Educazione: Journal of Education and Learning
Vol 03, No 01, 2025 | Page (67-75)
Doi: https://doi.org/10.61987/educazione.2504

OPEN ACCESS

Bridging *Pesantren* Pedagogy and Digital Assessment through Wordwall Gamification: A Hybrid Model for Arabic Language Learning

Achmad Farouq Abdullah*, Sofwan Manaf, Tutik Lestari

Universitas Darunnajah, Indonesia

Email Corresponding : farouqabdullah@darunnajah.ac.id

Abstract

This study aims to test the effectiveness of Wordwall gamification within the Hybrid Evaluation Model for Arabic language learning at the Rafah Bogor Islamic Boarding School, as a dissemination of the *Durusullughah* 4.0 innovation based on Problem-Based Learning (PBL) that was previously developed. Using a quantitative approach with a quasi-experimental, nonequivalent control group design, this study involved 60 students of class I TMI, divided into experimental groups (n=30) who received Wordwall-PBL-based learning and control groups (n=30) who used conventional methods. The instrument is an Arabic-language competency test (*mufrodat* and *qawa'id*) that has been validated with an alpha coefficient (α) of 0.84. The results of the Independent Samples t-test analysis showed a significant difference between the two groups (t (58) =5.812, p<0.001), with the average posttest score of the experimental group (M=81.23; SD=6.41) significantly higher than that of the control group (M=70.87; SD=7.54). The N-Gain Score of the experimental group was 0.65 (moderate-high category) compared to 0.36 in the control group. The perceptions of students and teachers towards the media showed very high acceptance rates (86.9% and 89%, respectively). These findings confirm that the PBL Wordwall model can be effectively disseminated in other Islamic boarding schools and strengthen Arabic language learning without compromising the typical values of Islamic boarding schools.

Article History
Received : July 2025
Revised : August 2025
Accepted : November 2025

Keywords:
Wordwall, Problem-Based Learning, Hybrid Assessment Model, Arabic, Pesantren,

How to Cite:
Abdullah, A. F., Manaf, S., & Lestari, T. (2025). Bridging *Pesantren* Pedagogy and Digital Assessment through Wordwall Gamification: A Hybrid Model for Arabic Language Learning. *Educazione: Journal of Education and Learning, 3*(1), 67–75.

INTRODUCTION

Learning Arabic in Islamic boarding schools is the main foundation for students to understand Islamic sources directly from the original text. However, the real challenge faced by many Islamic boarding schools today is the gap between the evaluation methods used and the learning needs of the digital generation of students. Survey results from several Islamic boarding schools show that 78% of students want more interactive digital assessment tools, while 76% of teachers acknowledge the limitations of conventional methods in fostering critical thinking and problem-solving skills (Zulkifli & Lestari, 2024; Hidayat & Syarifuddin, 2020). This condition indicates an urgent need for pedagogical innovation to bridge the scientific tradition of Islamic boarding schools and the demands of 21st-century digital literacy.

In response to this need, a research team from Darunnajah University Jakarta has developed an innovation, *Durusullughah* 4.0, a hybrid evaluation model that integrates the Wordwall platform with a Problem-Based Learning (PBL) approach based on the MDA (Mechanics, Dynamics, Aesthetics) framework through Research and Development (R&D) research with the ADDIE model at the Darunnajah Islamic Boarding School in Jakarta. The results of the development showed a high level of feasibility, with the validation of media experts of 86.9% (very good) and material experts of 77% (good), as well as an increase in the

average score of students from 8.0 to 9.3 in a limited trial (Abdullah et al., 2025). This success opens up opportunities to disseminate the model in other Islamic boarding schools to test its effectiveness and generalizability.

The Rafah Bogor Islamic Boarding School was chosen as the location for dissemination because it has representative characteristics: a modern Islamic boarding school with an integrative curriculum that combines general and religious education, with a large student base and adequate technological infrastructure. More importantly, the Rafah Islamic Boarding School has never implemented a gamification-based evaluation model before, making it an ideal context to test the effectiveness of the Wordwall PBL model in a setting different from the original development site. This selection is also in line with the need for external validation (external validity) of an educational innovation before it can be recommended for wider application.

The Wordwall platform provides a variety of interactive game formats, including quizzes, matching, puzzles, and anagrams, that can be customized using native Arabic characters without transliteration (Mardiyah & Rachman, 2025; Chandra & Kun, 2024). Combined with the PBL approach, which encourages students to solve contextual problems collaboratively and reflectively, this model not only measures learning outcomes in a formative way but also creates a meaningful and enjoyable learning experience (Othman & Shah, 2013; Lestari & BB, 2024). Theoretically, this integration rests on the principle of social constructivism (Vygotsky), which emphasizes that optimal learning occurs through active interaction, scaffolding, and shared reflection (Zuo et al., 2023).

Previous studies on gamification in language learning have shown consistently positive results. Alatas et al. (2024) found that the use of Wordwall significantly improved vocabulary mastery in a quasi-experimental setting. Do and Huynh (2024) reported on students' positive perceptions of Wordwall in strengthening vocabulary. Maulina and Mawardi (2022) confirm that Wordwall is effective in improving critical thinking skills when combined with the right learning design. However, no research has specifically tested the effectiveness of the integrated Wordwall–PBL model in the context of dissemination in Islamic boarding schools, distinct from the development sites, using a more rigorous quantitative design.

Based on these gaps, this study aims to: (1) analyze the differences in Arabic language learning outcomes of Rafah Bogor Islamic Boarding School students between the group that uses the PBL Wordwall model (experimental group) and the group that uses conventional methods (control group); (2) measure the amount of improvement in Arabic language competence (N-Gain Score) in both groups; and (3) describe the perception of students and teachers towards the implementation of the PBL Wordwall model as a hybrid assessment model at the Rafah Bogor Islamic Boarding School. The findings of this study are expected to provide strong empirical evidence to support the wider dissemination of the PBL Wordwall model within the Indonesian *pesantren* ecosystem.

RESEARCH METHOD

Research Design

This study uses a quantitative approach with a quasi-experimental design of nonequivalent control groups. This design was chosen in light of the research context in *pesantren*, which does not allow full randomization in group division due to the administrative limitations of the classes that have already been formed. Nonetheless, the use of pretests allows control over differences in initial abilities between the two groups.

Table 1. Nonequivalent Control Group Research Design

Groups	Pretest	Treatment	Posttest
Experiment (E)	O_1	X: Wordwall PBL	O_2
Control (K)	O_3	– (Conventional)	O_4

Description: O_1, O_3 = pretest; O_2, O_4 = posttest; X = Wordwall–PBL treatment

Population and Sample

The research population comprises all students in class I TMI (*Tarbiyatul Mu'allimin Al-Islamiyah*) at the Rafah Bogor Islamic Boarding School for the 2024/2025 academic year, totaling 120 students across four parallel classes. Sampling was carried out using a purposive sampling technique based on class equality: two classes were selected based on the equivalence of the average Arabic score of the previous semester (difference < 2 points), namely Class I-A as the experimental group (n=30) and Class I-C as the control group (n=30), so that the total sample was 60 students.

The inclusion criteria for the sample include: (1) actively registered as a class I student of TMI; (2) have never used the Wordwall platform before; (3) have Arabic language scores in the previous semester in the range of 60-80 (fairly good category); and (4) willing to participate in all research sessions. The equivalence of the two groups was verified using Levene's Test for homogeneity and the Mann-Whitney U test on the pretest data administered before treatment.

Research Instruments

Data collection instruments are of two types. First, the standardized Arabic competency test consists of: (a) a *mufrodat* (vocabulary) test with 25 multiple-choice questions that measure the ability to identify, comprehend, and use vocabulary according to the context of the *Durusul Lughah* Volume 1 material; and (b) a *qawa'id* (grammar) test with 25 questions that measure understanding of the basic sentence structure (*mubtada-khabar*, number of *fi'liyah*, and preposition). A total score of 50 items is converted to a scale of 0-100.

The instrument was validated through two stages: (1) the validity of the content by two Arabic language experts using a grid that refers to the Core Competencies of the *Durusul Lughah* Curriculum; (2) trials on 30 non-sample students for the analysis of item validity (Pearson's Product Moment correlation, r-criticism = 0.361), reliability (Alpha Cronbach α = 0.84), difficulty level (P = 0.30-0.70), and differentiating power (D $\geq$ 0.30). The result: 47 out of 50 items were declared valid and met all eligibility criteria. Second, a perception questionnaire (Likert scale of 1-5) for students (21 items) and teachers (10 items) which measured aspects of attractiveness, ease of use, relevance, and impact on learning motivation.

Research Procedure

The research was carried out for ten effective weeks (February-April 2025) with the following stages. Weeks 1-2: Pretest in both groups, research socialization, and technical training on the use of Wordwall for teachers and facilitators of the experimental group. Weeks 3-9: Learning implementation; the two groups received identical *Durusul Lughah* Volume 1 material, but the experimental group used eight PBL Wordwall activities that had been developed from previous research (*Mufradat Al-Madrasah*, *Tarkib Mubtada-Khabar, Al-Adad Wa Al-Ma'dud, Asma Al-Adawat, Al-Alwan, Ayna Qalamuka, Al-Jam' wa al-Mufrad, Muraja'ah al-Mufrada*t) as many as three sessions per week (45 minutes/session), while the control group used the lecture, memorization and practice written questions. Week 10: Posttest for both groups and a perception questionnaire.

Data Analysis Techniques

Data were analyzed with SPSS version 26.0 through the following stages: (1) Descriptive statistics (mean, standard deviation, min-max value) for the pretest and posttest of both groups; (2) Prerequisite test: normality with Shapiro-Wilk and homogeneity with Levene's Test; (3) Different tests: Independent Samples t-test to compare the posttest of the two groups, and Paired Samples t-test to see intra-group improvement; (4) Effect size with Cohen's d; (5) N-Gain Score (g = (posttest – pretest) / (Max-pretest score)) with criteria: g > 0.70 (high), 0.30 $\leq$ g $\leq$ 0.70 (moderate), g < 0.30 (low); (6) Descriptive analysis of perception questionnaires on a percentage scale.

RESULT AND DISCUSSION

Result

Prerequisite Test Results

Because the validity of an Independent Samples *t*-test rests on the assumptions of normality and homogeneity, both were examined before any hypothesis testing proceeded. The Shapiro-Wilk test was applied to the pretest and posttest distributions of each group, and the results are reported in Table 2.

Table 2. Shapiro-Wilk Normality Test Results

Groups	Tests	Statistics	df	Sig.
Experiments	Pretest	.971	30	.572
	Posttest	.968	30	.488
Controls	Pretest	.965	30	.417
	Posttest	.962	30	.359

Note: Sig. > 0.05 indicates normally distributed data

Every distribution returned a significance value well above the 0.05 threshold, indicating that the scores in both groups were normally distributed at pretest and at posttest. Levene's test then confirmed equality of variances, with no significant departure from homogeneity at pretest ($F = 1.187$, $p = 0.281$) or at posttest ($F = 1.934$, $p = 0.170$). With both assumptions satisfied, parametric comparison of the two groups was warranted.

Descriptive Statistics of Learning Outcomes

Table 3 summarises the central tendency and dispersion of scores for both groups before and after treatment.

Table 3. Descriptive Statistics of Pretest and Posttest of Both Groups

Group	N	Pretest Mean (SD)	Posttest Mean (SD)	Mean Difference (Δ)	Percentage Increase (%)
Experimental	30	60.87 (5.63)	81.23 (6.41)	20.36	33.45
Control	30	61.10 (5.92)	70.87 (7.54)	9.77	15.99

At the outset the two groups were statistically indistinguishable, separated by a mere 0.23 points on the pretest, which confirms that the class-matching procedure produced genuinely equivalent baselines. The divergence emerged only after instruction. Across the seven weeks of core learning, the experimental group advanced by 20.36 points, rising from 60.87 to 81.23, whereas the control group gained only 9.77 points, moving from 61.10 to 70.87. The resulting gap of 10.59 points between the two gains offers the first clear indication that the *Wordwall-PBL* model produced learning beyond what conventional instruction achieved over the same period.

Hypothesis Test Results

The main hypothesis test was carried out using an Independent-Samples t-test on posttest data. The hypotheses tested were H_0: there is no significant difference in learning outcomes between the experimental and control groups; H_a: there is a significant difference in learning outcomes between the experimental and control groups. The results of the analysis are presented in Table 4.

Table 4. Results of Independent Samples t-test Data Posttest

Comparison	t	df	p (2-tailed)	Δ Average	95% CI
Experiment vs Control	5.812	58	.000	10.36	[6.78; 13.94]

The test yielded t (58) = 5.812 with $p < 0.001$, far below the 0.05 criterion, so the null hypothesis was rejected. The posttest means differed by 10.36 points, and the 95% confidence interval of [6.78; 13.94] excludes zero throughout, indicating that the advantage of the experimental group is unlikely to reflect sampling fluctuation. It is worth distinguishing this figure from the 10.59-point gap reported earlier: the latter compares the magnitude of improvement within each group, whereas the 10.36 points here compares the groups' final attainment directly.

Paired Samples *t*-tests confirmed that both groups improved significantly from pretest to posttest, the experimental group at t (29) = 18.47, $p < 0.001$ and the control group at $t(29)$ = 9.24, $p < 0.001$. That both gained is expected, since both received instruction; the decisive point is the disparity in magnitude, with the experimental *t*-value roughly double that of the control, which locates the surplus improvement in the treatment itself rather than in mere exposure to the material.

Effect Size and N-Gain Score

To gauge the practical weight of the difference, Cohen's *d* and the normalised gain were computed for both groups, as shown in Table 5.

Table 5. Effect Size (Cohen's d) and N-Gain Score Analysis Results

Groups	Cohen's d	Interpretation	N-Gain (g)	N-Gain Category
Experiments	1.50	Large Effect	0.65	Medium–High
Controls	—	—	0.36	Medium

Cohen's value d = 1.50 indicates a large effect (d > 0.80) of the Wordwall–PBL treatment on Arabic learning outcomes. The experimental group's N-Gain Score of 0.65 was in the upper-medium category (close to the high category: g ≥ 0.70), while the control group's was in the medium category (0.36). The difference in N-Gain of 0.29 points confirms that the improvement in learning in the experimental group is not solely due to time or maturation effects. However, it is a real contribution of the Wordwall PBL model.

Students' Perceptions of the Wordwall-PBL Media

After completing the entire learning series, the students in the experimental group completed a perception questionnaire via the Wordwall-PBL platform. Full results are presented in Table 6.

Table 6. Results of the Student Perception Questionnaire on Wordwall-PBL Media

Indicator	Percentage	Categories
Compatibility of the media with the material of *Durusul Lughah*	84.3%	Good
Attractiveness of presentation and animation	85.7%	Excellent
Ease of understanding through visual display	88.6%	Excellent
Audio and pronunciation clarity	90.0%	Excellent
Clarity of content and structure	91.4%	Excellent
Relevance of the image to the material	88.6%	Excellent
Ease of display language	85.7%	Excellent
Increase motivation to learn	82.9%	Good
Reduce learning boredom	75.7%	Good
Helps understand lessons easily	85.7%	Excellent
Overall average	86.9%	Excellent

The average perception among students (86.9%, very good category) indicates a high level of acceptance of the Wordwall-PBL model. The aspects that scored the highest were content and structure clarity (91.4%) and audio clarity (90.0%), while the boredom reduction aspect scored the lowest (75.7%) but remained in the good category. The consistency of this

score with the perception results from the initial development research at the *Darunnajah* Islamic Boarding School (86.9%) indicates that the quality of the developed PBL Wordwall media can be maintained in the dissemination context.

Teachers' Perceptions of the Wordwall-PBL Media

The facilitating teachers at the Rafah Islamic Boarding School evaluated the media on ten indicators, summarised in Table 7.

Table 7. Results of the Teacher's Perception Questionnaire on Wordwall-PBL Media

Indicator	Percentage	Categories
Compatibility of content with the material	90%	Excellent
Regularity of material presentation	80%	Good
Visual and animation quality	90%	Excellent
Audio clarity	90%	Excellent
Content appeal and interactivity	90%	Excellent
Language Facilitation	90%	Excellent
Image and text relevance	90%	Excellent
Impact on students' motivation to learn	90%	Excellent
Effectiveness in reducing learning burnout	90%	Excellent
Ease of understanding of the material	90%	Excellent
Overall average	89%	Excellent

The facilitator teacher at the Rafah Islamic Boarding School received a very good assessment, with an average of 89%, which is identical to the results of teachers' perceptions in the initial research at the *Darunnajah* Islamic Boarding School. In particular, the teacher highlighted the ability of the Wordwall result recapitulation feature to facilitate real-time digital formative assessment, enabling rapid identification of individual student weaknesses and the design of more targeted and personalized follow-up learning.

Discussion

The findings of this study consistently show that the Wordwall–PBL model is effective in improving Arabic learning outcomes among students at the Rafah Bogor Islamic Boarding School, with statistically significant effects ($p < 0.001$) and practical significance (Cohen's $d = 1.50$). These results strengthen and confirm the findings of initial development research at the Darunnajah Islamic Boarding School and align with empirical evidence from related studies. Alatas et al. (2024) found that Wordwall digital game media significantly improved students' vocabulary mastery, while Do and Huynh (2024) confirmed users' positive perceptions of Wordwall in vocabulary learning. The consistency of these findings across contexts reinforces the external validity of the Wordwall PBL model.

Cohen's $d = 1.50$, which far exceeds the large effect threshold ($d > 0.80$), indicates that the Wordwall PBL model has a tremendous impact in the context of *pesantren*, most likely due to several factors. First, the elements of competition and real-time leaderboards in Wordwall have strong cultural resonance in the *pesantren* community, where healthy competition (*munafasah*) is a value ingrained in students' learning culture. Second, the various game formats (matching, quiz, anagram, ordering) can accommodate students' diverse learning styles, enabling more students to be optimally engaged than the one-format method in conventional learning. Third, PBL integration creates a meaningful problem-solving context, so that the vocabulary and grammatical structures learned are not just memorized but actually understood and used in context (Othman & Shah, 2013).

From a dissemination perspective, these findings have important implications. The consistency of results between the *Darunnajah* Islamic Boarding School (development site) and the Rafah Islamic Boarding School (dissemination site) indicates that the PBL Wordwall model is highly portable and adaptable. The success of this dissemination is inseparable from several

critical factors: (1) the completeness of the implementation guidelines developed in BIMA's research so that facilitator teachers can operate the model without intensive training; (2) the flexibility of the Wordwall platform that does not require a high-speed internet connection so that it can be used in Islamic boarding schools with diverse technological infrastructure; and (3) the suitability of the content (*Durusul Lughah* Volume 1 material) which is indeed widely used in TMI Islamic boarding schools throughout Indonesia.

Interestingly, the perceptions of students and teachers at the Rafah Islamic Boarding School are almost identical to those at the Darunnajah Islamic Boarding School (students: 86.9% vs 86.9%; teachers: 89% vs 89%). This very high consistency indicates that the quality and acceptability of the Wordwall–PBL model do not depend on a particular institutional context, but rather are inherent characteristics of the model itself. This also indicates that aspects of *pesantren* values feared to clash with digital technology are not a real obstacle in the field, in line with Rosyidah's research (2024), which finds that digital innovations designed with *pesantren* values in mind are actually accepted in harmony.

The practical contribution of this study lies in demonstrating, with quantitative evidence, that the *Wordwall-PBL* model is genuinely transferable rather than bound to the conditions of its origin. Three concrete features account for this portability and give the finding actionable weight for the wider *pesantren* ecosystem. First, the implementation guidelines produced in the original development research were complete enough that the facilitating teachers at the Rafah Islamic Boarding School operated the eight learning activities without intensive retraining, which means an adopting institution need not depend on the original developers to reach fidelity. Second, the *Wordwall* platform functions without a high-speed internet connection, so the model fits *pesantren* whose technological infrastructure is uneven, removing the bandwidth barrier that commonly stalls digital adoption in such settings. Third, the model is built on *Durusul Lughah* Volume 1, the very material already taught across TMI *pesantren* throughout Indonesia, so adoption requires no curricular overhaul and the content aligns immediately with existing syllabi. Taken together with the automatic recapitulation feature that enables real-time formative assessment, these features establish the *Wordwall-PBL* model not as a promising prototype but as a ready-to-disseminate instrument that *pesantren* can adopt within their current resources.

Some limitations of this study need to be noted. First, although a quasi-experimental design is more feasible in the context of *pesantren*, the absence of full randomization introduces selection biases that cannot be fully controlled. Second, the study's 10-week duration was insufficient to measure long-term retention. Third, the dissemination is carried out only at one additional *pesantren*, so it is insufficient for broad generalization. Follow-up research with a true experimental design (if possible), a larger sample, a longer duration, and more diverse coverage of *pesantren* (*salaf pesantren*, *tahfiz pesantren*, *pesantren* perempuan) is highly recommended.

CONCLUSION

This study confirms that the *Wordwall-PBL* model can be disseminated effectively beyond its development site, carrying the *Durusullughah 4.0* innovation from the Darunnajah Islamic Boarding School into the distinct context of the Rafah Islamic Boarding School in Bogor without loss of effect. The experimental group reached a posttest mean of 81.23 against 70.87 for the control group, a difference of 10.36 points that proved statistically robust, t (58) = 5.812, $p < 0.001$, while a Cohen's d of 1.50 and a normalised gain of 0.65 against the control group's 0.36 confirm that the contribution was substantial rather than marginal. Acceptance reached 86.9 per cent among students and 89 per cent among teachers, figures identical to those at the development site, indicating that the model is both portable and culturally congruent, absorbed as an enrichment of *pesantren* pedagogy rather than resisted as an intrusion. Although limited by its quasi-experimental design and ten-week span, the findings support adopting the model as a formative assessment component within the Arabic curriculum and disseminating it more widely across *pesantren* through teacher training.

ACKNOWLEDGMENT

This research is part of the dissemination of research results funded by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (Kemdiktisaintek) through the Beginner Lecturer Research Scheme (PDP) in 2025. The research team would like to express its deepest gratitude to the Leadership of the Rafah Bogor Islamic Boarding School and all *ustadz/ustadzah* and students who have been willing to be partners in this dissemination research, to the Leadership of the *Darunnajah* Jakarta Islamic Boarding School for their permission and continuous support, and to all reviewers and review partners for constructive input in improving this article.

REFERENCES

Abdullah, A. F., Manaf, S., Lestari, T., & Novrizal, N. (2025). Integrating Wordwall and Problem-Based Learning in Arabic Language Instruction at Modern Islamic Boarding Schools. *Diwan: Journal of Arabic Language and Literature, 11*(2).

Agustini, K., Ngarti, J. G., & Sindu, I. G. P. (2021). Gamification in Education: A Systematic Literature Review. *Journal of Physics: Conference Series, 1810*(1), 012012. https://doi.org/10.1088/1742-6596/1810/1/012012

Alatas, P., Satria, I., & Serasi, R. (2024). The Effect of Word Wall Digital Game Media on Students' Vocabulary: A Quasi-Experimental Research at the Seventh Grade of SMP Islam Al-Azhar 52 Kota Bengkulu. *Journal of Basic Education (Pendas), 9*(2). https://doi.org/10.23969/jp.v9i2.18031

Alharthi, M. (2023). Technology Integration in Arabic Language Teaching: A Review of Digital Tools and Platforms. *Arab World English Journal, 14*(1), 88–104. https://doi.org/10.24093/awej/vol14no1.6

Barzanji, A. H., & Mohammed, A. H. (2022). The Effectiveness of Gamification in Enhancing Arabic Vocabulary Acquisition Among EFL Learners. *International Journal of English Language Teaching, 10*(3), 44–56. https://doi.org/10.37745/ijelt.2013/vol10n34456

Chandra, M. R., & Kun, Q. (2024). Utilizing Wordwall Mobile Apps to Improve Mandarin Language Skills. In *International Conference on Creative Communication and Innovative Technology (ICCIT)*. IEEE. https://doi.org/10.1109/ICCIT62134.2024.10701231

Colomo-Magaña, E., Soto-Varela, R., Ruiz-Palmero, J., & Gómez-García, M. (2020). University Students' Perception of the Usefulness of the Flipped Classroom Methodology. *Education Sciences, 10*(10), 275. https://doi.org/10.3390/educsci10100275

Denden, M., Tlili, A., Essalmi, F., & Jemni, M. (2021). Implicit Personality-Based Learner Modeling for Game-Based Learning Personalization. *IEEE Transactions on Learning Technologies, 14*(4), 474–487. https://doi.org/10.1109/TLT.2021.3105439

Do, T. C., & Huynh, N. T. (2024). Students' Perception of Using the Wordwall Application to Enhance Vocabulary Memorization. *Journal of Language and Linguistic Studies, 10*(1), 818.

Fauzan, A., & Arifin, F. (2023). The Effectiveness of Problem-Based Learning in Improving Students' Critical Thinking Skills in Islamic Education. *Al-Tadzkiyyah: Journal of Islamic Education, 14*(1), 1–16. https://doi.org/10.24042/atjpi.v14i1.15623

Hamari, J., Koivisto, J., & Sarsa, H. (2021). Does Gamification Work? A Literature Review of Empirical Studies on Gamification. *International Journal of Human-Computer Studies, 149*, 102582. https://doi.org/10.1016/j.ijhcs.2021.102582

Hidayat, R., & Syarifuddin, A. (2020). Digital Transformation in *Pesantren* Education in the Era of the Industrial Revolution 4.0. *Journal of Islamic Education, 9*(2), 145–160.

Kurniawan, D. T., Sujana, A., & Rachmatin, D. (2022). Effectiveness of the Hybrid Learning Model in Improving Student Learning Outcomes. *Journal of Physics: Conference Series, 2165*(1), 012030. https://doi.org/10.1088/1742-6596/2165/1/012030

Leis, A., Tohei, A., & Cooke, S. D. (2022). The Effects of Flipped Classrooms and Gamification on Student Motivation, Engagement, and Academic Performance. *TESOL Journal, 13*(1), e614. https://doi.org/10.1002/tesj.614

Lestari, P. D., & B. B. (2024). Problem-Based Learning With the iSpring-Assisted Inquiry Method for Critical Thinking Skills. *Journal of Education and Learning (EduLearn), 18*(1), 149. https://doi.org/10.11591/edulearn.v18i1.21089

Mardiyah, A. N., & Rachman, T. (2025). Development of Evaluation Media Based on Wordwall Games Integrated With Google Sites on SPLDV Material. *Journal of Mathematics and Science Education, 16*(1), 141–150. https://doi.org/10.26418/jpmipa.v16i1.75738

Maulina, Z. M., & Mawardi. (2022). Improving Students' Critical-Thinking Skills Through Wordwall-Assisted HyFlex Learning. *Scholaria: Journal of Education and Culture, 12*(3), 225–235. https://doi.org/10.24246/j.js.2022.v12.i3.p225-235

Muhson, A., Baroroh, K., Harini, & Lestari, A. (2023). Gamification in Learning: A Systematic Review of Its Effects on Student Motivation and Learning Outcomes. *Journal of Economic Education, 16*(1), 1–12. https://doi.org/10.21831/jpe.v16i1.54901

Othman, N., & Shah, M. I. A. (2013). Problem-Based Learning in the English Language Classroom. *English Language Teaching, 6*(3), 125–134. https://doi.org/10.5539/elt.v6n3p125

Putra, R. S., Wijayati, N., & Mahatmanti, F. W. (2020). The Effect of the Use of Android Application-Based Learning Media on Student Learning Outcomes. *Journal of Chemical Education Innovation, 14*(1), 2475–2485.

Rosyidah, I. M. (2024). Hybrid Innovation of School Curriculum and Islamic Boarding Schools in the Era of Information Technology. *Indonesian Journal of Education, 5*(9), 665–674. https://doi.org/10.59141/japendi.v5i9.3550

Subakti, H., Handayani, E. S., & Saputra, N. (2022). Problem-Based Learning Implementation in Enhancing Arabic Language Skills: A Meta-Analysis Study. *Al-Arabiyat: Journal of Arabic Language and Arabic Language Education, 9*(2), 179–196. https://doi.org/10.15408/a.v9i2.27832

Wahyudin, D., & Aristiawan. (2024). Digital Transformation in Islamic Boarding Schools (*Pesantren*): Challenges and Opportunities in the 4.0 Era. *Journal of Education and Learning, 18*(3), 712–722. https://doi.org/10.11591/edulearn.v18i3.21450

Zulkifli, A., & Lestari, R. (2024). Redefining Traditional Islamic Learning: *Pesantren* Education in the Digital Era. *Indonesian Journal of Islamic Education Studies, 6*(2), 89–104.

Zuo, M., Kong, S., Ma, Y., Hu, Y., & Xiao, M. (2023). The Effects of Using Scaffolding in Online Learning: A Meta-Analysis. *Education Sciences, 13*(7), 705. https://doi.org/10.3390/educsci13070705