
Innovating *maharah kalam* instruction through philosophical story-based role play: A quasi-experimental studyFaiqotul Millah^{1*}, Lailatul Mathoriyah²^{1,2} Universitas KH. A. Wahab Hasbullah, Jombang, Indonesia*Corresponding Author: faiqotulmillah1@gmail.com

Submission: 20 June 2026

Revision: 7 July 2026

Accepted: 14 July 2026

Abstract

Maharah kalam remains a major challenge in Arabic language learning because instructional practices are still dominated by dialogue memorization, while the integration of philosophical stories as the contextual basis for *role play* has received limited scholarly attention. This study aims to examine the effectiveness of a *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon* in improving students' *maharah kalam*. A quantitative approach employing a *quasi-experimental pretest-posttest control group design* was conducted with 65 students. Data were analyzed using descriptive statistics, normality and homogeneity tests, *Normalized Gain (N-Gain)*, an *independent-samples t-test*, Cohen's *d* effect size, and a student response questionnaire. The findings revealed that the experimental group achieved a higher posttest mean (90.92) than the control group (84.55), a moderate *N-Gain* (0.5606), a significant difference ($p < 0.001$), a very large effect size (3.48), and positive student responses (77.35%). This study contributes by proposing a conceptual model explaining how philosophical storytelling strengthens the effectiveness of *role play* in improving *maharah kalam*.

Keywords: *Maharah kalam*; *Role play*; Philosophical story; *Hayy bin Yaqdzon*; Arabic language learning.**Abstrak**

Kemampuan *maharah kalam* masih menjadi tantangan dalam pembelajaran bahasa Arab karena pendekatan yang digunakan cenderung berorientasi pada hafalan dialog, sementara kajian mengenai integrasi cerita filosofis sebagai konteks *role play* masih sangat terbatas. Penelitian ini bertujuan menganalisis efektivitas pendekatan *role play* berbasis cerita filosofis *Hayy bin Yaqdzon* dalam meningkatkan *maharah kalam* siswa. Penelitian menggunakan pendekatan kuantitatif dengan desain *quasi-experimental pretest-posttest control group design* yang melibatkan 65 siswa, dengan analisis meliputi statistik deskriptif, uji normalitas, uji homogenitas, *Normalized Gain (N-Gain)*, *independent sample t-test*, *effect size* (Cohen's *d*), dan angket respons siswa. Hasil penelitian menunjukkan bahwa kelas eksperimen memperoleh rata-rata *posttest* lebih tinggi (90,92) dibandingkan kelas kontrol (84,55), *N-Gain* kategori sedang (0,5606), perbedaan signifikan ($p < 0,001$), *effect size* sangat besar (3,48), serta respons siswa positif (77,35%). Penelitian ini berkontribusi dengan menawarkan model konseptual yang menjelaskan mekanisme integrasi cerita filosofis dalam memperkuat efektivitas *role play* untuk meningkatkan *maharah kalam*.

Kata Kunci: *Maharah kalam*; *Role play*; Cerita filosofis; *Hayy bin Yaqdzon*; Pembelajaran bahasa Arab.

Introduction

Maharah kalam is one of the core competencies in Arabic language learning, as it enables students to express ideas, articulate opinions, and engage in active communication (Buska et al., 2018). Nevertheless, the teaching of *maharah kalam* in various educational institutions continues to face fundamental challenges. Most instructional practices remain focused on vocabulary acquisition and grammatical structures, while opportunities for students to use Arabic in authentic communicative situations are still limited (Abidah et al., 2026). As a result, many students possess adequate theoretical knowledge but struggle to communicate spontaneously in real-life contexts. This condition suggests that improving speaking proficiency cannot be achieved solely through linguistic mastery; rather, it requires meaningful learning experiences that connect language use with authentic communicative contexts, thereby encouraging students to use Arabic actively across diverse learning situations.

These challenges indicate that the primary issue in teaching *maharah kalam* extends beyond students' limited speaking proficiency to the lack of instructional strategies capable of fostering authentic communication (Kim et al., 2022). In practice, classroom instruction is still dominated by lectures, scripted dialogue reading, and memorization exercises, leaving students with limited opportunities to explore ideas, develop speaking confidence, and interact using Arabic in meaningful ways (Almelhes, 2024). Consequently, the learning process tends to be passive and provides little exposure to genuine communicative experiences. Therefore, an instructional approach is needed that not only encourages students to speak but also promotes their cognitive, affective, and social engagement throughout the learning process. One promising approach is *role play*, as it creates contextualized communicative situations through structured role enactment (Sarifudin & Setyawan, 2025).

However, the effectiveness of *role play* depends not only on the enactment of roles itself but also on the quality of the context in which communication takes place (Othlinghaus-Wulhorst & Hoppe, 2020). Simple and mechanical dialogues often encourage students to memorize conversations without fully understanding their underlying meanings (Ericsson et al., 2023). To address this limitation, the present study integrates the philosophical story of *Hayy bin Yaqdzon* as the instructional context, providing students with opportunities to explore characters, engage in reflection, and construct meaning before participating in speaking activities. This classical philosophical narrative was selected because it embodies themes of knowledge seeking, character development, the relationship between humans and nature, and critical thinking, all of which are closely related to the development of communicative competence. Accordingly, *role play* functions not merely as a speaking exercise but as a learning process that connects language use with reflective thinking and meaningful knowledge construction.

Research on the implementation of *role play* in Arabic language learning has been widely conducted. A study by Suchada and Asama (2018) demonstrated that the *role-play* approach enhances students' speaking confidence, communication fluency, and active participation during classroom instruction. However, the study relied primarily on simple thematic dialogues, limiting students' opportunities to engage in deeper meaning construction. Another study by Naul and Liu (2020) found that story-based learning media enhanced students' motivation and engagement by presenting more compelling narrative experiences. Nevertheless, stories in that study functioned merely as reading materials or supporting media rather than serving as the primary foundation for *role-play* activities. These findings indicate an important research gap regarding the integration of philosophical stories into *role-play*-based instruction to create more reflective, contextualized, and meaningful communicative learning experiences.

Based on these considerations, this study aims to examine the effectiveness of a *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon* in improving the *maharah kalam* of Madrasah Aliyah students. Beyond measuring learning improvement through statistical analysis, this study also proposes a conceptual model explaining how philosophical storytelling establishes meaningful learning contexts, promotes active participation, strengthens communicative interaction, and ultimately enhances speaking proficiency. The significance of this research lies in providing an alternative approach to Arabic language instruction that extends beyond conventional dialogue practice by integrating philosophical, reflective, and communicative dimensions within a single learning process. Accordingly, the findings are expected to contribute both empirically and conceptually to the development of more meaningful, contextualized, and learner-centered strategies for teaching *maharah kalam*.

Method

This study employed a quantitative approach using a *quasi-experimental pretest-posttest control group design* to examine the effectiveness of a *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon* in improving students' *maharah kalam* (Rogers & Revesz, 2019). The research was conducted at Madrasah Aliyah Unggulan KH. Abd. Wahab Hasbulloh, Jombang, Indonesia, involving a total of 65 students, comprising 34 students in the experimental group and 31 students in the control group. The research site was selected because the school actively implemented Arabic language instruction but had not previously adopted a *role-play* approach based on philosophical storytelling as a strategy for teaching *maharah kalam*. The participants were selected through purposive sampling, as class assignments had been administratively determined at the beginning of the academic year, making random assignment infeasible (Etikan, 2016). Nevertheless, the two groups demonstrated comparable academic characteristics based on their pretest results, making them

suitable as the experimental and control groups within the *quasi-experimental* design. The experimental group received instruction through a *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon*, whereas the control group received conventional instruction.

Data were collected through oral tests and a student response questionnaire. The oral test served as the primary instrument for assessing *maharah kalam* during both the pretest and posttest using an expert-validated scoring rubric, ensuring consistent evaluation of pronunciation, fluency, comprehension, meaning construction, and creativity. Upon completing the instructional intervention, students in the experimental group completed a closed-ended questionnaire using a Likert scale to evaluate their responses to the implementation of the *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon*. The collected data were analyzed sequentially using descriptive statistics to summarize students' learning outcomes, followed by normality and homogeneity tests to verify the assumptions for parametric analysis. The effectiveness of the intervention was subsequently examined using the *Normalized Gain (N-Gain)*, an *independent-samples t-test*, and *effect size* (Cohen's *d*), while questionnaire data were analyzed using percentage statistics to describe students' acceptance of the implemented instructional approach (Creswell, 2014). This combination of analytical techniques was employed to comprehensively evaluate the effectiveness of the instructional approach from both the perspective of learning improvement and students' responses to the learning process.

Results and Discussion

Students' maharah kalam proficiency

An initial overview of students' *maharah kalam* development was obtained through descriptive statistical analysis. This analysis summarizes the characteristics of the data by presenting the minimum and maximum scores, means, and standard deviations, thereby providing an objective description of students' speaking proficiency before and after the instructional intervention. In addition to illustrating the overall learning performance of each group, descriptive statistics also provide information on data distribution as a preliminary basis before conducting inferential statistical analyses. Accordingly, the findings are presented sequentially, beginning with the descriptive statistics of the control group, followed by those of the experimental group and a comparison of the *Normalized Gain (N-Gain)* scores. The descriptive statistical summary for the control group is presented in Table 1.

Table 1. Descriptive statistics of the control group

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Pretest	31	76.00	82.80	79.4194	1.76303
Posttest	31	80.40	88.20	84.5484	1.89699

Posttest – Pretest	31	2.00	8.90	5.1290	1.62833
Ideal Score – Pretest	31	17.20	24.00	20.5806	1.76303
N-Gain Score	31	0.10	0.39	0.2485	0.07393
N-Gain (%)	31	10.36	38.86	24.8473	7.39292
Valid N (listwise)	31				

Source: SPSS 26 output, processed by the authors (2026).

As shown in Table 1, the control group achieved a mean pretest score of 79.42, which increased to 84.55 after the instructional process. The improvement of 5.13 points indicates a positive development in students' speaking proficiency, although the magnitude of the improvement remained relatively limited. This finding is further reflected in the *N-Gain* score of 0.2485 (24.85%), which falls within the low category. The ranges of both the pretest and posttest scores shifted toward higher values, while the standard deviations remained relatively small. These results suggest that learning improvement occurred fairly consistently across most students; however, the effectiveness of the conventional instructional approach was still insufficient to produce substantial gains in *maharah kalam* proficiency.

A different pattern emerged in the experimental group following the implementation of the *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon*. Although students in this group demonstrated an initial level of proficiency comparable to that of the control group, the improvement observed after the intervention was considerably greater. The mean score increased substantially, accompanied by a noticeable upward shift in the range of posttest scores. The descriptive statistical summary for the experimental group is presented in Table 2.

Table 2. Descriptive statistics of the experimental group

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Pretest	34	77.00	83.00	79.3900	1.53100
Posttest	34	87.00	94.00	90.9200	1.76600
Posttest – Pretest	34	8.50	14.80	11.5353	1.57536
Ideal Score – Pretest	34	17.20	23.00	20.6147	1.53132
N-Gain Score	34	0.44	0.71	0.5606	0.07268
N-Gain (%)	34	43.91	71.43	56.0585	7.26777
Valid N (listwise)	34				

Source: SPSS 26 output, processed by the authors (2026).

As presented in Table 2, the experimental group obtained a mean pretest score of 79.39, which was nearly identical to that of the control group. Following instruction using the *role-play* approach, the mean posttest score increased to 90.92, representing an

average improvement of approximately 11.54 points. This substantial increase indicates greater learning gains than those observed in the control group. Furthermore, the posttest scores ranged from 87 to 94, while the standard deviation remained relatively low, suggesting that the improvement was not limited to a small number of students but was distributed consistently across the group. These descriptive findings indicate that the implementation of the *role-play* approach was associated with a more pronounced improvement in students' learning outcomes.

The difference in learning improvement between the two groups becomes more apparent when examined using the *Normalized Gain (N-Gain)* score. This analysis measures the magnitude of learning improvement while accounting for students' initial performance, thereby providing a more proportional evaluation of instructional effectiveness. A comparison of the *N-Gain* results for the control and experimental groups is presented in Table 3.

Table 3. Comparison of *N-Gain* scores

Group	Mean <i>N-Gain</i>	<i>N-Gain</i> (%)	Category
Control Group	0.2485	24.85	Low
Experimental Group	0.5606	56.06	Moderate

Source: SPSS 26 output, processed by the authors (2026).

The comparison presented in Table 3 shows that the control group achieved a mean *N-Gain* score of only 0.2485, whereas the experimental group obtained a considerably higher mean score of 0.5606. This difference indicates that the improvement in *maharah kalam* proficiency was substantially greater in the experimental group than in the control group. From a descriptive perspective, these findings suggest that the *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon* was more effective in enhancing students' speaking proficiency than conventional instruction. Nevertheless, these findings remain descriptive in nature; therefore, inferential statistical analyses are required to determine whether the observed differences are statistically significant.

Effectiveness of the philosophical story-based role-play approach

Following the descriptive analysis of students' *maharah kalam* proficiency, the next stage involved conducting statistical tests to verify that the data satisfied the assumptions required for parametric analysis before hypothesis testing. The analysis began with a normality test to determine whether the data distributions in each group followed a normal distribution. Establishing normality is an essential prerequisite for applying an *independent-samples t-test*, ensuring that the resulting statistical inferences are valid and reliable. Accordingly, normality was examined for both the pretest and posttest scores of the control and experimental groups using the Kolmogorov-Smirnov and Shapiro-Wilk tests. The results of the normality test are presented in Table 4.

Table 4. Test of normality

Group		Kolmogorov-Smirnov Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Control Group Pretest		0.145	31	0.095	0.967	31	0.450
Control Group Posttest		0.114	31	0.200*	0.986	31	0.946
Experimental	Group Pretest	0.141	34	0.086	0.956	34	0.180
Experimental	Group Posttest	0.130	34	0.155	0.976	34	0.632

Source: SPSS 26 output, processed by the authors (2026).

As presented in Table 4, all significance values exceeded the threshold of 0.05 in both the Kolmogorov-Smirnov and Shapiro-Wilk tests. The significance values were 0.450 for the control group pretest, 0.946 for the control group posttest, 0.180 for the experimental group pretest, and 0.632 for the experimental group posttest. These results indicate that the data in each group satisfied the assumption of normality, suggesting the absence of distributional violations that could affect the validity of subsequent parametric analyses. Having established normality, the next step was to examine the homogeneity of variance between the two groups using Levene's test. The results are presented in Table 5.

Table 5. Test of homogeneity of variance

Basis of Test	Levene Statistic	df1	df2	Sig.
Based on Mean	0.001	1	63	0.979
Based on Median	0.000	1	63	0.994
Based on Median and with Adjusted df	0.000	1	60.017	0.994
Based on Trimmed Mean	0.001	1	63	0.981

Source: SPSS 26 output, processed by the authors (2026).

As shown in Table 5, all significance values obtained from Levene's test exceeded 0.05, with a significance value of 0.979 based on the mean. These findings indicate that the variances of the control and experimental groups were homogeneous, suggesting that both groups exhibited comparable levels of variability. Satisfying the homogeneity assumption further confirmed the appropriateness of using the *independent-samples t-test* to examine the research hypothesis. In other words, the data fulfilled the two principal assumptions required for parametric analysis, namely normality and homogeneity of variance. After confirming these assumptions, hypothesis testing was conducted to determine whether significant differences existed between the learning outcomes of the two groups. The results are summarized in Table 6.

Table 6. Independent-samples *t*-test

Statistic	Value
Levene's <i>F</i>	0.001
Levene's Sig.	0.979
<i>t</i>	-14.024
df	63
Sig. (2-tailed)	0.000
Mean Difference	-6.372
Std. Error Difference	0.454
95% CI Lower	-7.280
95% CI Upper	-5.464

Source: SPSS 26 output, processed by the authors (2026).

The results presented in Table 6 indicate a significance value of 0.000, which is below the significance level of 0.05. This finding indicates that the null hypothesis was rejected, demonstrating a statistically significant difference in learning outcomes between the control and experimental groups. The mean difference of 6.372 points indicates that students who received instruction through the *role-play* approach achieved higher *maharah kalam* proficiency than those who experienced conventional instruction. Beyond statistical significance, the magnitude of the intervention's impact was further examined using *effect size*, which provides information regarding the practical strength of the instructional treatment. The results of the *effect size* analysis are presented in Table 7.

Table 7. Effect size (Cohen's *d*)

Statistic	Value
Mean Posttest of the Control Group	84.5484
SD of the Control Group	1.89699
Mean Posttest of the Experimental Group	90.9200
SD of the Experimental Group	1.76600
Pooled Standard Deviation	1.831
Cohen's <i>d</i>	3.48
Interpretation	Very Large Effect

Source: Calculated from the SPSS output, processed by the authors (2026).

The Cohen's *d* value of 3.48 presented in Table 7 indicates a *very large effect*, demonstrating that the instructional intervention had a substantial practical impact on improving students' *maharah kalam* proficiency. These findings suggest that the observed difference was not only statistically significant but also educationally

meaningful. Therefore, the *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon* exerted a considerably stronger influence on students' speaking proficiency than conventional instruction. In addition to learning outcomes, instructional effectiveness was also evaluated through students' responses to the learning experience. A summary of the questionnaire results is presented in Table 8.

Table 8. Student responses to the instructional approach

Component	Value
Total Score Obtained	1,315
Maximum Score	1,700
Percentage	77.35%
Category	Good

Source: Questionnaire data processed by the authors (2026).

As presented in Table 8, students' responses reached 77.35%, which falls within the good category. This finding indicates that most students responded positively to the implementation of the *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon*. The high level of positive responses suggests that the instructional approach not only improved learning outcomes but was also well received by students throughout the learning process. Overall, the statistical analyses demonstrated that the data satisfied all assumptions required for parametric testing, revealed significant differences in learning outcomes between the two groups, showed a *very large* treatment effect, and confirmed positive student perceptions of the instructional approach. These findings provide a robust empirical foundation for the subsequent discussion on the effectiveness of the *role-play* approach in enhancing students' *maharah kalam* proficiency.

Effectiveness of the role-play approach based on the philosophical story of Hayy bin Yaqdzon in improving maharah kalam

The improvement of speaking proficiency in Arabic language learning has traditionally been explained as a direct consequence of applying particular instructional methods (Fadhlan et al., 2024). While this perspective adequately accounts for changes in learning outcomes, it does not fully explain the mechanisms through which such improvements occur. In other words, the success of an instructional method depends not only on the learning activities it involves but also on the context that shapes students' learning experiences. The findings of this study revealed that the control and experimental groups began with comparable levels of proficiency, with mean pretest scores of 79.42 and 79.39, respectively, which subsequently increased to 84.55 and 90.92 following the intervention. These improvements were accompanied by *N-Gain* scores of 0.2485 and 0.5606, a significance value of 0.000, a Cohen's *d* of 3.48, and a positive student response rate of 77.35%.

Collectively, these findings indicate that improvements in *maharah kalam* cannot be explained solely by changes in test scores but should instead be understood through the interaction among multiple instructional components. Based on these findings, the mechanism underlying instructional effectiveness was synthesized into the conceptual model presented in Figure 1.

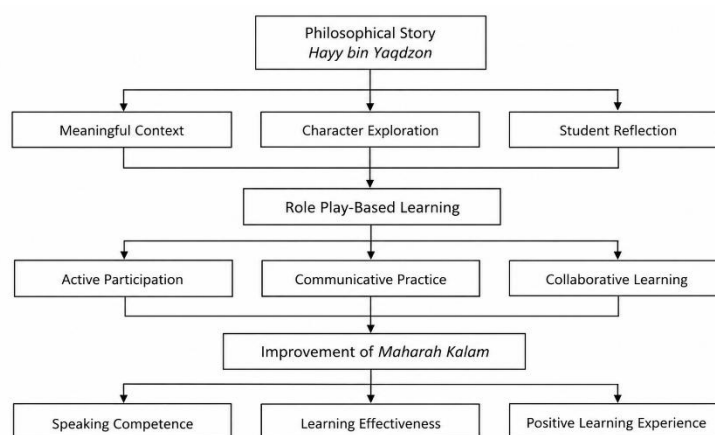


Figure 1. Conceptual findings of philosophical story-based *role play* in improving *maharah kalam*

Source: Developed by the authors based on the research findings (2026).

Figure 1 presents a conceptual synthesis derived from the overall findings of the study, with each component representing relationships identified throughout the instructional process. Rather than merely demonstrating that the *role-play* approach enhances *maharah kalam*, the model illustrates how a philosophical story establishes a meaningful learning context, how that context shapes instructional activities, and how these activities ultimately lead to improved speaking proficiency. Accordingly, the primary function of this conceptual model is to connect empirical evidence with the instructional mechanisms underlying the observed learning outcomes. This perspective is consistent with the findings of Johnston et al. (2018), who argued that conceptual models are essential for explaining relationships among variables that cannot be fully captured through statistical analyses alone. Similarly, Burns et al. (2020) emphasized that conceptual visualization provides a more comprehensive explanation of how educational phenomena develop than numerical data alone.

The philosophical story of *Hayy bin Yaqdzon* serves as the instructional foundation by fostering *meaningful context*, *character exploration*, and *student reflection*. These three components indicate that the story functions not merely as supplementary learning material but as an instructional instrument that prepares students intellectually before engaging in communicative activities. A meaningful context enables students to develop a deeper understanding of dialogue, character exploration encourages them to internalize the roles they perform, and reflection helps them connect the narrative with their own intellectual experiences. This mechanism explains why two groups with nearly identical initial proficiency ultimately demonstrated substantially

different learning outcomes following the instructional intervention. These findings are consistent with Hiver et al. (2020), who reported that meaning-rich narratives enhance students' cognitive engagement during language learning. Likewise, Yasin et al. (2025) found that character exploration strengthens learners' comprehension of dialogue. Unlike these previous studies, however, the present research demonstrates that philosophical storytelling not only improves dialogue comprehension but also cultivates intellectual reflection that serves as a foundation for meaningful communication.

The philosophical context subsequently evolves into a *role-play*-based instructional process characterized by *active participation*, *communicative practice*, and *collaborative learning*. These three components interact synergistically, enabling students not only to practice the language but also to engage in more authentic communicative interactions throughout the learning process. Active participation provides greater opportunities for language production, communicative practice situates language use within meaningful contexts, and collaborative learning facilitates the negotiation of meaning through peer interaction. The effectiveness of this mechanism is reflected in the experimental group's *N-Gain* score of 0.5606, substantially higher than the control group's score of 0.2485, and is further supported by the statistically significant result of the *independent-samples t-test* ($p = 0.000$). These findings are consistent with Mroz (2018), who identified authentic communication as the central element of speaking instruction. Similarly, Z. Zhang et al. (2026) demonstrated that collaborative learning increases students' confidence in using the target language. The present study extends these findings by showing that authentic communication becomes even more effective when it is grounded in a philosophical context that students have already understood and internalized.

The instructional process ultimately culminates in improved *maharah kalam*, as reflected in enhanced speaking competence, greater instructional effectiveness, and more positive learning experiences. These outcomes are not isolated achievements but rather the consequence of the interaction among the instructional components established from the beginning of the learning process. This conclusion is further supported by the study's *effect size* (Cohen's $d = 3.48$), which falls within the *very large effect* category, indicating that the intervention produced not only statistically significant differences but also substantial practical educational benefits. Moreover, the positive student response rate of 77.35% demonstrates that improved learning outcomes were accompanied by favorable perceptions of the instructional experience. These findings are consistent with Saito et al. (2018), who reported that positive learning experiences contribute to improved oral communication performance. Likewise, Q. Zhang (2025) concluded that active student engagement enhances the effectiveness of language instruction. Unlike these previous studies, however, the present research demonstrates that positive learning experiences emerged through the

integration of philosophical storytelling and *role-play* activities rather than merely from the adoption of a more active instructional method.

Overall, the findings indicate that the effectiveness of the *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon* does not result from a simple cause-and-effect relationship but from a mutually reinforcing interaction among philosophical context, instructional processes, and student engagement. This interaction explains why the experimental group achieved greater learning gains, higher *N-Gain* scores, statistically significant differences, a very large *effect size*, and more positive student responses. Consequently, the primary contribution of this study lies not only in demonstrating the effectiveness of the *role-play* approach in improving *maharah kalam* but also in proposing a conceptual model that explains the mechanisms underlying this effectiveness. The model suggests that the quality of the instructional context is just as important as the instructional method itself. Therefore, integrating philosophical storytelling into *role-play*-based instruction represents a promising pedagogical strategy for developing Arabic language learning that is more communicative, reflective, and meaningful.

Conclusion

This study demonstrates that the *role-play* approach based on the philosophical story of *Hayy bin Yaqdzon* is effective in improving students' *maharah kalam*. Its effectiveness is evidenced not only by the statistically significant improvement in learning outcomes but also by the instructional mechanism established throughout the learning process. Although the control and experimental groups began with comparable levels of proficiency, with mean pretest scores of 79.42 and 79.39, respectively, the experimental group achieved a substantially higher mean posttest score (90.92) than the control group (84.55). This improvement was further supported by a moderate *N-Gain* score (0.5606), a statistically significant difference ($p < 0.001$), a *very large effect size* (Cohen's $d = 3.48$), and positive student responses (77.35%). These findings indicate that improvements in *maharah kalam* were influenced not only by the implementation of the *role-play* method itself but also by the integration of a philosophical story that established a meaningful learning context, promoted character exploration and student reflection, and fostered more active and collaborative communicative interactions throughout the instructional process.

This study contributes conceptually by proposing an instructional model that explains how philosophical storytelling enhances the effectiveness of *role play* in improving *maharah kalam*. The proposed model demonstrates that the quality of the instructional context is as important as the instructional method itself in developing students' speaking competence. Nevertheless, this study has several limitations, including the use of a *quasi-experimental* design conducted in a single school with a relatively small sample size and the use of only one philosophical story as the instructional medium.

Therefore, future research is encouraged to examine the proposed model across different educational levels and learner characteristics, explore a wider range of philosophical stories or literary works as the basis for *role-play* activities, and integrate quantitative and qualitative approaches to provide a more comprehensive understanding of the mechanisms underlying the development of speaking proficiency.

References

- Abidah, A. L., Setiawan, A., S., M. N. A. Q. F., & Putri, N. A. A. (2026). Teaching *maharah al-kalam* through digital media in higher education. *Arabiyatuna: Jurnal Bahasa Arab*, 10(1), 187–206. <https://doi.org/10.29240/jba.v10i1.13553>
- Almelhes, S. (2024). Enhancing Arabic language acquisition: Effective strategies for addressing non-native learners' challenges. *Education Sciences*, 14(10), 1116. <https://doi.org/10.3390/educsci14101116>
- Burns, A., Xiong, C., Franconeri, S., Cairo, A., & Mahyar, N. (2020). How to evaluate data visualizations across different levels of understanding. *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, 19–28. <https://doi.org/10.1109/BELIV51497.2020.00010>
- Buska, W., Prihartini, Y., & Hasnah, N. (2018). Analysis of students' Arabic proficiency for vocabulary mastery in State Islamic Junior High School in Muaro Jambi. *INNOVATIO: Journal for Religious Innovation Studies*, 18(1), 51–62. <https://doi.org/10.30631/innovatio.v18i1.40>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.
- Ericsson, E., Sofkova Hashemi, S., & Lundin, J. (2023). Fun and frustrating: Students' perspectives on practising speaking English with virtual humans. *Cogent Education*, 10(1). <https://doi.org/10.1080/2331186X.2023.2170088>
- Etikan, I. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fadhlan, M., Asrori, I., Sutaman, & Setiyadi, A. C. (2024). The improvement of students' Arabic speaking skills: The role of interactive strategies and the establishment of a language environment. *Izdiyar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 7(2). <https://doi.org/10.22219/jiz.v7i2.30137>
- Hiver, P., Zhou, S., Tahmouresi, S., Sang, Y., & Papi, M. (2020). Why stories matter: Exploring learner engagement and metacognition through narratives of the L2 learning experience. *System*, 91, 102260. <https://doi.org/10.1016/j.system.2020.102260>

- Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, 52(4), 1957–1976. <https://doi.org/10.1007/s11135-017-0584-6>
- Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies*, 27(5), 6069–6104. <https://doi.org/10.1007/s10639-021-10831-6>
- Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, 51(3), 617–637. <https://doi.org/10.1111/flan.12348>
- Naul, E., & Liu, M. (2020). Why story matters: A review of narrative in serious games. *Journal of Educational Computing Research*, 58(3), 687–707. <https://doi.org/10.1177/0735633119859904>
- Othlinghaus-Wulhorst, J., & Hoppe, H. U. (2020). A technical and conceptual framework for serious role-playing games in the area of social skill training. *Frontiers in Computer Science*, 2. <https://doi.org/10.3389/fcomp.2020.00028>
- Rogers, J., & Revesz, A. (2019). Experimental and quasi-experimental designs. In *The Routledge handbook of research methods in applied linguistics* (pp. 133–143). Routledge.
- Saito, K., Dewaele, J., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, 68(3), 709–743. <https://doi.org/10.1111/lang.12297>
- Sarifudin, S., & Setyawan, W. H. (2025). The effect of using role-play method in improving speaking skills of basic level students at central course. *Prima Magistra: Jurnal Ilmiah Kependidikan*, 6(1), 29–35. <https://doi.org/10.37478/jpm.v6i1.5190>
- Suchada, T., & Asama, T. (2018). The effects of role-playing simulation activities on the improvement of EFL students' business English oral communication. *The Journal of AsiaTEFL*, 15(3), 735–749. <https://doi.org/10.18823/asiatefl.2018.15.3.11.735>
- Yasin, M. I., Fikri, M. H., Sudrajad, W., & Dewi, M. (2025). Understanding character dialogue in modern English literature: A qualitative study on students' cognitive processes in literary learning. *Journal of Development Research*, 9(1), 92–100. <https://doi.org/10.28926/jdr.v9i1.453>
- Zhang, Q. (2025). Active language learning strategies in Chinese university English education: Effects on proficiency, motivation, and teacher engagement. *Acta Psychologica*, 261, 105914. <https://doi.org/10.1016/j.actpsy.2025.105914>

Zhang, Z., Liu, T., Gao, X., & Lee, C. B. (2026). A longitudinal examination of language learners' group-level enjoyment and emotion regulation in online collaborative learning. *Language Teaching Research*, 30(6), 3111–3133.
<https://doi.org/10.1177/13621688241227584>