ASIA PACIFIC JOURNAL
OF EDUCATORS AND EDUCATION

Volume 40, Issue 3, 2025
https://doi.org/10.21315/apjee2025.40.3.26

Research Article:

Development and Validation of a Construct 2-Based Gamified Assessment for Arabic Language Learning

Mohammad Ahsanuddin[1] , Zawawi Ismail[2], Muhammad Ahsan Thoriq[1], Farika Riskiyah[1], Rico Fenda Pradana[1], Arya Wahyu Pratama[1], Najma Auliya' Maulidah Ahsan[3]

[1]Arabic Languange Education, Universitas Negeri Malang, Indonesia
[2]Department of Language and Literacy Education, Faculty of Education, Universiti Malaya, Malaysia
[3]Communication Sciences Department, Universitas Negeri Malang, Indonesia

*Corresponding author: mohammad.ahsanuddin.fs@um.ac.id

ABSTRACT

Students' perceptions of gamified assessment significantly influence their motivation and engagement in learning Arabic. Measuring these perceptions offers insight into how students evaluate the effectiveness of assessment strategies in Arabic language education. However, the development of suitable gamified assessment instruments remains underexplored. This study aims to develop and validate a gamified assessment tool using Construct 2 for Arabic language instruction at SMP Darul Faqih, Indonesia. The assessment dimensions and items were derived from a conceptual analysis and open-ended survey responses. Content validity was established through expert judgement involving five experts and three Arabic language teachers. The questionnaire was piloted with 60 students, divided evenly for exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), comprising both control and experimental groups. The analysis identified four dimensions—interactivity, user interface, feedback, and learning motivation—which collectively explained 66.8% of the variance. The finalised instrument consists of 12 items and demonstrated satisfactory reliability, validity, and internal consistency. Among the dimensions, learning motivation exhibited the strongest path coefficient, while the user interface dimension had the weakest. The experimental group showed significant improvements in interactivity, feedback, and learning motivation, whereas no significant difference was found in the user interface dimension. These findings suggest that the developed questionnaire is a valid and reliable tool for assessing students' perceptions of gamified assessment in Arabic language learning. Future research should explore the application of this instrument in examining the relationship between students' perceptions and their learning outcomes.

Keywords: Gamified assessment, Arabic language learning, Construct 2, student perception

Accepted: 23 October 2025; **Published:** 5 January 2026

To cite this article: Ahsanuddin, M., Ismail, Z., Thoriq, M. A., Riskiyah, F., Pradana, R. F., Pratama, A. W., Ahsan, N. A. M. (2025). Development and validation of a construct 2-based gamified assessment for Arabic language learning. *Asia Pacific Journal of Educators and Education, 40*(3), 671–706. https://doi.org/10.21315/apjee2025.40.3.26

INTRODUCTION

Arabic holds a central role in Islamic education. As the language of the Qur'an and Hadith, Arabic serves not only as a means of communication but also as a vital tool for accessing Islamic teachings in their original form. In Indonesia, Arabic is widely taught from primary to secondary levels, particularly within Islamic educational institutions such as madrasahs, *pesantren*, and religious-based schools. However, despite extensive policy support and curriculum development, Arabic instruction at the junior high school (SMP) level continues to face persistent challenges that are inadequately addressed by existing pedagogical approaches. Critical examination of current practices reveals that teaching methods remain predominantly traditional, relying heavily on lectures, rote memorisation, and grammar-focused instruction. These approaches have consistently failed to engage students or sustain their motivation, particularly among 21st-century learners who are immersed in digital and multimedia environments (Fannakhosrow et al., 2022; Hanafy et al., 2021).

Previous attempts to integrate technology into Arabic language instruction have yielded mixed results, largely due to theoretical and methodological limitations. Many digital interventions have focused primarily on content delivery rather than pedagogical transformation, failing to address fundamental issues of learner engagement and motivation (Ritonga et al., 2023). Furthermore, the absence of theoretically grounded frameworks for designing and evaluating technology-enhanced Arabic instruction has hindered systematic progress in this domain. This theoretical gap is particularly evident in the lack of validated instruments for measuring student perceptions of innovative assessment approaches—a critical oversight given that learner acceptance and engagement are key predictors of educational technology success.

Another persistent issue in Arabic language instruction at the junior high schools is students' low level of active engagement during classroom activities. The complex morphological and syntactic structure of Arabic, combined with limited exposure to authentic language use outside the classroom, creates significant cognitive and motivational barriers for learners. These challenges highlight the urgent need for innovative teaching media and strategies that are not only pedagogically sound but also psychologically engaging. One approach that has shown promise in enhancing student participation and fostering a more enjoyable learning environment is gamification—the integration of game-based elements into non-game learning contexts (Khoo et al., 2023).

Gamification, when grounded in Self-Determination Theory (SDT), offers a theoretically robust approach to addressing the motivational deficits observed in traditional Arabic instruction. SDT posits that intrinsic motivation is fostered through the satisfaction of three basic psychological needs: autonomy (sense of volition and

choice), competence (sense of effectiveness and mastery), and relatedness (sense of connection with others) (Ryan & Deci, 2020). Game elements such as adaptive challenges, immediate feedback, and progress visualisation can potentially satisfy these needs, thereby enhancing learner engagement. However, critical questions remain regarding which specific gamification elements most effectively support these psychological needs in the context of Arabic language learning, and how learners themselves perceive and respond to gamified assessment approaches. These questions constitute the theoretical gaps that the present study seeks to address.

In the field of education, gamification has been widely studied and applied across various disciplines. Game elements such as points, levels, badges, leaderboards, and instant feedback have been shown to foster both competitive and collaborative learning environments, thereby increasing student engagement and motivation (Zeybek & Saygı, 2024). In the context of foreign language learning, gamification has proven effective in helping students retain vocabulary, understand sentence structures, and develop listening and speaking skills in a more interactive manner (Oliveira et al., 2023). Vrcelj et al., (2023) demonstrated that mobile learning-based gamification systems significantly improve student motivation and learning achievement in language education. These findings suggest that gamified learning encourages a more student-centered, adaptive, and responsive learning experience that aligns with diverse learner needs.

In terms of media development for gamified instruction, Construct 2 offers a practical and efficient solution for creating interactive HTML5-based assessments. Designed to be accessible to users without advanced programming skills, Construct 2 enables educators to develop educational games through an intuitive visual interface (L. Li et al., 2024). Teachers can design game-based learning scenarios that incorporate evaluative components such as quizzes, simulations, and level-based challenges. A key advantage of this platform is its ability to produce applications compatible with various devices, including PCs, tablets, and smartphones (Mazeas et al., 2022).

Despite the growing potential of gamification and platforms like Construct 2 in education, their application in Arabic language learning—particularly in the form of gamified assessment—remains limited. This gap is largely due to the lack of evaluation tools specifically designed to meet the needs of Arabic instruction, considering its linguistic content, pedagogical approach, and the characteristics of its learners (Rohan et al., 2021). Yet, assessment plays a critical role in the learning process, serving to measure learning outcomes, guide instruction, and provide constructive feedback. When implemented through gamification, assessments can function not only as evaluative tools but also as engaging and meaningful learning instruments (Oke et al., 2024).

The success of gamified assessment in educational contexts is determined not only by its technical design or game features, but also by how well it is received and perceived as effective by students. Learners' perceptions of the media used in instruction serve

as a critical indicator of the quality and practical value of educational innovations (Silic & Lowry, 2020). These perceptions encompass various aspects, including ease of use, visual appeal, clarity of instructions, and the extent to which the tool supports content comprehension. Consequently, a valid and reliable instrument is essential for measuring students' perceptions of gamified assessments in Arabic language learning. Prior studies in education have consistently shown that positive student perceptions of instructional media correlate with improved learning outcomes and greater learner satisfaction (Schöbel et al., 2020).

This study developed a gamified assessment tool using Construct 2, tailored to the characteristics of Arabic language instruction at the junior secondary school level. The development process was guided by principles of Arabic pedagogy and incorporated relevant game elements to ensure alignment with learning objectives. In addition, a validation process was undertaken for the perception questionnaire to ensure that the tool was not only technically effective but also psychologically acceptable to its end users—students (Reyes et al., 2021). The resulting assessment tool is expected to offer Arabic language teachers an innovative and meaningful alternative for evaluating student learning, while also contributing to the broader development of technology-based instruments in language education.

More broadly, the development of this Construct 2-based gamified assessment addresses the growing need for evaluation tools that are responsive to students' digital learning styles and the demands of 21st-century education. Within the framework of the Merdeka Curriculum, which promotes differentiated and character-based learning, such assessments can enrich Arabic language pedagogy by offering a more enjoyable, challenging, and motivational approach. By integrating gamification into the assessment process, students are evaluated not only in terms of cognitive performance but are also encouraged to become more active, reflective, and autonomous in their learning (Limantara et al., 2023). Accordingly, this study represents an important step toward transforming Arabic language assessment into a more innovative, contextualised, and student-centered experience.

This research has two primary objectives. First, to develop a questionnaire to measure students' perceptions of a gamified assessment created with Construct 2 for Arabic language learning at the junior secondary level. Second, to investigate differences in perception between a control group (taught using conventional methods) and an experimental group (taught using the gamified assessment). The instrument developed through this study is expected to help educators identify the most effective gamification elements, thereby enhancing student engagement, motivation, and learning outcomes in Arabic language instruction.

LITERATURE REVIEW

Gamification in Language Learning

Gamification has emerged as a widely studied pedagogical approach for enhancing language learning outcomes, with substantial research demonstrating its positive effects on motivation, engagement, and knowledge retention (Gao, 2021; Guerandel et al., 2021; Khasawneh et al., 2024; Yoshigami et al., 2019). Game elements such as points, levels, leaderboards, and reward systems have been shown to create more engaging learning environments that promote active participation (Min et al., 2025; Reyes et al., 2021; Shakoor et al., 2025; Wainwright et al., 2020).

However, critical examination of the gamification literature reveals several important limitations and unresolved tensions. First, empirical evidence suggests that gamification effects are highly context-dependent and not universally positive (Bovermann & Bastiaens, 2020). Some studies have found that extrinsic rewards can undermine intrinsic motivation under certain conditions, particularly when game elements are perceived as controlling rather than informational (Ryan & Deci, 2020). Second, implementation challenges related to technological infrastructure, teacher training, and curriculum integration often limit the practical effectiveness of gamified interventions in real educational settings (Limantara et al., 2023). Third, cultural differences in gaming preferences and learning styles may moderate the effectiveness of specific gamification strategies, yet cross-cultural research in this area remains limited.

Within the specific context of Arabic language learning, gamification research has predominantly focused on vocabulary acquisition and reading skills, with less attention to comprehensive language proficiency development. Furthermore, while gamification can reinforce the four key language skills—listening (*istima'*), reading (*qira'ah*), writing (*kitabah*), and speaking (*kalam*)—the mechanisms through which game elements support different skill domains remain undertheorised. This study addresses these gaps by examining student perceptions of gamified assessment across multiple dimensions of the learning experience.

Digital Assessment and Its Role in Learning

Digital assessment technologies have evolved considerably, ranging from traditional computer-based tests to fully immersive game-based evaluation systems. Research indicates that game-based assessments can provide immediate feedback, reduce test anxiety, and measure authentic competencies more effectively than traditional paper-based formats (AlSaif & Alsenany, 2015). However, critical questions remain regarding the validity, reliability, and fairness of these innovative assessment approaches.

For Arabic language learning, digital assessments offer potential advantages including adaptive difficulty adjustment, multimedia integration, and real-time performance analytics (Ng & Raghbir, 2021; Sinar et al., 2024). Yet these affordances also introduce new challenges. Technical complexity may create access barriers for students with limited digital literacy. Cultural assumptions embedded in game design may disadvantage learners from non-Western backgrounds. Assessment authenticity may be compromised if game elements distract from or distort the construct being measured. These tensions highlight the need for rigorous validation of gamified assessment instruments that examines not only their psychometric properties but also their perceived appropriateness and effectiveness from students' perspectives.

Construct 2 as a Development Platform

Construct 2 represents a category of visual programming environments that enable educational content creation without requiring extensive coding expertise (Ahsanuddin et al., 2025; Hestiningsih, 2019). From a theoretical perspective, such platforms democratise educational technology development by reducing technical barriers, potentially allowing teachers to become designers of their own digital learning resources. The platform's event-based programming logic and drag-and-drop interface align with constructionist learning principles by enabling iterative design and rapid prototyping.

However, critical evaluation reveals several limitations that warrant consideration. First, while Construct 2 reduces programming complexity, effective educational game design still requires substantial pedagogical expertise and multimedia skills that many teachers lack. Second, the platform's focus on 2D game development may limit the sophistication of learning interactions compared to more advanced development environments. Third, games created with Construct 2 depend on browser compatibility and device capabilities, potentially creating technical barriers for students with limited technology access.

Despite these constraints, Construct 2 offers practical advantages for educational research, particularly its capacity for rapid development and cross-platform deployment (Mazeas et al., 2022). Critically, however, empirical research validating Construct 2-based educational applications remain limited, especially for language learning contexts. This study addresses this gap by developing and validating a gamified assessment instrument using Construct 2, thereby contributing both a practical tool and empirical evidence regarding the platform's pedagogical potential for Arabic language education.

Gamified Assessment in Arabic Language Learning

Emerging research suggests that game-based evaluation can reduce anxiety and enhance engagement in Arabic language learning contexts (Hesham et al., 2024). Interactive

quiz formats have demonstrated effectiveness in improving vocabulary retention and reading comprehension (Baso, 2022; Roumei & Feng, 2012). However, several critical gaps limit the field's theoretical and practical advancement.

First, most existing studies have examined gamification effects on learning outcomes without systematically investigating student perceptions of gamified assessment—a significant oversight given that learner acceptance strongly predicts technology adoption and sustained engagement (Rohan et al., 2021). Second, the theoretical mechanisms through which gamification influences Arabic learning remain underspecified, with limited integration of motivation theory or cognitive science frameworks. Third, methodological limitations including small sample sizes, short intervention periods, and lack of validated measurement instruments constrain the generalisability and interpretability of existing findings.

Most critically, no previous research has developed and validated a psychometrically sound instrument for measuring student perceptions of gamified assessment specifically designed for Arabic language learning contexts. This represents a fundamental gap, as the development of such an instrument is prerequisite for systematic evaluation of gamification effectiveness across different implementations, populations, and instructional contexts. The present study addresses this gap by developing and validating the Gamified Assessment in Arabic Learning Questionnaire (GA-ALQ), thereby providing both a practical evaluation tool and theoretical insights into the dimensional structure of student perceptions of gamified assessment.

Validation of Educational Media

Establishing validation routines for technology-enhanced assessments is indispensable to safeguard instrument integrity (Muzdalifah et al., 2021; Rohman, 2019). Drawing from Ma and Feng (2021), the multidimensional nature of validity can be disaggregated into content validity (degree of conformity with specified learning content), construct validity (degree of correspondence between measurement indicators and articulated competencies), and face validity (intuitive usability and intelligibility for intended examinees). Within the present study, engagement of bilingual Arabic linguists, systematic input from multimedia-design specialists, and iterative trial deployments with target learners collectively furnish converging evidence, thereby enhancing the psychometric robustness and pedagogical appropriateness of the instrument.

METHODOLOGY

To achieve the research objectives, we developed a questionnaire titled the Gamified Assessment in Arabic Learning Questionnaire (GA-ALQ). The development process

followed a systematic approach comprising four key stages, consistent with established procedures for educational instrument development (Magistretti et al., 2020).

Stage 1: Initial Item Development

Item development began with an in-depth literature review on gamification in education and Arabic language learning. Using keywords such as "gamification in education," "Arabic language learning," and "interactive assessment," we searched indexed educational journals. The first author conducted thematic coding of relevant literature to identify key dimensions underlying students' perceptions of gamified assessment. This process initially yielded three dimensions: interactivity (the assessment's ability to engage students actively), user interface (visual design and ease of navigation), and learning motivation (the impact of the assessment on students' learning interest). The coding was validated through discussions with two co-authors to ensure consistency and accuracy.

To further refine these dimensions, we conducted an open-ended survey with 10 junior high school students enrolled in Arabic classes. Students were asked: "What do you expect from a game-based assessment to support your Arabic language learning?" Respondents provided free-form responses about the assessment features they considered important.

Qualitative Data Analysis Framework

The open-ended responses were analysed. The coding process involved three stages (Schüttpelz-Brauns et al., 2020):

1. Initial Coding (First Cycle): All 10 student responses were transcribed verbatim, the first author conducted line-by-line open coding to identify initial concepts, a total of 47 preliminary codes were generated (e.g., "wants instant results," "needs visual appeal," "prefers game-like challenges").

2. Axial Coding (Second Cycle): Preliminary codes were grouped into higher-order categories based on conceptual similarity, five preliminary themes emerged: Interactivity, Visual Design, Immediate Feedback, Motivational Elements, and Ease of Use, two independent coders (co-authors) reviewed the coding, with inter-coder agreement of 89% (Cohen's kappa = 0.84).

3. Selective Coding (Final Integration): The five preliminary themes were refined into four final dimensions through discussion with three Arabic language teachers, visual Design and Ease of Use were merged into "User Interface", immediate Feedback emerged as a distinct fourth dimension (mentioned by 8/10 students, final dimensions: (i) Interactivity, (ii) User Interface, (iii) Feedback, (iv) Learning Motivation.

Table 1. Coding framework from student open-ended responses

Final dimension	Preliminary codes	Example student quote	Frequency (n = 10)
Interactivity	Active participation; Game challenges; Direct engagement; Can try multiple times	I want to directly try questions, not just read	9
User Interface	Visual appeal; Easy navigation; Clear layout; Attractive colours	The display should be attractive so I don't get bored	7
Feedback	Instant results; Know mistakes immediately; Get correction; See score right away	Immediately know if answer is right or wrong	8
Learning motivation	More interested; Want to keep learning; Feel challenged; More enthusiastic	Become more enthusiastic about learning Arabic	10

This systematic coding process ensured that the four dimensions were grounded in students' actual expectations rather than solely derived from literature, enhancing the ecological validity of the instrument. These four dimensions were subsequently validated by three Arabic language teachers to ensure their relevance to the instructional context.

Based on this process, we developed an initial pool of 20 questionnaire items: 6 items for interactivity, 5 for user interface, 5 for learning motivation, and 4 for feedback. Each item used a five-point Likert scale (1 = strongly disagree, 5 = strongly agree), following recommendations by Yamashita (2022) for measuring affective dimensions in education.

Stage 2: Content Validity Evaluation

Content validity was assessed to ensure that each item accurately reflected its intended dimension and aligned with the research objectives. A panel of five experts—three specialists in Arabic language education and two in gamification—along with three experienced Arabic teachers, evaluated the 20-item draft. Panel members were asked to rate each item using two response options: *agree* (with suggestions for revision, if needed) or *disagree* (item to be deleted). Agreement rates were calculated, and items with less than 80% agreement were excluded from the final instrument (Naye et al., 2022).

The evaluation led to the removal of 3 items from interactivity, 2 from user interface, and 3 from learning motivation due to redundancy or lack of specificity. For example, the item "Gamified assessment helps me learn faster" was removed for being too general

and overlapping with other items in the learning motivation dimension. Following these revisions, the questionnaire was reduced to 12 items: 3 items each for interactivity (INT1–3), user interface (UI1–3), feedback (FB1–3), and learning motivation (ML1–3).

Stage 3: Factorial Validity and Internal Reliability Testing

Preliminary analyses involved descriptive statistics (means and standard deviations) and Pearson product-moment correlations between individual items and total scores. Items with significant correlations ($p < 0.05$) and coefficients below 0.80 were retained to avoid multicollinearity (Lehmann & Vogt, 2024).

Exploratory Factor Analysis (EFA) was conducted using Principal Components Analysis (PCA) with varimax rotation and Kaiser normalization to identify the underlying factor structure. Items with factor loadings below 0.50 or cross-loadings (loading on two or more factors) were removed (Park, 2023). Eigenvalues greater than 1.00 were used to determine the number of retained factors. Subsequently, Confirmatory Factor Analysis (CFA) was performed to test the model fit derived from the EFA using maximum likelihood estimation. Fit indices included: $X^2/df \leq 3.00$ (Hair et al., 2020), SRMR < 0.06, RMSEA < 0.08 (Hooper et al., 2008), and CFI, GFI, NFI, and TLI ≥ 0.90 (Alavi et al., 2020a).

Internal reliability was assessed using Cronbach's alpha ($\alpha \geq 0.60$), Construct Reliability (CR ≥ 0.70), and Average Variance Extracted (AVE ≥ 0.50) (Hair, 2010). Concurrent validity was examined through correlations between each dimension and the total score, with correlation coefficients (r) below 0.80 indicating adequate dimensional independence (Lai, 2021).

The study sample consisted of 60 junior high school students from SMP Darul Faqih, Indonesia, all of whom were enrolled in Arabic language classes, yielding a 100% response rate. The sample was evenly split: 30 students (15 control, 15 experimental) for EFA, and 30 students (15 control, 15 experimental) for CFA. While this sample size meets minimum requirements for exploratory analysis, we acknowledge that it falls below optimal recommendations for CFA, which typically require 100 to 200 participants for stable parameter estimates (Marsh et al., 2020).

This limitation must be explicitly recognised. The relatively small sample size, particularly for CFA ($n = 30$), increases the risk of parameter estimation instability and limits the generalisability of our findings. Confidence intervals for model parameters are likely wider than they would be with larger samples, and the risk of Type II error (failing to detect true misspecifications) is elevated. Furthermore, model fit indices in

small samples can be unreliable, potentially producing either overly optimistic or overly pessimistic assessments of model adequacy.

Despite these limitations, several factors support the provisional validity of our approach. First, this study represents an initial instrument development effort, with the expectation that subsequent validation with larger samples will be necessary before widespread application. Second, our KMO value of 0.78 indicates adequate sampling adequacy for factor analysis, suggesting that the correlation matrix is appropriate for factoring despite sample size constraints. Third, the consistency between EFA and CFA results provides converging evidence for the four-factor structure. Nevertheless, readers should interpret findings cautiously, recognising that replication with larger, more diverse samples is essential for establishing the instrument's robustness and generalisability.

Post-hoc power analysis using Monte Carlo simulation (1,000 replications) indicated that our sample size provided adequate power (≥ .80) to detect large effects ($d \geq 0.80$) but insufficient power (< .60) to detect small to medium effects ($d < 0.50$). This suggests that null findings should be interpreted with particular caution, as they may reflect inadequate statistical power rather than true absence of effects.

Content validity evaluation involved systematic documentation procedures to ensure replicability. Each expert received a validation form containing: operational definitions of the four dimensions, the complete item pool with accompanying rating scales, structured evaluation criteria (clarity, relevance, representativeness), and space for qualitative feedback and revision suggestions.

Inter-rater agreement was calculated using Fleiss' kappa for multiple raters, yielding $\kappa = 0.76$ (95% CI [0.68, 0.84]), indicating substantial agreement. Items with agreement rates below 80% across expert panels were flagged for revision or removal. For example, the item "Gamified assessment helps me learn faster" received only 62.5% agreement (5 of 8 experts), with qualitative feedback noting that it was "too general" and "confounds speed with quality of learning." This item was consequently removed from the final instrument. All expert evaluation forms and inter-rater reliability calculations are available in supplementary materials to enable full replication of the validation process.

Stage 4: Analysis of Student Perceptions

To assess differences in students' perceptions of gamified assessment, total scores and subscale scores were compared between the control and experimental groups using one-way ANOVA. This analysis aimed to determine whether the gamified assessment had a statistically significant effect on student perceptions compared to conventional methods.

RESULTS

The mean scores of individual GA-ALQ items ranged from 3.02 to 4.25, with standard deviations between 0.72 and 1.12. Item–total correlations ranged from 0.38 to 0.74, all falling within the moderate range and below the multicollinearity threshold of 0.80. No item deviated more than 2.5 standard deviations from the mean. Therefore, all 12 items were retained for further analysis.

Exploratory Factor Analysis

The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was 0.78, indicating that the sample was suitable for factor analysis. Bartlett's test of sphericity was significant, $X2(66) = 892.4$, $p = 0.000$, confirming the presence of sufficient correlations among variables and multivariate normality. EFA identified four distinct dimensions—interactivity, user interface, feedback, and learning motivation—comprising all 12 items and explaining 66.8% of the total variance. All items exhibited factor loadings above 0.50, with no evidence of cross-loading. As a result, no items were excluded from the model.

The EFA results for the 12 GA-ALQ items revealed a clear factor structure distributed across four main dimensions: Interactivity, User Interface, Feedback, and Learning Motivation. Only factor loadings above 0.50 are reported. The Interactivity (INT) dimension included three items: INT1 (0.762), INT2 (0.789), and INT3 (0.715), with communalities of 0.712, 0.724, and 0.658, respectively, indicating strong internal consistency. The User Interface (UI) dimension comprised UI1 (0.698), UI2 (0.724), and UI3 (0.701), with communalities of 0.632, 0.678, and 0.645. The Feedback (FB) dimension included two items: FB1 (0.831) and FB2 (0.865), with communalities of 0.792 and 0.821. The Learning Motivation (ML) dimension contained ML1 (0.882), ML2 (0.856), and ML3 (0.819), with communalities of 0.845, 0.812, and 0.789, respectively, indicating a strong contribution to the overall factor structure. The eigenvalues for each factor were 5.892, 1.923, 1.584, and 1.112, respectively, accounting for a cumulative variance of 66.8%, which suggests that the four-factor model explains the majority of variance in the data. Overall, the EFA results indicate a robust and well-structured model, supporting the factorial validity of the instrument and the underlying dimensional structure of students' perceptions of gamified assessment.

Table 2. Exploratory factor analysis of 12 items

Dimensions	Component				Communality
	1	2	3	4	
Interactivity					
INT 1	0.762				0.712
INT 2	0.789				0.724
INT 3	0.715				0.658
User Interface					
UI 1		0.698			0.632
UI 2		0.724			0.678
UI 3		0.701			0.645
Feedback					
FB 1			0.831		0.792
FB 2			0.865		0.821
FB 3			0.794		0.763
Learning Motivation					
ML 1				0.882	0.845
ML 2				0.856	0.812
ML 3				0.819	0.789
Eigenvalue	5.892	1.923	1.584	1.112	
% of variance	36.825	12.019	9.900	7.056	
Cumulative %	36.825	48.844	58.744	66.800	

Notes: Only factor loadings greater than .50 are reported. INT = Interactivity, UI = User Interface, FB = Feedback, ML = Learning Motivation.

Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) using a correlated-factor model (Figure 1a) was conducted to evaluate the 12 questionnaire items (Appendix A) identified through EFA. The results indicated a strong model fit, with $X^2/df = 2.098$, RMSEA = 0.042, SRMR = 0.069, CFI = 0.952, GFI = 0.927, NFI = 0.914, and TLI = 0.938. A hierarchical model (Figure 1b) yielded similarly acceptable fit indices: $X^2/df = 2.132$, RMSEA = 0.045, SRMR = 0.070, CFI = 0.950, GFI = 0.924, NFI = 0.912, and TLI = 0.936. All standardised regression weights (λ) were greater than 0.50 and statistically significant ($p < 0.05$). Among the four dimensions, Learning Motivation had the strongest path coefficient (0.92, $p < 0.001$), while User Interface had the lowest (0.60, $p > 0.05$), suggesting that although all dimensions contributed meaningfully to the construct, their relative impact varied.

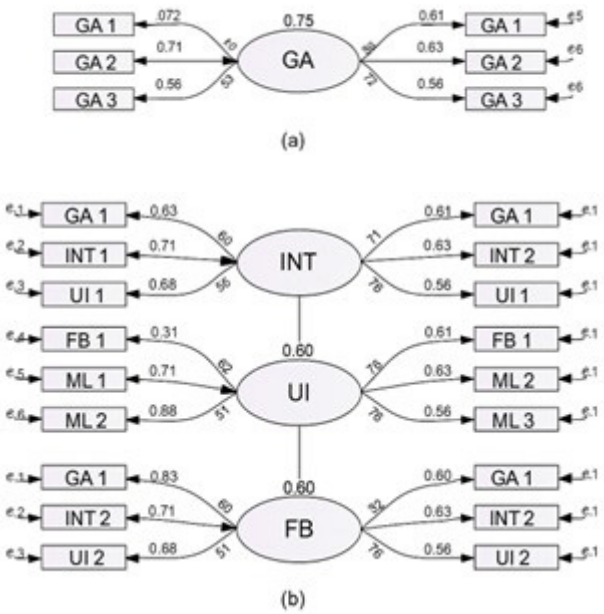

Figure 1. Confirmatory Factor Analysis based on 12 items and four dimensions: (a) Correlated-factor model; (b) Hierarchical model. GA = Gamified Assessment; INT = Interactivity; UI = User Interface; FB = Feedback; ML = Learning Motivation.

In the correlated-factor model (Figure 1a), the Gamified Assessment (GA) dimension demonstrated strong correlations with its associated items: GA1 (0.75), GA2 (0.71), and GA3 (0.56). Inter-item correlations were also high—GA1–GA2 (0.88), GA1–GA3 (0.72), and GA2–GA3 (0.56)—with consistently low measurement errors (e5 to e6), indicating strong internal consistency. The hierarchical model (Figure 1b) reflected a structured relationship in which the INT, UI, and FB dimensions contributed to a second-order latent factor. INT showed high loadings on INT1 (0.71) and INT2 (0.63), UI loaded strongly on UI1 (0.68), and FB on FB1 (0.31). The ML dimension was supported by ML1 (0.71) and ML2 (0.88). Correlations between dimensions ranged from 0.50 to 0.76. Item-level correlations in the hierarchical model were also substantial, such as GA1–INT2 (0.83) and INT2–UI2 (0.71), and measurement errors (e1 to e6) remained consistently low. These results indicate that both models demonstrate acceptable fit, though the hierarchical model provides a more organised structure in which dimensions are interrelated under a higher-order factor.

Based on the values of Cronbach's alpha (α), Composite Reliability (CR), Average Variance Extracted (AVE), and inter-dimensional correlations (ρ), the instrument demonstrates strong reliability and convergent validity. For the Interactivity dimension, α = 0.672, CR = 0.811, and AVE = 0.80, indicating good reliability and strong convergent validity, with an inter-dimensional correlation (ρ) of 0.81. The User

Interface dimension showed α = 0.645, CR = 0.782, and AVE = 0.76, with ρ = 0.78, suggesting adequate internal consistency, albeit slightly below the ideal threshold for α. The Feedback dimension demonstrated excellent psychometric properties, with α = 0.802, CR = 0.849, AVE = 0.89, and ρ = 0.88, indicating high reliability and convergent validity. The Learning Motivation dimension performed the best, with α = 0.890, CR = 0.862, and AVE = 0.90, and a correlation coefficient of ρ = 0.89. Overall, the total instrument yielded α = 0.873, CR = 0.873, and AVE = 0.90, with ρ = 0.90, confirming that the overall model possesses high internal consistency and convergent validity, thus supporting the robustness of the construct assessed in this study.

Table 3. CR, AVE, and α of 12 items and 4 dimensions of the gamified assessment

Dimension	Criteria			
	λ	CR	AVE	α
Interactivity				
INT 1	0.672			
INT 2	0.811	0.80	0.49	0.81
INT 3	0.729			
User Interface				
UI 1	0.645			
UI 2	0.782	0.76	0.46	0.78
UI 3	0.710			
Feedback				
FB 1	0.802			
FB 2	0.849	0.89	0.70	0.88
FB 3	0.873			
Learning motivation				
ML 1	0.890			
ML 2	0.862	0.90	0.72	0.89
ML 3	0.828			
Total				0.90

Table 4. Concurrent validity

	Dimensions			
	1	2	3	4
Dimension 1	1			
Dimension 2	0.412**	1		
Dimension 3	0.428**	0.645**	1	
Dimension 4	0.395**	0.552**	0.467**	1

Note: $p < 0.01$

The analysis of concurrent validity for the four identified dimensions revealed a correlation matrix indicating statistically significant relationships among dimensions at the $p < 0.01$ level. Dimension 1 showed moderate but significant positive correlations with Dimension 2 (0.412), Dimension 3 (0.428), and Dimension 4 (0.395), suggesting consistent interrelations. Dimension 2 demonstrated stronger correlations with Dimension 3 (0.645) and Dimension 4 (0.552), indicating higher consistency with the other dimensions in the context of concurrent validity. Dimension 3 was also significantly correlated with Dimension 4 (0.467), reflecting a solid interconnection among constructs. The diagonal values of 1.00 represent perfect internal validity within each dimension. Collectively, the significant inter-dimensional correlations (all $p < 0.01$) support the conclusion that the four dimensions are meaningfully related and reinforce one another, confirming that the construct under investigation possesses adequate concurrent validity within the framework of this study.

Table 5. Means (SD) GA-ALQ scores by group

Dimension	Total	Control	Experiment	*F*	*p*
Total	3.78 (0.62)	3.65 (0.64)	3.91 (0.59)	5.128	0.027*
INT	3.45 (0.80)	3.30 (0.77)	3.60 (0.82)	4.872	0.031*
UI	3.28 (0.75)	3.25 (0.73)	3.31 (0.78)	0.092	0.762
FB	4.05 (0.70)	3.90 (0.72)	4.20 (0.67)	6.214	0.015*
ML	4.10 (0.68)	3.95 (0.70)	4.25 (0.65)	7.109	0.010**

Notes: $^{**}p < 0.01$, $^{*}p < 0.05$, INT = Interactivity, UI = User Interface, FB = Feedback, ML = Learning Motivation

An analysis of means and standard deviations (SD) was conducted to compare GA-ALQ scores between the total sample, control group, and experimental group across four dimensions. Statistical comparisons were performed using F-tests, with significance evaluated at the $p < 0.01$ and $p < 0.05$ levels. The overall mean GA-ALQ score was 3.78 (SD = 0.62), with the control group scoring 3.65 (SD = 0.64) and the experimental group scoring 3.91 (SD = 0.59). This difference was statistically significant, $F(1, N) = 5.128$, $p = 0.027$. For the INT (Interactivity) dimension, the total mean was 3.45 (SD = 0.80), with the control group at 3.30 (SD = 0.77) and the experimental group at 3.60 (SD = 0.82). The difference approached significance, $F = 4.872$, $p = 0.031$. The UI (User Interface) dimension yielded a total mean of 3.28 (SD = 0.75), with scores of 3.25 (SD = 0.73) in the control group and 3.31 (SD = 0.78) in the experimental group. No statistically significant difference was found, $F = 0.092$, $p = 0.762$. For the FB (Feedback) dimension, the total mean score was 4.05 (SD = 0.70), with the control group at 3.90 (SD = 0.72) and the experimental group at 4.20 (SD = 0.67). This difference was statistically significant, $F = 6.214$, $p = 0.015$. In the ML (Learning Motivation) dimension, the total mean was 4.10 (SD = 0.68), with the control group

scoring 3.95 (SD = 0.70) and the experimental group scoring 4.25 (SD = 0.65). This difference was also statistically significant, F = 7.109, p = 0.010.

This study confirms that the gamified assessment (GA-ALQ) has a robust factor structure, with four main dimensions—Interactivity, User Interface, Feedback, and Learning Motivation—explaining 66.8% of the data variance (Table 2). Further CFA validated this structure with excellent goodness-of-fit (χ^2/df = 2.098, RMSEA = 0.042, CFI = 0.952 for the correlated model; Table 3), indicating that the GA-ALQ questionnaire is both reliable (α = 0.90, CR = 0.873) and valid (AVE = 0.90). The Learning Motivation dimension showed the strongest contribution (λ = 0.92, p < 0.001), followed by Feedback, demonstrating that gamification elements such as immediate feedback and intrinsic motivation are crucial in enhancing student engagement. Group comparisons between the control and experimental groups (Table 5) revealed that the gamification intervention significantly improved scores in the Feedback (p = 0.015) and Learning Motivation (p = 0.010) dimensions, with the experimental group scoring higher on average (4.20 and 4.25) than the control group (3.90 and 3.95). These findings reinforce that gamified assessment is effective in enhancing motivation and feedback, although the User Interface did not show significant differences (p = 0.762), possibly due to a sub-optimised interface design. Overall, this study contributes to the development of a valid gamification measurement tool and highlights the importance of feedback and motivation in gamification-based learning contexts. Because of the limited sample size, these findings are more relevant within the experimental context of this study and cannot be generalised to the broader population without further research involving a larger and more representative sample.

The contribution of this study lies in the validation of the GA-ALQ questionnaire as a reliable instrument for evaluating gamified assessment, with practical implications for educational developers to focus on feedback and motivation elements. Theoretically, these findings support self-determination theory (Deci & Ryan, 2000) by demonstrating that relevant feedback and interactivity enhance students' intrinsic motivation.

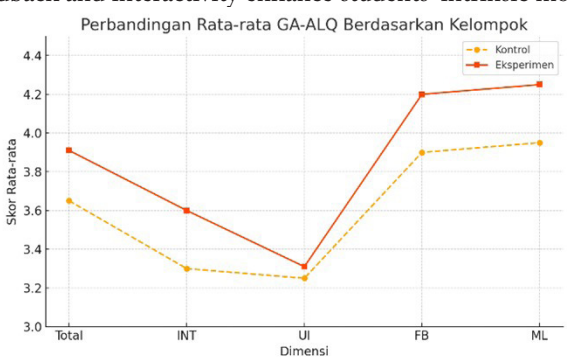

Figure 2. Average dimension scores by group

DISCUSSION

This study represents an early effort to develop and validate a questionnaire to measure students' perceptions of gamified assessment in Arabic language learning at the junior high school level in Indonesia, using Construct 2 as the development platform. The research is based on the belief that understanding how students perceive gamified assessment is essential for designing more effective and engaging learning experiences. Before choosing any learning strategies or assessment tools, teachers need to know how students respond to them. The resulting instrument, the Gamified Assessment in Arabic Learning Questionnaire (GA-ALQ), has shown strong validity and reliability in capturing key aspects of students' perceptions.

Validity and Reliability of the Questionnaire

The GA-ALQ went through a thorough process to test its validity and reliability. Although the sample size (60 students) was relatively small compared to recommendations in other studies (which often suggest 100–200 participants for stronger model stability) (Marsh et al., 2020), it met the minimum requirements for both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Despite this, the Kaiser-Meyer-Olkin (KMO) value of 0.78 falls within the "good" range (0.70–0.80), as recommended by Alavi et al. (2020b), and the result of Bartlett's Test of Sphericity (p = 0.000) indicates that the sample correlation matrix is suitable for factor analysis. The total variance explained reached 66.8%, exceeding the recommended minimum threshold of 60% (Tavakol & Wetzel, 2020). All factor loadings were above 0.50, meeting the criterion for meaningful interpretation (van Zyl & ten Klooster (2022). Confirmatory Factor Analysis (CFA) further demonstrated excellent model fit for both the correlated and hierarchical models. Fit indices such as the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) likely approached or exceeded the 0.90 benchmark, while the Root Mean Square Error of Approximation (RMSEA) remained below 0.08 (Shi & Maydeu-Olivares, 2020). These findings suggest that the four dimensions—interactivity, user interface, feedback, and learning motivation—are interrelated yet distinct, supporting a robust conceptualisation of the Gamified Assessment construct.

The internal reliability of the questionnaire was also strong. Cronbach's alpha reached 0.90, well above the commonly accepted threshold of 0.70 (Hayes & Coutts, 2020). Composite Reliability (CR) values were all ≥ 0.76, satisfying the minimum standard of 0.70 (Goodboy & Martin, 2020). The Average Variance Extracted (AVE) values were ≥ 0.46, approaching the conventional 0.50 cutoff (Kline et al., 2020). Although the AVE for the user interface dimension was slightly below 0.50, it remains acceptable given the sufficiently high CR, aligning with more flexible interpretations that permit lower AVE when composite reliability is strong (Cuzzocrea et al., 2020).

The moderate intercorrelations among the dimensions, ranging from 0.395 to 0.645, suggest that the dimensions are distinct yet mutually reinforcing in capturing students' perceptions—consistent with a multidimensional approach to psychometric measurement (Ruggeri et al., 2020). These correlation values indicate that each dimension contributes uniquely to the overall construct, while also highlighting the coherence among gamification elements. This finding aligns with recent research by Jaramillo-Mediavilla et al. (2024), which showed that gamification elements such as interactivity and feedback have a synergistic relationship with learning motivation. The questionnaire can be used either in its entirety to obtain a comprehensive view of students' perceptions or by analysing each dimension independently for more focused insights. However, strong recommendations support the full-scale application to ensure greater accuracy and consistency with the original psychometric design (Garofalo et al., 2021). This approach is also supported by recent findings from Raykov and Bluemke (2021), who emphasise the importance of holistic instrument use in evaluating educational technologies to minimise measurement bias.

Furthermore, the construct validity of this questionnaire is reinforced by its theoretical alignment with the educational gamification framework. Interactivity and feedback, which demonstrated high factor loadings, represent key elements shown to enhance student engagement (Bovermann & Bastiaens, 2020). Learning motivation, which contributed the most to the construct, underscores gamification's impact on intrinsic motivation, consistent with the Self-Determination Theory (Ryan & Deci, 2020). Although the user interface dimension showed the weakest path coefficient, it remains relevant—as intuitive visual design can support learning experiences, particularly in digital environments like Construct 2 (Ryan et al., 2021). Despite the limited sample size, this questionnaire provides a solid foundation for developing gamified assessments. Future studies could extend the validation with larger samples and incorporate advanced technologies, such as AI (Artificial Intelligence)-based analysis, to enhance measurement precision (Bao et al., 2023).

Analysis of User Interface Dimension Findings

The User Interface (UI) dimension demonstrated the weakest path coefficient in our analysis and showed no significant difference between control and experimental groups ($F = 0.092$, $p = 0.762$). This finding requires critical examination, as it contradicts expectations from user experience design literature suggesting that interface quality significantly impacts user satisfaction and technology acceptance (Ortiz-Crespo et al., 2021).

Several theoretical explanations may account for this unexpected finding. First, from a cognitive load perspective, when students are primarily focused on mastering complex Arabic language content, they may allocate limited attentional resources to evaluating aesthetic interface features. Functional affordances that directly support learning tasks

(interactivity, feedback) may become more salient than purely visual design elements. This interpretation aligns with Barana, Marchisio, and Sacchet's (Barana et al., 2021) finding that in educational contexts, interface functionality often trumps aesthetic appeal.

Second, cultural design preferences may moderate the relationship between UI quality and user perception. The visual design conventions employed in our Construct 2-based assessment reflected Western game design standards, which may not align with Indonesian students' aesthetic preferences or cultural expectations for educational materials. Cross-cultural interface design research has consistently demonstrated that visual preferences, color symbolism, and layout conventions vary significantly across cultures (Clinton-Lisell et al., 2023), yet our design process did not incorporate culturally-specific design principles.

Finally, the measurement model itself may be problematic. The UI dimension showed the lowest AVE (0.46), suggesting that our three UI items may not adequately capture the construct. Alternative conceptualisations of interface quality—incorporating dimensions such as aesthetic appeal, navigational efficiency, information architecture, and responsiveness—might reveal different patterns. Future research should develop more comprehensive UI measurement models informed by human-computer interaction theory. These contradictions between theoretical expectations and empirical findings highlight the complexity of technology-enhanced learning evaluation. Rather than dismissing the null finding, researchers should view it as theoretically informative, suggesting that relationships between interface design and learning experience may be more contextually contingent and multidimensional than currently conceptualised.

Items and Domains of the Questionnaire

The Gamified Assessment in Arabic Learning Questionnaire (GA-ALQ) comprises four core dimensions—interactivity, user interface, feedback, and learning motivation—which collectively reflect key gamification elements in educational contexts, specifically within Arabic language learning at the junior high school level in Indonesia. The interactivity dimension captures students' active engagement through features such as challenges, levels, and direct interaction. This is supported by Hailikari et al. (2022), who found that such game mechanics significantly enhance student involvement. Integrating these features into assessments built on platforms like Construct 2 allows for dynamic student participation, aligning with contemporary gamification strategies that emphasize contextual engagement (Pedro et al., 2020). The user interface dimension, encompassing visual design, layout, and navigation, plays a crucial role in ensuring a seamless and enjoyable user experience. This is echoed in gamified assessment development guidelines by Ohlms et al., (2024). However, findings from this study show that this dimension had the lowest path coefficient—possibly because students prioritize functionality over visual aesthetics. A similar pattern was observed in a recent

study by J. Li et al. (2024), which noted that when interactive elements dominate, interface design tends to be perceived as secondary.

The instant feedback dimension, which provides students with immediate input on their performance, represents a core element of gamification that helps learners identify both their strengths and areas for improvement (Barends et al., 2022). Within the gamified assessment developed using Construct 2, this feedback mechanism supports adaptive learning, consistent with data-driven learning approaches (Maraza-Quispe, 2024). The learning motivation dimension, which yielded the highest path coefficient in the analysis, underscores the significant impact of gamified assessments on students' interest, persistence, and intrinsic motivation in learning Arabic (Nicholson, 2015). This finding aligns with the Self-Determination Theory framework by Ryan and Deci (2020), which posits that intrinsic motivation can be enhanced through gamification elements such as rewards and progression—a conclusion further supported by Alzubi and Nazim (2024) and recent studies examining the influence of gamification on cognitive engagement (Meulenbroeks et al., 2024).

As the most influential dimension within the overall construct, learning motivation highlights the centrality of intrinsic motivation in educational gamification, reinforcing its role as a key predictor of successful game-based learning outcomes (Chen & Tu, 2021). In contrast, the user interface dimension demonstrated the lowest path coefficient, suggesting that students prioritise interactive and feedback-driven components that directly influence their learning experience over purely aesthetic features (Barana et al., 2021). In digital learning environments, functional interactivity tends to be more valued than graphical design (Clinton-Lisell et al., 2023), especially in subjects such as Arabic that demand intensive practice. Consequently, for Arabic language learning contexts, gamified assessments should prioritise enhancements in interactivity and feedback mechanisms, while still considering basic improvements in user interface design to ensure an optimal user experience (Jääskä et al., 2022)

Student Perceptions

ANOVA results revealed significant differences in students' perceptions of the gamified assessment, with the experimental group reporting more favourable views of interactivity, feedback, and learning motivation dimensions compared to the control group. Gamification has been shown to enhance student engagement and motivation through game elements such as challenges, points, and levels (Rivera & Garden, 2021). Statistically significant F-values for interactivity ($F = 4.872$, $p = 0.031$), feedback ($F = 6.214$, $p = 0.015$), and learning motivation ($F = 7.109$, $p = 0.010$) indicate that the gamified intervention had a meaningful impact on students' active engagement and interest in Arabic learning. On the other hand, the absence of a significant difference in the user interface dimension ($F = 0.092$, $p = 0.762$) suggests that visual features such

as layout and graphic design exert minimal influence on student perception compared to more functional interactive elements (Chen & Tu, 2021).

Game elements like simulation and instant feedback can transform classroom dynamics into more interactive and engaging experiences (Govender & Arnedo-Moreno, 2021). Gamification technologies, such as the one employed in this study, have the potential to improve content retention by up to 20% over traditional methods, particularly in linguistically complex subjects like Arabic. These findings underscore the transformative potential of gamified assessments in Arabic language instruction, particularly through the reinforcement of interactivity and motivation (Ritonga et al., 2023). To maximise impact, however, user interface design must be improved through User-Centered Design (UCD) principles that take student preferences into account—such as the use of high-contrast colors or more intuitive navigation features (Ortiz-Crespo et al., 2021). Moreover, the integration of learning analytics can support real-time monitoring of student perceptions, allowing for personalised content adjustments (Guzmán-Valenzuela et al., 2021). Future research should explore the relationship between these perceptions and measurable learning outcomes—such as vocabulary acquisition or grammatical proficiency—while also incorporating demographic variables like age or prior technology experience to enrich the analysis.

Limitations

This study has several significant limitations that need to be considered critically. First, the small sample size (n = 60, with 30 students for EFA and 30 for CFA) falls below the ideal recommendations for CFA (100–200 participants), increasing the risk of parameter instability, such as large standard errors or potential overfitting (Kline, 2015). Although the EFA indicated data adequacy (KMO = 0.78) and the simplicity of the model (12 items, 4 dimensions) helped reduce risk, the CFA results should be interpreted with caution. The generalisability of the findings is limited to the context of students at SMP Darul Faqih Indonesia and the Construct 2 platform; therefore, further validation with larger and more diverse samples is required. Second, this study did not account for other variables, such as students' initial proficiency levels or technological experience, which may influence perceptions. Third, the relationship between students' perceptions and learning outcomes (e.g., vocabulary mastery or grammar acquisition) was not examined, limiting the practical implications. Fourth, the weak influence of the User Interface dimension indicates shortcomings in platform design, which were not analysed in depth in this study. Future research should address these limitations by employing larger samples, analysing additional variables, and improving interface design.

Implications and Future Research

The GA-ALQ questionnaire provides a strong foundation for evaluating gamified assessment in Arabic language learning, with significant theoretical and pedagogical implications. Theoretically, this study reinforces Self-Determination Theory (Ryan & Deci, 2020) by demonstrating that gamification elements such as instant feedback and interactivity enhance students' intrinsic motivation. The findings also contribute to the educational gamification literature by validating a multidimensional measurement tool that encompasses interactivity, feedback, and motivation, while highlighting interface design weaknesses as an area requiring further attention. Pedagogically, educators can use this questionnaire to evaluate the effectiveness of gamified assessments and optimise instructional design, for example, by prioritising instant feedback and interactive challenges. Specific recommendations include adopting User-Centered Design to improve the interface, such as simplifying navigation or adding visually engaging elements for junior high school students (Ortiz-Crespo et al., 2021). The integration of learning analytics can also be employed to monitor student perceptions in real time, enabling content personalisation (Guzmán-Valenzuela et al., 2021).

Future research is recommended to: (1) revalidate the GA-ALQ questionnaire with larger and more diverse samples; (2) explore the relationship between student perceptions and learning outcomes, such as vocabulary mastery or grammar acquisition; (3) examine the influence of demographic variables such as age or technological experience; and (4) develop more responsive interface designs using a User-Centred Design approach to address the weaknesses of the User Interface dimension. Integrating gamification with conventional methods may create a more effective hybrid approach, while maintaining a strong emphasis on interactivity and feedback.

CONCLUSION

This study successfully developed and validated the Gamified Assessment in Arabic Learning Questionnaire (GA-ALQ) to measure students' perceptions of gamified assessment in Arabic language instruction at the junior high school level in Indonesia. The questionnaire comprises four core dimensions—Interactivity, User Interface, Feedback, and Learning Motivation—which were confirmed to be both valid and reliable through rigorous statistical analysis. However, several critical limitations must be acknowledged that constrain the generalisability and interpretability of our findings. Most significantly, the small sample size (n = 60, with only 30 for CFA) falls below optimal standards for confirmatory factor analysis, increasing the risk of parameter instability and limiting statistical power. Findings should therefore be considered preliminary, requiring replication with substantially larger and more diverse samples before the instrument can be confidently applied in high-stakes contexts.

Exploratory Factor Analysis (EFA) indicated that these four dimensions explained 66.8% of the total variance, with factor loadings exceeding 0.50. CFA further supported the model's adequacy, showing strong fit indices ($X^2/df < 2.5$, RMSEA < 0.08, CFI > 0.90). Internal reliability was excellent, with a Cronbach's alpha of 0.90, CR ≥ 0.76, and AVE ≥ 0.46, although the AVE for the User Interface dimension fell slightly below the ideal threshold.

The results indicated that Learning Motivation contributed the most to the overall gamification construct, followed by Feedback and Interactivity, while User Interface exerted the least influence. ANOVA analysis revealed significant differences between the experimental and control groups, with the experimental group reporting more positive perceptions of Interactivity ($p = 0.031$), Feedback ($p = 0.015$), and Learning Motivation ($p = 0.010$), though no significant difference was found for User Interface ($p = 0.762$). These findings affirm that gamification elements such as challenges, instant feedback, and rewards are effective in enhancing student engagement and motivation in Arabic language learning, whereas the visual aspects of the interface appear to have a lesser impact.

From a theoretical perspective, this study contributes to gamification research by demonstrating that student perceptions of game-based assessment are multidimensional, with motivational elements exerting stronger influence than interface aesthetics. This finding challenges technology-deterministic assumptions and highlights the importance of psychological mechanisms in educational technology effectiveness. The unexpected weakness of the UI dimension suggests that visual design may be less critical than functional interactivity in educational contexts—a finding that warrants theoretical elaboration and further empirical investigation. For educational practice, our findings suggest that educators developing gamified assessments for Arabic learning should prioritise design features that enhance interactivity, provide meaningful feedback, and support intrinsic motivation, while treating interface aesthetics as secondary (though still important for basic usability). However, these recommendations must be tentative given the study's limitations.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS

The researcher would like to express deep appreciation to the LPPM (Lembaga Penelitian dan Pengabdian Masyarakat) Universitas Negeri Malang for the funding they provided, as well as the administrative support that was granted during the period of this study. Specifically, this aid helped in the meticulous planning, implementation, and finalisation of this study. It is anticipated that the outcomes of this research will be significant in driving forward scholarly discourse, especially within the context of educational pedagogy on the Arabic language.

CONTRIBUTIONS OF AUTHORS

Mohammad Ahsanuddin: Conceptualisation, Validation, writing—original draft preparation.

Zawawi Ismail: Methodology, writing—review and editing.

Muhammad Ahsan Thoriq: Software, project administration.

Farika Riskiyah: Formal Analysis.

Rico Fenda Pradana: Investigation.

Arya Wahyu Pratama: Data curation, supervision.

Najma Auliya' Maulidah Ahsan: Visualisation, funding acquisition

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE (AI)

Artificial Intelligence (Claude.ai Sonnet 4.5) was utilised in the preparation of this manuscript to support various aspects of text development. Specifically, the AI assisted in drafting initial versions of certain sections, refining sentence structures, enhancing linguistic clarity and coherence, and ensuring optimal readability for the audience. Additionally, Claude.ai Sonnet 4.5 provided suggestions for more precise word choices and improved narrative flow to meet academic/professional writing standards. All AI-generated outputs were thoroughly reviewed and verified by the authors to ensure factual accuracy, contextual relevance, and alignment with the research objectives. No generative content was incorporated without human oversight, and the authors bear full responsibility for the accuracy, originality, and integrity of the entire content of this manuscript.

REFERENCES

Ahsanuddin, M., Ainin, M., Fauzan, M., Hassan, A. R., & Thoriq, M. A. (2025). The effectiveness gamification as an assessment method to enhance Arabic language learning competency in Indonesian junior high schools. *Forum for Linguistic Studies, 7*(8), 764–775. https://doi.org/10.30564/fls.v7i8.10417

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020a). Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing, 76*(9), 2209–2211. https://doi.org/10.1111/jan.14399

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020b). Exploratory factor analysis and principal component analysis in clinical studies: Which one should you use? *Journal of Advanced Nursing, 76*(8), 1886–1889. https://doi.org/10.1111/jan.14377

AlSaif, A. A., & Alsenany, S. (2015). Effects of interactive games on motor performance in children with spastic cerebral palsy. *Journal of Physical Therapy Science, 27*(6), 2001–2003. https://doi.org/10.1589/jpts.27.2001

Alzubi, A. A. F., & Nazim, M. (2024). Students' intrinsic motivation in EFL academic writing: Topic-based interest in focus. *Heliyon, 10*(1), e24169–e24169. https://doi.org/10.1016/j.heliyon.2024.e24169

Bao, H., Zhang, K., Yu, C., Li, H., Cao, D., Shu, H., Liu, L., & Yan, B. (2023). Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence. *BMC Oral Health, 23*(1), 191. https://doi.org/10.1186/s12903-023-02881-8

Barana, A., Marchisio, M., & Sacchet, M. (2021). Interactive feedback for learning mathematics in a digital learning environment. *Education Sciences, 11*(6), 279. https://doi.org/10.3390/educsci11060279

Barends, A. J., de Vries, R. E., & van Vugt, M. (2022). Construct and predictive validity of an assessment game to measure honesty-humility. *Assessment, 29*(4), 630–650. https://doi.org/10.1177/1073191120985612

Baso, Y. S. (2022). Reduce the foreign accent of Indonesian Arabic learners through a minimal pairs strategy. *Arabiyatuna: Jurnal Bahasa Arab, 6*(1), 193. https://doi.org/10.29240/jba.v6i1.4145

Bovermann, K., & Bastiaens, T. J. (2020). Towards a motivational design? Connecting gamification user types and online learning activities. *Research and Practice in Technology Enhanced Learning, 15*(1), 1. https://doi.org/10.1186/s41039-019-0121-4

Chen, C.-C., & Tu, H.-Y. (2021). The effect of digital game-based learning on learning motivation and performance under social cognitive theory and entrepreneurial thinking. *Frontiers in Psychology, 12*, 750711. https://doi.org/10.3389/fpsyg.2021.750711

Clinton-Lisell, V., Seipel, B., Gilpin, S., & Litzinger, C. (2023). Interactive features of e-texts' effects on learning: A systematic review and meta-analysis. *Interactive Learning Environments, 31*(6), 3728–3743. https://doi.org/10.1080/10494820.2021.1943453

Cuzzocrea, A., Scemama, A., Briels, W. J., Moroni, S., & Filippi, C. (2020). Variational principles in quantum Monte Carlo: The troubled story of variance minimization. *Journal of Chemical Theory and Computation, 16*(7), 4203–4212. https://doi.org/10.1021/acs.jctc.0c00147

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Fannakhosrow, M., Nourabadi, S., Ngoc Huy, D. T., Dinh Trung, N., & Tashtoush, M. A. (2022). A comparative study of information and communication technology (ICT)-based and conventional methods of instruction on learners' academic enthusiasm for L2 learning. *Education Research International, 2022,* 1–8. https://doi.org/10.1155/2022/5478088

Gao, R. (2021). he vocabulary teaching mode based on the theory of constructivism. *Theory and Practice in Language Studies, 11*(4), 442–446. https://doi.org/10.17507/tpls.1104.14

Garofalo, C., Weller, J. A., Kirisci, L., & Reynolds, M. D. (2021). Omega over alpha for reliability estimation of unidimensional communication measures. *Psychological Assessment, 33*(9), 890–903. https://doi.org/10.1037/pas0001023

Goodboy, A. K., & Martin, M. M. (2020). Omega over alpha for reliability estimation of unidimensional communication measures. *Annals of the International Communication Association, 44*(4), 422–439. https://doi.org/10.1080/23808985.2020.1846135

Govender, T., & Arnedo-Moreno, J. (2021). An approach to teaching psychiatry to medical students in the time of COVID-19. *Sustainability, 13*(12), 6679. https://doi.org/10.3390/su13126679

Guerandel, A., McCarthy, N., McCarthy, J., Mulligan, D., Lane, A., & Malone, K. (2021). An approach to teaching psychiatry to medical students in the time of COVID-19. *Irish Journal of Psychological Medicine, 38*(4), 293–299. https://doi.org/10.1017/ipm.2020.87

Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., & Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: A preponderance of analytics but very little learning? *International Journal of Educational Technology in Higher Education, 18*(1), 23. https://doi.org/10.1186/s41239-021-00258-x

Hailikari, T., Virtanen, V., Vesalainen, M., & Postareff, L. (2022). Student perspectives on how different elements of constructive alignment support active learning. *Active Learning in Higher Education, 23*(3), 217–231. https://doi.org/10.1177/1469787421989160

Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research, 109*, 101–110. https://doi.org/10.1016/j.jbusres.2019.11.069

Hanafy, S. M., Jumaa, M. I., & Arafa, M. A. (2021). A comparative study of online learning in response to the coronavirus disease 2019 pandemic versus conventional learning. *Saudi Medical Journal, 42*(3), 324–331. https://doi.org/10.15537/smj.2021.42.3.20200741

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hesham, S., Afsah, O., & Baz, H. (2024). Exploring phonological awareness skills in Arabic-speaking children with borderline intellectual functioning and poor reading. *Egyptian Journal of Otolaryngology, 40*(1), 58. https://doi.org/10.1186/s43163-024-00605-5

Hestiningsih, I. (2019). Rancang bangun game pembelajaran bahasa Arab dasar "Aladin" (Arabic learning by exploring) dengan konsep petualangan berbasis Android. *Orbith: Majalah Ilmiah Pengembangan Rekayasa Dan Sosial, 14*(3), 139. https://doi.org/10.32497/orbith.v14i3.1311

Jääskä, E., Lehtinen, J., Kujala, J., & Kauppila, O. (2022). Game-based learning and students' motivation in project management education. *Project Leadership and Society, 3*, 100055. https://doi.org/10.1016/j.plas.2022.100055

Jaramillo-Mediavilla, L., Basantes-Andrade, A., Cabezas-González, M., & Casillas-Martín, S. (2024). Impact of gamification on motivation and academic performance: A systematic review. *Education Sciences, 14*(6), 639. https://doi.org/10.3390/educsci14060639

Khasawneh, Y. J. A., Khasawneh, N., & Khasawneh, M. A. S. (2024). Exploring the long-term effects: Retention and transfer of skills in gamified learning environment. *International Journal of Data and Network Science, 8*(1), 195–200. https://doi.org/10.5267/j.ijdns.2023.10.004

Khoo, E., Le, A., & Lipp, M. J. (2023). Learning games: A new tool for orthodontic education. *International Journal of Environmental Research and Public Health, 20*(3), 2039. https://doi.org/10.3390/ijerph20032039

Kline, P., Saggio, R., & Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica, 88*(5), 1859–1898. https://doi.org/10.3982/ECTA16410

Lai, M. H. C. (2021). Composite reliability of multilevel data: It's about observed scores and construct meanings. *Psychological Methods, 26*(1), 90–102. https://doi.org/10.1037/met0000287

Lehmann, R., & Vogt, B. (2024). Improving Likert scale big data analysis in psychometric health economics: Reliability of the new compositional data approach. *Brain Informatics, 11*(1), 19. https://doi.org/10.1186/s40708-024-00232-z

Li, J., Bai, J., Zhu, S., & Yang, H. H. (2024). Game-based assessment of students' digital literacy using evidence-centered game design. *Electronics, 13*(2), 385. https://doi.org/10.3390/electronics13020385

Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review. *Educational Technology Research and Development, 72*(2), 765–796. https://doi.org/10.1007/s11423-023-10337-7

Limantara, N., Meyliana, Gaol, F. L., & Prabowo, H. (2023). Designing gamified learning management systems for higher education. *International Journal of Information and Education Technology, 13*(1), 25–32. https://doi.org/10.18178/ijiet.2023.13.1.1776

Ma, J., & Feng, B. (2021). Integrated design of graduate education information system of universities in digital campus environment. *Wireless Communications and Mobile Computing, 2021*, 8357488. https://doi.org/10.1155/2021/8357488

Magistretti, S., Dell'Era, C., & Verganti, R. (2020). Searching for the right application: A technology development review and research agenda. *Technological Forecasting and Social Change, 151*, 119879. https://doi.org/10.1016/j.techfore.2019.119879

Maraza-Quispe, B. (2024). Impact of the use of gamified online tools: A study with kahoot and quizizz in the educational context. *International Journal of Information and Education Technology, 14*(1), 132–140. https://doi.org/10.18178/ijiet.2024.14.1.2033

Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research, 55*(1), 102–119. https://doi.org/10.1080/00273171.2019.1602503

Mazeas, A., Duclos, M., Pereira, B., & Chalabaev, A. (2022). Evaluating the effectiveness of gamification on physical activity: Systematic review and meta-analysis of randomized controlled trials. *Journal of Medical Internet Research, 24*(1), e26779. https://doi.org/10.2196/26779

Meulenbroeks, R., van Rijn, R., & Reijerkerk, M. (2024). Fostering secondary school science students' intrinsic motivation by inquiry-based learning. Research in Science Education, 54(3), 339–358. https://doi.org/10.1007/s11165-023-10139-0

Min, S., Atan, N. A., & Habibi, A. (2025). Gamification with self-determination theory to foster intercultural communicative competence and intrinsic motivation. International Journal of Evaluation and Research in Education, 14(3), 1985–1994. https://doi.org/10.11591/ijere.v14i3.29858

Muzdalifah, Z., Khasairi, M., & Kholisin, K. (2021). Development of the arabic grammar (nahwu) textbook al-ajurumiyyah al-qur'aniyyah based on the scaffolding-structure. Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature, 4(2), 153–164. https://doi.org/10.22219/jiz.v4i2.16229

Naye, F., Décary, S., & Tousignant-Laflamme, Y. (2022). Development and content validity of a rating scale for the pain and disability drivers management model. Archives of Physiotherapy, 12(1), 14. https://doi.org/10.1186/s40945-022-00137-2

Ng, L. L., & Raghbir, R. S. A. (2021). Learning english vocabulary via computer gaming. Issues in Language Studies, 10(1), 93–109. https://doi.org/10.33736/ils.2708.2021

Ohlms, M. L., Melchers, K. G., & Kanning, U. P. (2024). Can we playfully measure cognitive ability? Construct-related validity and applicant reactions. International Journal of Selection and Assessment, 32(1), 91–107. https://doi.org/10.1111/ijsa.12450

Oke, A. E., Aliu, J., Tunji-Olayeni, P., & Abayomi, T. (2024). Application of gamification for sustainable construction: An evaluation of the challenges. Construction Innovation, 24(4), 1066–1084. https://doi.org/10.1108/CI-09-2022-0247

Oliveira, W., Hamari, J., Shi, L., Toda, A. M., Rodrigues, L., Palomino, P. T., & Isotani, S. (2023). Tailored gamification in education: A literature review and future agenda. Education and Information Technologies, 28(1), 373–406. https://doi.org/10.1007/s10639-022-11122-4

Ortiz-Crespo, B., Steinke, J., Quirós, C. F., van de Gevel, J., Daudi, H., Gaspar Mgimiloko, M., & van Etten, J. (2021). User-centred design of a digital advisory service: enhancing public agricultural extension for sustainable intensification in Tanzania. International Journal of Agricultural Sustainability, 19(5–6), 566–582. https://doi.org/10.1080/14735903.2020.1720474

Park, C. G. (2023). Implementing alternative estimation methods to test the construct validity of Likert-scale instruments. Korean Journal of Women Health Nursing, 29(2), 85–90. https://doi.org/10.4069/kjwhn.2023.06.14.2

Pedro, A., Pham, H. C., Kim, J. U., & Park, C. (2020). Development and evaluation of context-based assessment system for visualization-enhanced construction safety education. International Journal of Occupational Safety and Ergonomics, 26(4), 811–823. https://doi.org/10.1080/10803548.2018.1553377

Raykov, T., & Bluemke, M. (2021). Examining multidimensional measuring instruments for proximity to unidimensional structure using latent variable modeling. Educational and PsychologicalMeasurement,81(2),319–339.https://doi.org/10.1177/0013164420940764

Reyes, E., Gálvez, J. C., & Enfedaque, A. (2021). Learning course: Application of gamification in teaching construction and building materials subjects. Education Sciences, 11(6), 287. https://doi.org/10.3390/educsci11060287

Ritonga, M., Wahyuni, S., & Novigator, H. (2023). The future of Arabic language learning for non-Muslims as an actualization of Wasathiyah Islam in Indonesia. F1000Research, 12, 27. https://doi.org/10.12688/f1000research.125760.1

Rivera, E. S., & Garden, C. L. P. (2021). Gamification for student engagement: A framework. Journal of Further and Higher Education, 45(7), 999–1012. https://doi.org/10.1080/0309877X.2021.1875201

Rohan, R., Pal, D., Funilkul, S., Chutimaskul, W., & Eamsinvattana, W. (2021). How gamification leads to continued usage of MOOCs? A theoretical perspective. *IEEE Access, 9*, 108144–108161. https://doi.org/10.1109/ACCESS.2021.3102293

Rohman, H. (2019). Pengembangan media construct 2 dalam pembelajaran Qira'ah di Madrasah Tsanawiyah Negeri 1 Yogyakarta. *EDULAB: Majalah Ilmiah Laboratorium Pendidikan, 4*(1), 25–46. https://doi.org/10.14421/edulab.2019.41-03

Roumei, F., & Feng, M. (2012). The use of multimedia in language teaching at Binus University. *Humaniora, 3*(2), 392–399. https://doi.org/10.35824/sjrs.v6i1.24967

Ruggeri, K., Garcia-Garzon, E., Maguire, Á., Matz, S., & Huppert, F. A. (2020). Well-being is more than happiness and life satisfaction: A multidimensional analysis of 21 countries. *Health and Quality of Life Outcomes, 18*(1), 192. https://doi.org/10.1186/s12955-020-01423-y

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Ryan, R. M., Deci, E. L., Vansteenkiste, M., & Soenens, B. (2021). Building a science of motivated persons: Self-determination theory's empirical approach to human experience and the regulation of behavior. *Motivation Science, 7*(2), 97–110. https://doi.org/10.1037/mot0000194

Schöbel, S. M., Janson, A., & Söllner, M. (2020). Capturing the complexity of gamification elements: A holistic approach for analysing existing and deriving novel gamification designs. *European Journal of Information Systems, 29*(6), 641–668. https://doi.org/10.1080/0960085X.2020.1796531

Schüttpelz-Brauns, K., Hecht, M., Hardt, K., Karay, Y., Zupanic, M., & Kämmer, J. E. (2020). Institutional strategies related to test-taking behavior in low stakes assessment. *Advances in Health Sciences Education, 25*(2), 321–335. https://doi.org/10.1007/s10459-019-09928-y

Shakoor, A., Mohansingh, C., van der Boon, R. M. A., Brugts, J. J., & Schaap, J. (2025). Gamification and its potential for better engagement in the management of heart failure or quality of care registries: A viewpoint. *Current Heart Failure Reports, 22*(1), 10–15. https://doi.org/10.1007/s11897-024-00692-8

Shi, D., & Maydeu-Olivares, A. (2020). The effect of estimation methods on SEM fit indices. *Educational and Psychological Measurement, 80*(3), 421–445. https://doi.org/10.1177/0013164419885164

Silic, M., & Lowry, P. B. (2020). Using design-science based gamification to improve organizational security training and compliance. *Journal of Management Information Systems, 37*(1), 129–161. https://doi.org/10.1080/07421222.2019.1705512

Sinar, T. S., Zein, T. T., Huszka, B., Yusuf, M., Maharani, P., & Sanjaya, D. (2024). Learning English literacy through video games: A multimodal perspective. *Journal of Curriculum and Teaching, 13*(1), 102–118. https://doi.org/10.5430/jct.v13n1p102

Tavakol, M., & Wetzel, A. (2020). Factor analysis: A means for theory and instrument development in support of construct validity. *International Journal of Medical Education, 11*, 245–247. https://doi.org/10.5116/ijme.5f96.0f4a

van Zyl, L. E., & ten Klooster, P. M. (2022). Exploratory structural equation modeling: Practical guidelines and tutorial with a convenient online tool for mplus. *Frontiers in Psychiatry, 12*, 795672. https://doi.org/10.3389/fpsyt.2021.795672

Vrcelj, A., Hoić-Božić, N., & Dlab, M. H. (2023). Use of gamification in primary and secondary education: A systematic literature review. *International Journal of Educational Methodology, 9*(1), 13–27. https://doi.org/10.12973/ijem.9.1.13

Wainwright, B. R., Allen, M. L., & Cain, K. (2020). Narrative comprehension and engagement with e-books vs. paper-books in autism spectrum condition. *Autism and Developmental Language Impairments, 5*. https://doi.org/10.1177/2396941520917943

Yamashita, T. (2022). Analyzing Likert scale surveys with Rasch models. *Research Methods in Applied Linguistics, 1*(3), 100022. https://doi.org/10.1016/j.rmal.2022.100022

Yoshigami, K., Hayashi, T., Tsunoda, M., Uwano, H., Sasaki, S., & Matsumoto, K. (2019). How does time conscious rule of gamification affect coding and review? *IEICE Transactions on Information and Systems, E102D*(12), 2435–2440. https://doi.org/10.1587/transinf.2019MPL0002

Zeybek, N., & Saygı, E. (2024). Gamification in education: Why, where, when, and how? A systematic review. *Games and Culture, 19*(2), 237–264. https://doi.org/10.1177/15554120231158625

APPENDIX

Appendix A: Items of Gamified Assessment in Arabic Learning Questionnaire (GA-ALQ)

Interactivity (INT)

1. The gamified assessment allows me to actively engage with Arabic language content.
2. I feel challenged to complete the tasks in the gamified assessment.
3. The gamified assessment makes learning Arabic more enjoyable.

User Interface (UI)

1. The layout of the gamified assessment is easy to understand and use.
2. I find the visual design of the gamified assessment appealing.
3. Navigation in the gamified assessment is clear and not confusing.

Feedback (FB)

1. The gamified assessment provides instant feedback on my answers.
2. The feedback helps me understand my mistakes.
3. Feedback in the gamified assessment motivates me to improve my learning.

Learning Motivation (ML)

1. The gamified assessment increases my interest in learning Arabic.
2. I feel more motivated to complete Arabic language tasks through gamified assessment.
3. The gamified assessment makes me want to keep learning Arabic.

Appendix B: Expert Evaluation

Evaluation Criteria:

1. Clarity: Is the item written clearly and easily understood by junior school students?
2. Relevance: Is the item relevant to the intended dimension?
3. Representativeness: Does the item represent an important aspect of that dimension?

Rating Scale:

1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree

DIMENSION 1: INTERACTIVITY

Operational Definition: The ability of gamified assessment to actively engage students through challenges, levels, and direct interaction with Arabic language content.

No	Item	Clarity	Relevance	Representativeness	Decision (Accept/ Revise/ Reject)	Revision suggestions
INT1	Gamified assessment allows me to interact actively with Arabic language content.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
INT2	I feel challenged to complete tasks in the gamified assessment.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
INT3	Gamified assessment makes Arabic language learning more enjoyable.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
INT4*	I can choose the order of questions I want to work on in the gamified assessment.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
INT5*	Gamified assessment encourages me to try again when I fail.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
INT6*	Game elements in the assessment make me more focused on learning.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	

Note: *Items deleted after expert evaluation

DIMENSION 2: USER INTERFACE

Operational Definition: Visual design, layout, and ease of navigation in gamified assessment.

No	Item	Clarity	Relevance	Representativeness	Decision	Revision suggestions
UI1	The layout of the gamified assessment is easy to understand and use.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	

UI2	I find the visual design of the gamified assessment attractive.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject
UI3	Navigation in the gamified assessment is clear and not confusing.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject
UI4*	The colours used in the gamified assessment attract my attention.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject
UI5*	The text size in the assessment is easy to read.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject

Note: *Items deleted after expert evaluation

DIMENSION 3: FEEDBACK

Operational Definition: Provision of immediate information about student performance that helps them understand errors and improve learning.

No	Item	Clarity	Relevance	Representativeness	Decision	Revision suggestions
FB1	Gamified assessment provides immediate feedback on my answers.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
FB2	Feedback helps me understand my mistakes.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
FB3	Feedback in the gamified assessment motivates me to improve my learning.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
FB4*	The feedback provided is detailed enough to help me learn.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	

Note: *Items deleted after expert evaluation

DIMENSION 4: LEARNING MOTIVATION

Operational Definition: The impact of gamified assessment on students' interest, persistence, and intrinsic motivation in learning Arabic.

No	Item	Clarity	Relevance	Representativeness	Decision	Revision suggestions
ML1	Gamified assessment increases my interest in learning Arabic.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
ML2	I feel more motivated to complete Arabic tasks through gamified assessment.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
ML3	Gamified assessment makes me want to continue learning Arabic.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
ML4*	Gamified assessment helps me learn faster.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
ML5*	I am more confident in my Arabic language abilities after using gamified assessment.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	
ML6*	Gamified assessment makes me more diligent in practicing Arabic.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	□ Accept □ Revise □ Reject	

Note: *Items deleted after expert evaluation

Appendix C: Expert Assessment Results and Inter-Rater Reliability

Summary of Evaluator Profiles

No	Code	Expertise	Experience (Years)
1	E1	Arabic Language Education	15
2	E2	Arabic Language Education	12
3	E3	Arabic Language Education	10
4	E4	Gamification & Educational Technology	8
5	E5	Gamification & Multimedia	9
6	T1	Arabic Language Teacher	7
7	T2	Arabic Language Teacher	11
8	T3	Arabic Language Teacher	6

Aggregate Expert Assessment Table (Percentage Agreement)

Dimension	Item	E1	E2	E3	E4	E5	T1	T2	T3	Agree (*n*)	Agreement (%)	Decision
Interactivity	INT1	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	INT2	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	INT3	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	INT4	✓	✗	✓	✓	✗	✓	✗	✓	5/8	62.5	REJECT
	INT5	✓	✓	✗	✓	✓	✗	✓	✗	5/8	62.5	REJECT
	INT6	✓	✓	✓	✗	✓	✓	✗	✓	6/8	75.0	REJECT
User Interface	UI1	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	UI2	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	UI3	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	UI4	✓	✗	✓	✗	✓	✓	✗	✓	5/8	62.5	REJECT
	UI5	✓	✓	✗	✓	✓	✗	✓	✓	6/8	75.0	REJECT
Feedback	FB1	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	FB2	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	FB3	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	FB4	✓	✓	✗	✓	✗	✓	✓	✗	5/8	62.5	REJECT
Learning Motivation	ML1	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	ML2	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	ML3	✓	✓	✓	✓	✓	✓	✓	✓	8/8	100	ACCEPT
	ML4	✓	✗	✓	✓	✗	✓	✗	✓	5/8	62.5	REJECT
	ML5	✓	✓	✗	✓	✓	✗	✓	✓	6/8	75.0	REJECT
	ML6	✓	✗	✓	✓	✗	✓	✓	✗	5/8	62.5	REJECT

Note: ✓ = Agree (score ≥ 4), ✗ = Disagree (score < 4); Acceptance Criteria: Agreement ≥ 80%

Pilot Test Results

Aspect	Findings
Completion time	Mean: 8.5 minutes (Range: 6–12 minutes)
Items difficult to understand	INT4, INT5, ML4, FB4 (4–6 students reported difficulty)
Items easy to understand	All items that were ultimately retained (INT1-3, UI1-3, FB1-3, ML1-3)
Technical Issues	None

Revisions Based on Pilot Test

Decisions:

1. Items frequently reported as difficult to understand will be evaluated more strictly in expert assessment
2. Items with negative feedback from ≥40% of students will be candidates for deletion
3. Questionnaire instructions will be clarified by adding a definition of "gamified assessment"