators have been built for English questions incompatible with Indonesian learning modules. So, in the previous study, we developed the Indonesian Automatic Question Generator (IAQG) based on the Indonesian Bidirectional Encoder Representations from the Transformers (IndoBERT) method. This new product has not been tested yet. We need to evaluate our product from the teachers' point of view and determine how teachers will accept the tool.

One of the most crucial aspects to be evaluated is usability. This is a fundamental thing to be fulfilled first as the user can assess whether the application is feasible to use or not based on their first use experience. For teachers, ideal software should be practical and easily integrated into teaching practice [14]. In addition, the generator should reduce the possibility of unanswered questions emerging [20]. Hence, IAQG should fulfil teachers' needs in learning evaluation, for instance, by accommodating many kinds of question formats.

In addition, employed IT applications might be challenging for learning activities as it needs various preparation, including adaptation to new teaching environments [23]. Improper IT application will waste so much time to be adapted and decrease students' learning motivation and self-discipline.

Several studies have been used to evaluate the educational applications [9,12,19,22]. Considering the usability and workload aspects, a study implemented the IBM Computer Usability Satisfaction Questionnaires (CSUQ) and the NASA Task Load Index (TLX) for evaluating a game-like engineering education application. The writers concluded that successful educational applications give good first impression and high levels of satisfaction to users [5]. Other research developed the CECAM survey to evaluate the usability of a learning application containing 56 items on pedagogical usability and user interface usability [18]. CECAM was designed to evaluate typical multimedia learning applications. One of the famous usability models is the Technology Acceptance Model (TAM) [6]. TAM was used to assess the utilization of social media for educational purposes. This study, which combined TAM with Task-Technology Fit (TTF) [11], evaluated user satisfaction and its impact on social media, both in academic and social aspects [1]. However, CECAM needs to be tested by more participants to reach certain levels of validity. In addition, choosing the proper dataset to train and test the automatic question generator model influences the model performance [25].

Several approaches, mostly sequence-to-sequence architecture models, have been proposed to build AQG in Indonesian, such as IndoBART [26], multilingual Text-to-Text Transfer Transformer (mT5) [10], mT5 with three different input schemes: paragraph, sentence and top 3 most relevant sentences [3], sequence-to-sequence model using CopyNet (GRU and Bi-GRU as encoder and LSTM is used as decoder) [13] and BiGRU, BiLSTM, Transformer [17]. Due to the lack of a large dataset that can be used for Automatic Question Generator, most researchers in Indonesia use the translated SQuAD v2.0 factoid question answering dataset, with additional Indonesian TyDiQA dev set for testing [21].

To comprehensively evaluate this brand-new question generator, we propose a novel approach that combines ISO 25023 standards with the TAM. We employed ISO25023 to evaluate all IAQG features and TAM for teacher acceptance through this application. The hybrid methodology allows us to identify the specific features teachers need from an automatic question generator while measuring their perceptions and likelihood of adopting the IAQG in their work.

The objectives of this study are threefold:

i. To evaluate the usability of the IAQG using ISO 25023 standards from the perspective of Indonesian teachers.
ii. To assess teachers' acceptance and intention to use the IAQG through the lens of the Technology Acceptance Model.
iii. To identify key features and improvements necessary to enhance the IAQG's effectiveness and adoption in Indonesian educational settings.

By addressing these objectives, this research aims to contribute to developing more effective, user-centred educational technology tools that can significantly alleviate teachers' workloads while maintaining high standards of assessment quality.

2. Methodology

2.1 Indonesian Automatic Question Generator (IAQG)

An AQG is a model or function that automatically produces several questions from a given text. Research on AQG deployment has been conducted in Indonesia for over ten years [17]. However, AQG research in Indonesia was improperly conducted because it lacked an Indonesian-educated quiz dataset [7]. Thus, the researchers must build the dataset manually.

The preprocessing started by separating the dataset from the paragraph context and the filtered question-answer [15]. Both pairs were assigned a number for reference and translated using an API such as Google Translate API. The researchers sorted out the unanswered questions and started to train the model.

Several approaches, mostly sequence-to-sequence architecture models, have been proposed to build AQG in Indonesian, such as BiGRU, BiLSTM and Transformer [17]. The sequence-to-sequence architecture has been designed following the task of translation, which consists of encoder and decoder, which are trained separately. The encoder includes several sub-encoders and is responsible for retrieving the subject atom, relationship atom and object atom, which are embedded as one fact and become an input for the decoder. A study revealed that the uncased version of the mentioned sequence-to-sequence models was more prominent than the cased version. The generated questions seemed like the expected questions. However, this implementation needs to be assessed by the teacher as the primary user of IAQG.

2.2 ISO/IEC 25023

To facilitate information system evaluation, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) developed ISO/IEC 25023 [8]. It represents all the standards modern ICT must fulfil to ensure and improve software products in all software development life cycles [4].

The introduction of ISO/IEC 25023 expanded the scope of evaluation by adding usability to the existing measurement metrics, which include system performance, reliability, security and maintainability. Usability, in this context, refers to how easily an ICT system can be learned and operated and how visually appealing it is to users. It is assessed across six key attributes, as detailed in Table 1.

Table 1
Usability attributes

Usability Attributes	Sub-attributes
Appropriateness recognizability	Description completeness, Demonstration coverage, Entry points self-descriptiveness
Learnability	User guidance completeness, Entry fields defaults, Error message understandability, Self-explanatory user interface
Operability	Operational consistency, Message clarity, Functional customizability, User interface customizability, Monitoring capability, Undo capability, Understandable categorization of information, Appearance consistency, Input device support
User error protection	Avoidance of user operation error, User entry error correction, User error recoverability
User interface aesthetics	Appearance aesthetics of user interfaces
Accessibility	Accessibility for users with disability, Supported languages adequacy

Usability attributes can be effective in evaluating an ICT system if all attributes are assessed positively in the system evaluation. It indicates the ICT system is worth to be used by users [24]. Evaluation can be done through direct testing with users, surveys or heuristic analysis by usability experts.

In this study, every sub-attribute was assessed using the Guttman scale, which has only two options, positive or negative, for each feature. We asked vocational high school teachers to try IAQG and fill in the instrument. The collected data were assessed to obtain the score of each sub-attribute. The attribute calculation function can be seen in Eq. (1):

$$Sub - attributes = \frac{Positive\ functions}{Total\ functions} \quad (1)$$

The calculated data will show whether IAQG is feasible or unsuitable and which features need improvement. As the metric is only suitable for usability, we combined it with the TAM to obtain and measure the acceptance aspects.

2.3 The Technology Acceptance Model

The TAM is a theoretical framework used to understand and predict user acceptance and usage of technology [4,8]. TAM is designed based on three factors influencing users to continue operating the system: cultural consideration, educational context and technology infrastructure. The factors are explained by measuring five aspects, as illustrated in Figure 1.

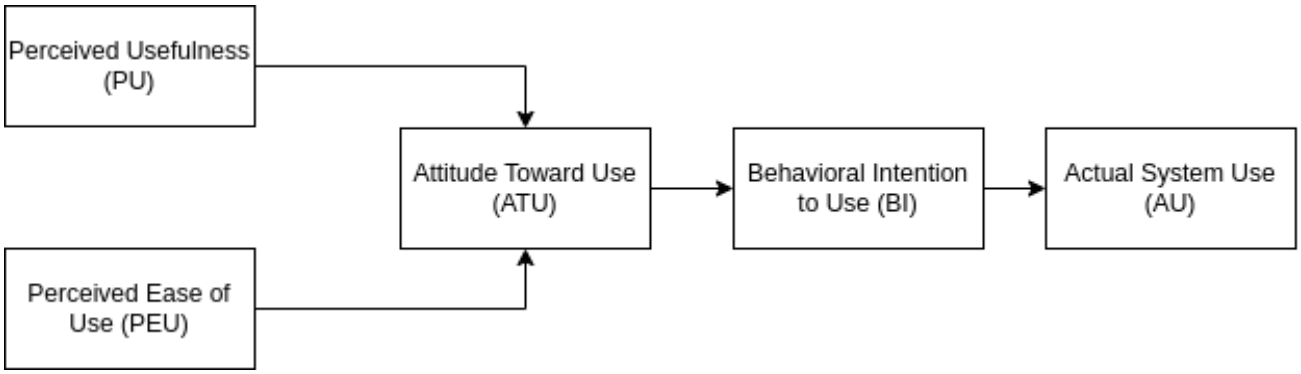

Fig. 1. TAM diagram [6]

Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) are fundamental aspects that influence users' decisions to continue or cease usage. Perception will influence users' attitudes and behavioural intentions and their assessment of the actual use [4,8]. Perception usually depends on the gender, age and experience of the users.

The IAQG has been developed, so the actual use aspect cannot be assessed. Therefore, we employed the revised TAM with additional conceptual frameworks such as Perceived Behaviour Control (PBC), Subjective Norm (SN), Personal Innovativeness (PIIT) and Behaviour Intention (BI) [16]. The questions about actual usage behaviour (AU) have been replaced by PIIT to obtain an interest in reimplementing IAQG. The expected benefits of IAQG based on the revised TAM are shown in Table 2. However, we assume all the IAQG users are educators, so we took areas of expertise as the only dependent factor.

Table 2
Expected benefits of IAQG based on the revised TAM attributes

TAM Attributes	Expected Benefit of IAQG
Perceived Usefulness (PU)	The generated questions are easy to understand and able to be answered.
Perceived Ease of Use (PEU)	The IAQG has a user-friendly interface with clear instructions, intuitive navigation and quick access to features.
Perceived Behaviour Control (PBC)	Ease of use encourages users to do a positive behaviour.
Subjective Norm (SN)	Ease of use encourages users to recommend IAQG to their colleagues.
Behavioural Intention to Use (BI)	Comprehensive training session and continuity of development strengthen the intention to use.
Attitude Towards Use (ATU)	Positive experience during using IAQG influences users to recommend IAQG to their relatives.
Personal Innovativeness (PIIT)	The characteristic of ingenious user increases certain skills that able to deal with certain challenge while operating new IT thing.

The TAM attributes are translated into 40 questions, with 4 to 9 questions for each aspect. Moreover, we asked the same respondents to fill out the questionnaire. To calculate the perceptions, we determined several hypotheses to test the correlation between the relevance aspects, deducing the following hypotheses (H):

H1: PU of IAQG correlated to ATU of IAQG
H2: PEU of IAQG influences to ATU of IAQG
H3: PBC influences to ATU of IAQG
H4: SN correlated to ATU of IAQG
H5: PIIT correlated to ATU of IAQG
H6: ATU of IAQG influences BI

Figure 2 illustrates the presumed relationships between the hypotheses. Hypotheses H1 through H5, shown on the left side of the figure, are expected to influence Attitude Towards Use (ATU), which is related to BI. We employed Pearson correlation and descriptive statistics to analyse these relationships and calculate the correlations between the relevant variables.

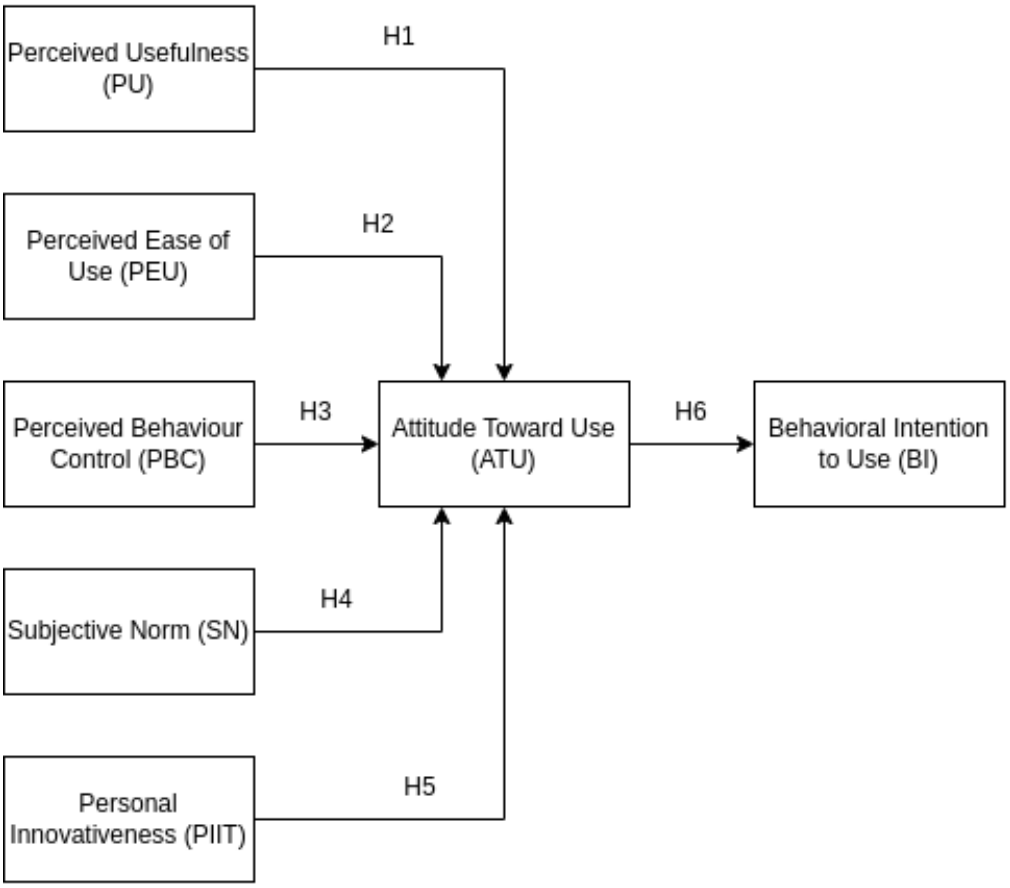

Fig. 2. The correlation between TAM attributes in this study

3. Results

3.1 Data Analysis

This study population is SMK teachers in Yogyakarta, Indonesia, who deliver subject-specific material in their fields. Several criteria were used to select participants to ensure the relevance and diversity of the data collected. They were asked to be active SMK teachers with experience designing and auditing students' learning outcomes, showing them the educational process firsthand. Teachers from multiple fields, including informatics, engineering, finance, mathematics, languages and social studies, were included to assess the applicability across various disciplines. Participation was optional, with teachers self-nominating or being named by their institutions. Participants were invited to the survey after confirming their willingness and availability. A different level of technological proficiency for educational purposes was more desirable as the tool needed to be adequately interacted with.

SMKN 1 Bantul is chosen for the teachers as representatives from a larger field of disciplines, as well as the related field discipline of teachers. SMK teachers are more experienced in information technology than other practitioners as they used to utilize various tools, especially IT applications. Such a decision is consistent with the TAM, which has the dimension of subjective norms and behavioural intentions, which shows that the ease of use of the IT applications influences people to learn it more intense and influence other people to use. Thus, we needed a school with diverse programs and has strong IT background to ensure all participants could provide useful feedback on the tool. All these considerations led us to choose SMKN 1 Bantul's teachers as our participants.

The subjects were selected from both technical and non-technical disciplines and purposive sampling was used; the demographic diversity of vocational schools in Yogyakarta is considered to balance age, gender and teaching experience.

We developed an instrument comprising 62 questions: 21 based on ISO 25023 and 40 derived from the revised TAM. This instrument was administered to 20 vocational high school teachers in Yogyakarta, Indonesia. Most respondents specialized in informatics engineering, followed by those in accounting and finance. Other participants came from diverse fields, including electrical engineering, civil engineering, English, Javanese, civic education, mathematics, office management, business and marketing and adaptive normative subjects. The distribution of respondents and their areas of expertise are presented in Figure 3.

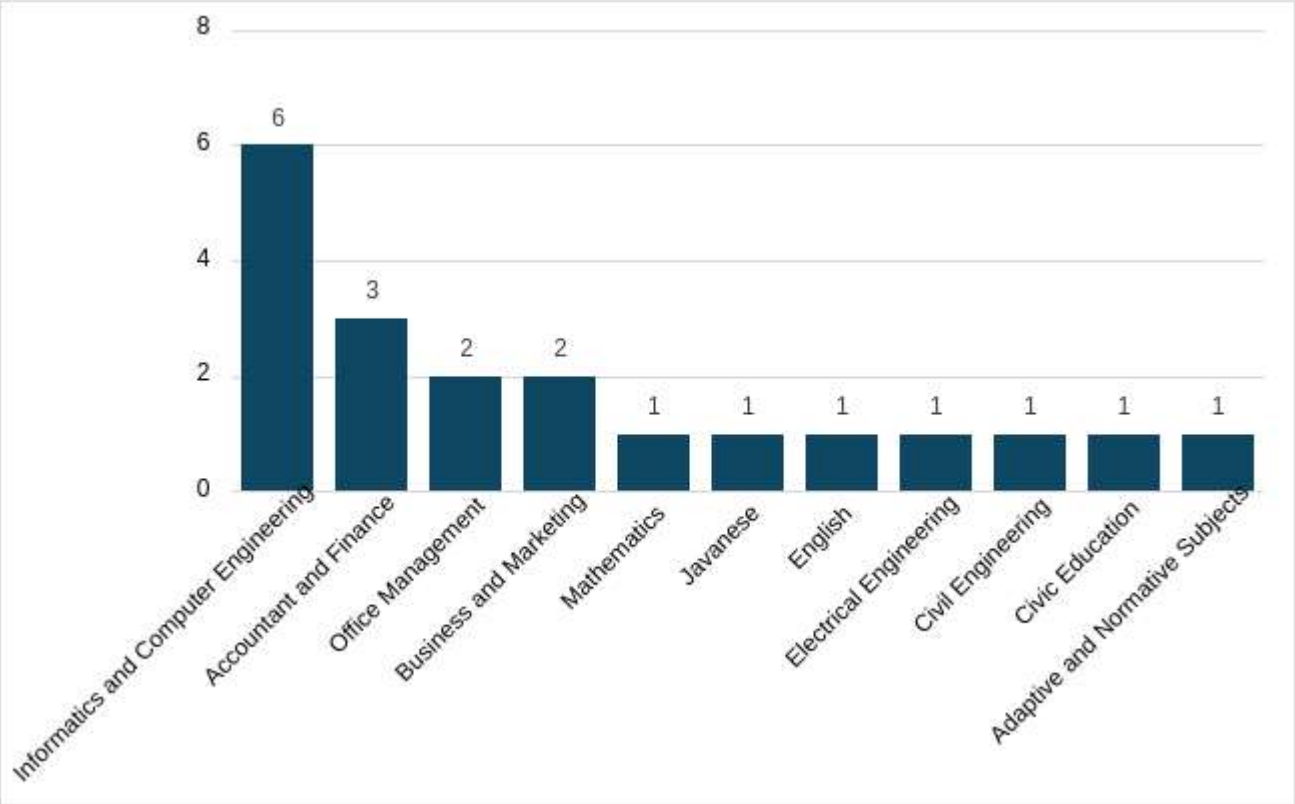

Fig. 3. Demographics of respondents

3.2 ISO/IEC 25023 Analysis

The respondents were asked to try the IAQG and then complete a questionnaire. The first part of the questionnaire was based on ISO 25023, focusing on the presence of several key features. We evaluated six main features: sign-in (F1), log-in (F2), question generator (F3), display generated questions (F4), print the generated questions (F5) and delete the questions (F6). Respondents were given only two options—true or false—for each feature. We were using ANOVA to check the difference between ISO attributes and the difference between IAQG features. Table 3 revealed that the data had both differences between means of 6 IAQG features and the differences between means of 22 ISO attributes. It means the quality of IAQG features are vary and based on ISO, there are several aspects need to be fixed. To find them, we calculated the percentage of "true" responses over the total number of responses for each feature.

Table 3
ANOVA analysis for ISO attributes and IAQG features

Factors		Sum of Square	df	Mean Square	F	F-critical	Result
ISO Attributes	Between Groups	0,9996	5	0,1999	15,445	2,2862	Reject H_0
	Within Groups	1,631	126	0,0129			
	Total	2,631	131				
IAQG Features	Between Groups	0,644	21	0,031	1,697	1,653	Reject H_0
	Within Groups	1,987	110	0,018			
	Total	2,631	131				

The summary of these results is presented in Table 4.

Table 4
Assessment of the features of IAQG

ISO Attributes	F1	F2	F3	F4	F5	F6	Average
Description Completeness	1	1	0.95	1	0.65	0.95	0.925
Demonstration Coverage	0.95	1	0.95	0.85	0.6	0.9	0.875
Entry Point Self-Descriptiveness	1	1	0.95	1	0.7	1	0.942
User Guidance Completeness	0.95	1	0.9	0.95	0.6	0.95	0.892
Entry Field Default	1	1	0.95	1	0.65	1	0.933
Error Message Understandability	1	1	0.9	0.95	0.65	0.95	0.908
Self-Explanatory User Interface	1	1	0.9	0.95	0.7	1	0.925
Operational Consistency	1	1	1	1	0.7	1	0.950
Message Clarity	1	1	0.95	1	0.65	1	0.933
Functional Customability	1	1	0.95	0.95	0.7	1	0.933
User Interface Customability	0.95	0.95	0.95	0.95	0.6	0.9	0.883
Monitoring Capability	0.95	1	0.9	0.95	0.5	0.9	0.867
Undo Capability	0.9	0.95	0.95	0.9	0.65	0.9	0.875
Understandable Categorization of Information	1	1	0.95	1	0.55	0.95	0.908
Appearance Consistency	1	1	1	0.95	0.7	1	0.942
Input Device Support	0.95	0.95	0.95	0.95	0.65	0.95	0.900
Avoidance of User Operation Error	0.95	0.95	0.95	0.65	0.95	1	0.908
User Entry Error Correction	1	0.95	1	0.7	1	0.85	0.917
User Error Recoverability	0.85	0.85	0.85	0.55	0.85	0.95	0.817
Appearance Aesthetic of User Interface	0.95	0.95	0.95	0.65	0.95	1	0.908
Accessibility for User with Disability	1	0.9	0.9	0.7	0.9	0.65	0.842
Supported Language Adequacy	0.65	0.65	0.65	0.4	0.7	0.6	0.608
Average	0.955	0.957	0.926	0.857	0.712	0.926	

Overall, the IAQG's features present minimal drawbacks, as the average scores for most attributes are recorded near 1. However, the printing feature received the lowest scores, as teachers expressed dissatisfaction with its functionality. Specifically, teachers are expected to export the generated questions into document formats, but the IAQG only displays a grouped list of questions, which sometimes includes other teachers' work. Additionally, teachers noted that the print function needs more precise instructions, more understandable error messages and better information on how to use it effectively.

Teachers also suggested that the IAQG should support more languages, as specific Indonesian terms, particularly in ICT topics, are ambiguous. Those specializing in literature hope that the IAQG can be further developed to generate better questions related to linguistic subjects.

3.3 TAM Analysis

To ensure the reliability of our instrument, we calculated Cronbach's alpha for each attribute. For every attribute, we first determined the total score of each response and then calculated the variance for the total score and each item. These variances were then used as inputs for the Cronbach's alpha function. The results of these calculations are presented in Table 5.

Table 5
Overview of Cronbach's Alpha results

TAM Attributes	Cronbach's Alpha	Number of Items	Reliability
Perceived Usefulness (PU)	0.790	9	Acceptable
Perceived Ease of Use (PEU)	0.767	6	Acceptable
Perceived Behaviour Control (PBC)	0.697	4	Questionable
Subjective Norm (SN)	0.910	7	Excellent
Behavioural Intention to Use (BI)	0.840	4	Good
Attitude Towards Use (ATU)	0.882	6	Good
Personal Innovativeness (PIIT)	0.809	4	Good

As shown in Table 5, only Perceived Behaviour Control (PBC) does not meet the acceptable threshold. PBC includes questions related to user behaviour, such as confidentiality, appropriate knowledge and resources for using IAQG and the freedom to operate the IAQG. These aspects are open to multiple interpretations, making them prone to misunderstandings. Consequently, we decided to exclude PBC and its correlation to other attributes from this analysis.

To test the hypotheses, we calculated the correlations between all attributes. Table 6 presents each attribute's mean, standard deviation and Pearson correlations. The standard deviations are less than the means, indicating that the data points are relatively close to the mean, with no extreme outliers. Respondents generally agreed with most attributes, except for Subjective Norm (SN) and BI, which had lower mean scores of 2.421 and 2.950, respectively. Respondents agreed to PBC, but we unconsidered it due to its unreliability.

Table 6
Pearson correlation and Descriptive statistics

	Mean	Std	PU	PEU	PBC	SN	BI	ATU	PIIT
PU	3.150	0.713	1.000	0.589	0.350	0.466	0.185	0.247	0.226
PEU	3.325	0.688	0.589	1.000	0.657	0.516	0.571	0.263	0.422
PBC	3.050	0.634	0.350	0.657	1.000	0.370	0.641	0.583	0.356
SN	2.421	0.898	0.466	0.516	0.370	1.000	0.523	0.533	0.046
BI	2.950	0.634	0.185	0.571	0.641	0.523	1.000	0.685	0.231
ATU	3.058	0.690	0.247	0.263	0.583	0.533	0.685	1.000	0.030
PIIT	3.350	0.597	0.226	0.422	0.356	0.046	0.231	0.030	1.000

Note: grey: considered to be ignored, yellow: moderate correlation, green: high correlation, blue: very high correlation

However, to confirm the hypotheses, ANOVA is needed to determine the statistical significance of these correlations. The results of ANOVA analysis can be seen in Table 7.

Table 7
ANOVA analysis for TAM attributes significance

Factors		Sum of Square	df	Mean Square	F	F-critical	Result
H1: Perceived Usefulness – Attitude Toward Use	Between Groups	55,632	9	6,181	0,487	0,853	insignificant
	Within Groups	126,918	10	12,691			
	Total	182,55	19				
H2: Perceived Ease of Use – Attitude Toward Use	Between Groups	85,371	6	14,229	1,903	0,156	insignificant
	Within Groups	97,179	13	7,475			
	Total	182,55	19				
H3: Perceived Behaviour Control – Attitude Toward Use	Between Groups	99,424	4	24,856	4,485	0,014	significant
	Within Groups	83,126	15	5,542			
	Total	182,55	19				
H4: Subjective Norm – Attitude Toward Use	Between Groups	126,848	7	18,121	3,904	0,019	significant
	Within Groups	55,702	12	4,642			
	Total	182,55	19				
H5: Personal Innovativeness – Attitude Toward Use	Between Groups	17,099	4	4,274	0,388	0,814	insignificant
	Within Groups	165,451	15	11,030			
	Total	182,55	19				
H6: Attitude Toward Use – Behavioural Intention to Use	Between Groups	56,018	6	9,336	4,159	0,015	significant
	Within Groups	29,182	13	2,248			
	Total	85,2	19				

Based on ANOVA analysis, out of the six hypotheses tested, H1, H2 and H5 were rejected despite having relatively high explanatory power, primarily PU and PEU. This indicates that PU, PEU and PIIT do not influence respondents to undergo positive experiences, which can influence them to promote the usage of IAQG to other people, especially teachers. On the other hand, PBC and SN strongly correlate to ATU with 54,5% and 69,5% explanatory power, respectively. In other words, ease of use encourages users to do positive behaviour like recommending to their colleagues to use IAQG. Ignoring the hypotheses, SN and BI do not correlate with teachers' intention to use IAQG (Table 6). Not all respondents realized that SN, ATU and BI correlate with IAQG's ease of use. Teachers assume they can operate IAQG with no error, so they think they need no advanced training to operate IAQG. This explains what makes BI has a low mean (Table 6).

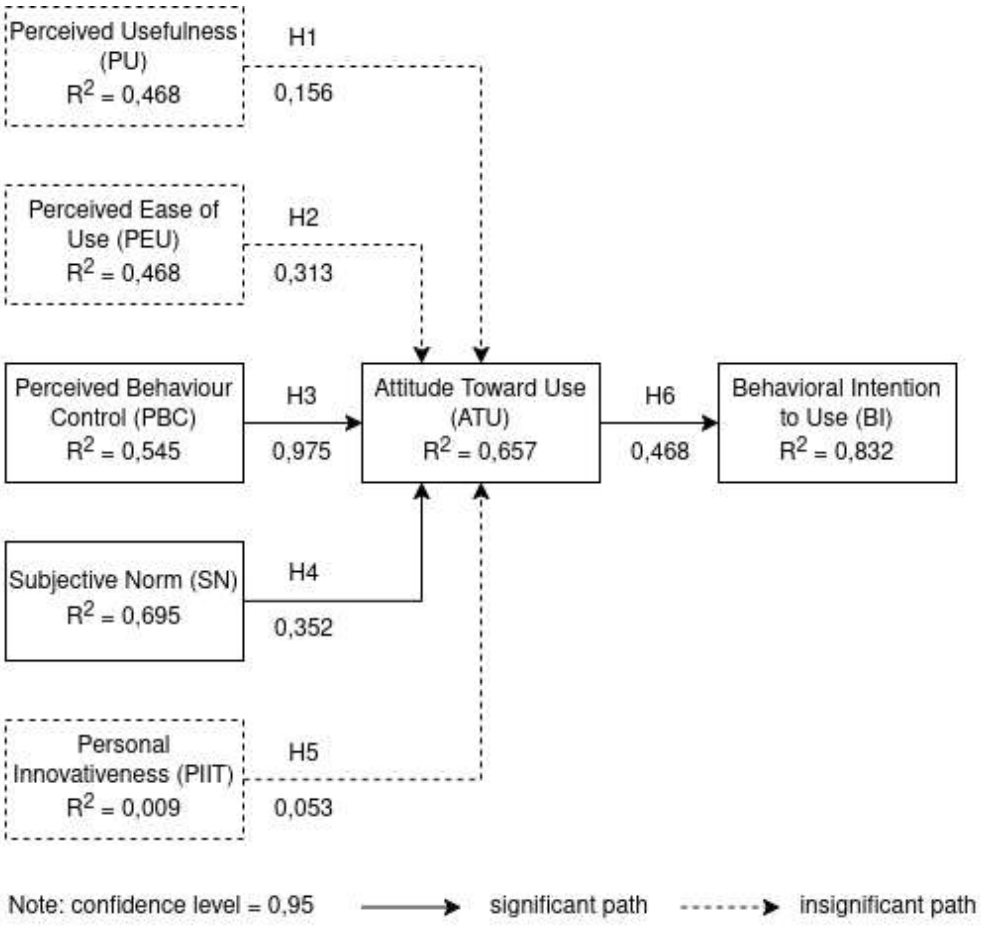

Fig. 4. Regression analysis

4. Discussion

The findings from this study underline the importance of aligning technological tools like the IAQG with educators' specific needs in learning outcome assessment and their expectations. IAQG's main features, such as the question generator and error messaging, are rated highly in usability, indicating that the tool can adequately support the teaching process if thoughtfully designed. However, the lower scores in accessibility and language support reveal broader challenges that must be addressed to ensure the inclusivity and adaptability of IAQG across diverse educational environments.

These results are particularly relevant in the broader context of the increasing digitization of educational tools. As Indonesian schools continue integrating more technology into their curricula, tools like IAQG could play a pivotal role in easing teachers' workloads and ensuring consistent assessment quality. However, usability challenges, such as enhanced language support and improved print functionality, must be addressed to realize these benefits fully. Addressing these areas is critical to ensuring that IAQG meets the diverse needs of Indonesian educators and can be seamlessly integrated into various educational settings.

The rejection of hypotheses H1, H2 and H5 suggests that Perceived Usefulness (PU), Perceived Ease of Use (PEU) and Personal Innovativeness (PIIT) were not significant factors in shaping teachers' attitudes toward using IAQG. Instead, Perceived Behaviour Control (PBC) and Subjective Norm (SN) influenced teachers' attitudes and, consequently, their behavioural intention to use the tool. This indicates that teachers are more likely to adopt the IAQG if it is easy to use and if their colleagues encourage its use, yet they remain doubtful that their usage influences others to adopt the tool.

Respondents highlighted several areas where IAQG could be improved, particularly its print functionality and language support. These issues could enhance the tool's effectiveness in real-world applications. For example, dissatisfaction with the print function—lacking clear instructions and error messages—indicates a need for better documentation and a more intuitive design. Additionally, the request for more robust language support reflects the linguistic diversity of Indonesian classrooms, where terms specific to ICT and other subjects may need to be better accommodated. Improvements in these areas could increase user satisfaction and positively affect other usability attributes, such as user interface aesthetics and operational consistency, thus driving higher adoption rates.

Future research could explore IAQG's application in different educational contexts, such as primary education or adult learning environments, to assess its adaptability and effectiveness across various settings. Longitudinal studies could provide valuable insights into how the sustained use of IAQG impacts teacher attitudes and student outcomes, offering a more comprehensive understanding of its long-term value. By focusing on enhancements such as improving documentation, accessibility features and language support, IAQG could become a more effective and widely adopted tool in educational settings.

The IAQG shows excellent potential as a supportive tool for educators. However, ongoing refinement and development are essential to fully meet teachers' needs and ensure its broader adoption and effectiveness in the classroom. Addressing the identified usability concerns and expanding the tool's capabilities will be vital in achieving this goal.

5. Conclusions

This study evaluated the usability of the IAQG, assessed teachers' acceptance and intention to use the tool and identified critical areas for improvement. The findings highlight IAQG's strengths, particularly its clear and understandable descriptions for generating and viewing questions, operational consistency and effective error messaging, indicating a user-friendly design. However, challenges such as improving accessibility for users with disabilities and expanding language support—especially for printing questions—must be addressed to increase IAQG's usability and inclusivity across diverse educational environments.

TAM analysis shows IAQG is generally well-received, with teachers expressing positive intentions to continue using it. However, Perceived Usefulness, Perceived Ease of Use and Personal Innovativeness did not significantly affect attitudes toward IAQG. Instead, Perceived Behaviour Control and Subjective Norm influenced teachers' attitudes and intentions to use the tool, suggesting that ease of use and colleague encouragement are key factors in adoption.

Respondents suggested several areas for improvement, notably the print functionality and language support. These enhancements, including more explicit documentation and intuitive design, would significantly increase IAQG's effectiveness in practical applications. Additionally, expanding language support to cover specific need and subject-specific terms would better serve the diverse needs of Indonesian classrooms and improve overall user satisfaction.

Future research could explore IAQG's application in primary education or adult learning environments to assess its adaptability across different educational contexts. Longitudinal studies could offer insights into the long-term impact of IAQG on teacher attitudes and student outcomes. By improving documentation, accessibility and language support, IAQG can become a more effective and widely adopted tool.

While IAQG shows excellent potential as a supportive tool for educators, ongoing refinement is crucial to fully meet teachers' needs and ensure its broader adoption and effectiveness in the classroom.

Acknowledgement
This research was funded by a grant from the Universitas Negeri Yogyakarta.

References
[1] Al-Maatouk, Qusay, Mohd Shahizan Othman, Ahmed Aldraiweesh, Uthman Alturki, Waleed Mugahed Al-Rahmi and Abdulmajeed A. Aljeraiwi. "Task-technology fit and technology acceptance model application to structure and evaluate the adoption of social media in academia." *Ieee Access* 8 (2020): 78427-78440. https://doi.org/10.1109/ACCESS.2020.2990420

[2] Asaloei, Sandra Ingried, Agustinus Kia Wolomasi and Basilius Redan Werang. "Work-Related stress and performance among primary school teachers." *International Journal of Evaluation and Research in Education* 9, no. 2 (2020): 352-358. https://doi.org/10.11591/ijere.v9i2.20335

[3] Awalurahman, Halim Wildan and Indra Budi. "Paragraph vs Sentence in Automatic Question Generation Fine-Tuning using Text-to-Text Transfer Transformer for Bahasa Indonesia." In *2024 10th International Conference on Education and Technology (ICET)*, pp. 155-161. IEEE, 2024. https://doi.org/10.1109/ICET64717.2024.10778465

[4] Barz, Nathalie, Manuela Benick, Laura Dörrenbächer-Ulrich and Franziska Perels. "Students' acceptance of e-learning: Extending the technology acceptance model with self-regulated learning and affinity for technology." *Discover Education* 3, no. 1 (2024): 114. https://doi.org/10.1007/s44217-024-00195-7

[5] Criollo-C, Santiago andrea Guerrero-Arias, Diego Buenaño-Fernández and Sergio Luján-Mora. "Usability and workload evaluation of a cybersecurity educational game application: a case study." *IEEE Access* 12 (2024): 12771-12784. https://doi.org/10.1109/ACCESS.2024.3352589

[6] Davis, Fred D. "Technology acceptance model: TAM." *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption* 205, no. 219 (1989): 5.

[7] Doxolodeo, Kerenza and Adila Alfa Krisnadhi. "AC-IQuAD: Automatically Constructed Indonesian Question Answering Dataset by Leveraging Wikidata." *Language Resources and Evaluation* 59, no. 1 (2025): 135-160. https://doi.org/10.1007/s10579-023-09702-y

[8] Esawe, Ahmed Taher, Karim Taher Esawe and Narges Taher Esawe. "Evaluating schoolteachers' acceptance of m-learning during Covid-19 using the Technology Acceptance Model: the moderating role of gender and age." *SN Social Sciences* 4, no. 2 (2024): 43. https://doi.org/10.1007/s43545-024-00830-y

[9] Faudzi, Masyura Ahmad, Zaihisma Che Cob, Masitah Ghazali, Ridha Omar and Sharul Azim Sharudin. "User interface design in mobile learning applications: Developing and evaluating a questionnaire for measuring learners' extraneous cognitive load." *Heliyon* 10, no. 18 (2024). https://doi.org/10.1016/j.heliyon.2024.e37494

[10] Fuadi, Mukhlish and Adhi Dharma Wibawa. "Automatic question generation from Indonesian texts using text-to-text transformers." In *2022 International conference on electrical and information technology (IEIT)*, pp. 84-89. IEEE, 2022. https://doi.org/10.1109/IEIT56384.2022.9967858

[11] Goodhue, Dale L. and Ronald L. Thompson. "Task-technology fit and individual performance." *MIS quarterly* (1995): 213-236. https://doi.org/10.2307/249689

[12] Hartono, Ferry. "Optimizing learning management system using UTAUT model: Environment, compatibility, material content and interactivity for enhanced usability." *Multidisciplinary Science Journal* 7, no. 3 (2025): 2025111-2025111. https://doi.org/10.31893/multiscience.2025111

[13] Henry, Matthew Martianus, Gregorius Natanael Elwirehardja and Bens Pardamean. "Automatic question generation for bahasa indonesia examination using copynet." *Procedia Computer Science* 245 (2024): 953-962. https://doi.org/10.1016/j.procs.2024.10.323

[14] Hu, Dongpin, Bei Yuan, Jiutong Luo and Minhong Wang. "A Review of Empirical Research on ICT Applications in Teacher Professional Development and Teaching Practice." *Knowledge Management & E-Learning* 13, no. 1 (2021): 1-20. https://doi.org/10.34105/j.kmel.2021.13.001

[15] Indrihapsari, Yuniar, Handaru Jati, N. Nurkhamid, Ratna Wardani, Pradana Setialana, Muhamad Izzudin Mahali, Danang Wijaya *et al.,* "A Comparison of OpenNMT Sequence Model for Indonesian Automatic Question Generation." *Elinvo (Electronics, Informatics and Vocational Education)* 8, no. 1 (2023): 55-63. https://doi.org/10.21831/elinvo.v8i1.56491

[16] Mertins, Birte and Jennifer Austermann. "Technology acceptance model revised: an investigation on the managerial attitudes towards using social media in innovation processes." (2014).

[17] Muis, Ferdiant Joshua and Ayu Purwarianti. "Sequence-to-sequence learning for indonesian automatic question generator." In *2020 7th international conference on advance informatics: Concepts, theory and applications (ICAICTA)*, pp. 1-6. IEEE, 2020. https://doi.org/10.1109/ICAICTA49861.2020.9429032

[18] Navarro-Cota, Christian X., Ana I. Molina, Miguel A. Redondo and Carmen Lacave. "A comprehensive usability measurement tool for m-learning applications." *IEEE Transactions on Education* 67, no. 2 (2024): 209-223. https://doi.org/10.1109/TE.2023.3347191

[19] Niromand, Elham, Meysam Siyah Mansoory, Ghobad Ramezani and Mohammad Rasool Khazaei. "Design, implementation and evaluation of e-learning program for common diseases to smartphone-based medical students: at a developing university." *BMC Medical Education* 24, no. 1 (2024): 52. https://doi.org/10.1186/s12909-023-05023-4

[20] Putri, Rifki Afina and Alice Oh. "Idk-mrc: Unanswerable questions for indonesian machine reading comprehension." *arXiv preprint arXiv:2210.13778* (2022). https://doi.org/10.18653/v1/2022.emnlp-main.465

[21] Richardson, Benny and Arya Wicaksana. "Comparison of indobert-lite and roberta in text mining for Indonesian language question answering application." *Int. J. Innov. Comput. Inf. Control* 18, no. 6 (2022): 1719-1734.

[22] Ruiz Fernández, María Luisa, Luis-Javier Márquez-Álvarez, Estíbaliz Jiménez-Arberas and Isabel Fernández Méndez. "Implementation and evaluation of a mobile application to support occupational therapy education." *Irish Journal of Occupational Therapy* 52, no. 2 (2024): 91-97. https://doi.org/10.1108/IJOT-03-2024-0005

[23] Si, Miao, Melor Md Yunus, Karmila Rafiqah M. Rafiq and Asep Sunandar. "A Systematic Review (2019–2023) on the Impact of Advanced Technologies in Higher Education Learning." *International Journal of Advanced Research in Future Ready Learning and Education* 37, no. 1 (2024): 26-40. https://doi.org/10.37934/frle.37.1.2640

[24] Suhartono, Derwin, Muhammad Rizki Nur Majiid and Renaldy Fredyan. "Towards automatic question generation using pre-trained model in academic field for Bahasa Indonesia." *Education and Information Technologies* 29, no. 16 (2024): 21295-21330. https://doi.org/10.1007/s10639-024-12717-9

[25] Steuer, Tim, Anna Filighera and Thomas Tregel. "Investigating educational and noneducational answer selection for educational question generation." *IEEE Access* 10 (2022): 63522-63531. https://doi.org/10.1109/ACCESS.2022.3180838

[26] Vincentio, Karissa and Derwin Suhartono. "Automatic Question Generation using RNN-based and Pre-trained Transformer-based Models in Low Resource Indonesian Language." *Informatica (03505596)* 46, no. 7 (2022). https://doi.org/10.31449/inf.v46i7.4236