al mubin
ISLAMIC SCIENTIFIC JOURNAL

P-ISSN: 2746-8240 E-ISSN: 2746-5535
DOI: 10.51192/almubin.v7i1.753

ENHANCING WRITING COMPREHENSION IN L2 ARABIC LEARNERS THROUGH AI-BASED TRANSLANGUAGING CHATBOTS

Nely Rahmawati Zaimah[1], Fatchiatuzahro[2], Eko Budi Hartanto[3]
[1]Al-Anwar Islamic Institute Rembang, Indonesia
[2]Ummul Quro Al-Islami Institute Bogor, Indonesia
[3]State Islamic Institute of Kediri, Indonesia
E-mail: nelyrahmawati@staialanwar.ac.id

ABSTRACT

The ability to speak a foreign language today is a skill that should be possessed by everyone, especially students, among whom are Arabic language skills. Nowadays, there are a lot of media and learning facilities that can be used to improve language skills. One of them is using AI chatbots, so this research is done to find out how effective these artificial intelligence chatbots are in improving the Arabic writing skills of L2 students. Second language acquisition poses a significant challenge when utilizing artificial intelligence chatbots for learning. Proficiency limitations in the second language can impede effective communication with chatbots. However, this challenge can be addressed through the practice of translanguaging in chatbot interactions. This study adopts a quantitative approach using pre-experimental methods to assess the efficacy of an artificial intelligence chatbot for enhancing Arabic writing comprehension within a translanguaging framework. The primary objective is to improve writing comprehension among Arabic learners as a second language (L2). The research involves 45 participants from Arabic language classes within the Department of Quranic Studies and Exegesis (IQT) at STAI Al-Anwar Sarang, Rembang. Statistical analyses and Bayesian inference are performed using JASP 0.18.1.0 software. Both classical and Bayesian analyses are employed to validate test results, augment probability and sustainability, while maintaining a focused analysis of the impact of chatbot-assisted learning within the translation framework. The results indicate a significant positive impact of utilizing AI chatbot-based Arabic writing comprehension among L2 learners. The researchers foresee the necessity for further exploration in the realms of translanguaging frameworks and their application in AI-assisted language learning.
***Keywords**:, Arabic Writing, Artificial Intelligence Chatbots, , Bayesian inferences, Second Language Acquisition*

INTRODUCTION

A foreign language learner is required to possess the ability to write, in addition to their proficiency in speaking and listening. The improvement of this writing skill can be achieved through consistent practice utilizing an assortment of educational resources. As a result of the advancement of media technology, artificial intelligence-based technologies can now be utilized in education. One such example is the AI chatbot.

Numerous researchers continue to engage in extensive discourse regarding the introduction of artificial intelligence (AI) models utilizing Natural Language Processing (NLP) and Large Language Models (LLMs), such as Chatbot Bing and ChatGPT, into the educational sphere (Kolhar & Alameen, 2021; Putri Supriadi et al., 2022; Renz & Vladova, 2021). Abdulkader and Al-Irhayim, (2022) have emphasized the unique characteristics of the Arabic language within the domain of Arabic education. This includes its wide range of variants, which consist of Classical Arabic (CAL), Modern Standard Arabic (MSA), and several Arabic dialects (DA), each serving a distinct linguistic purpose. By utilizing natural language processing (NLP) methods to improve writing comprehension, they contend that AI-powered chatbots can efficiently facilitate language learning. Additionally, Fuad and Al-Yahya

(2022) investigated the potential of AI chatbots for conversational Arabic learning, highlighting their significance in language education. Aligned with this, the prominence of the trend toward online and technology-enhanced second language instruction has become evident in the development of writing comprehension skills (Rumaisa et al., 2021; Saputri, 2021; Setiadi, 2021). In the context of Arabic language education, it is crucial to assess the impact and efficacy of LLM-based chatbots on the improvement of writing comprehension in Arabic as a second language.

Conversely, certain scholars have expressed concerns and particular difficulties pertaining to AI chatbot models, which continue to generate disagreement (Stahl, 2021; Veale & Zuiderveen Borgesius, 2021). Important scholars, including Perkins (2023) and Chan (2023), have emphasized that students' utilization of LLMs not only gives rise to apprehensions regarding academic integrity, but also calls into question their capacity to integrate this technology into their academic studies. Concerns are thus urged regarding the efficacy and consequences of utilizing such technology in the realm of language acquisition, particularly with regard to writing comprehension.

There has been an increase of concerns in numerous disciplines (Ashoori & Weisz, 2019; Brundage et al., 2020; Hatherley, 2020). Chan, (2023) and Miao et al. (2021) have thus highlighted the importance of quickly formulating policies to tackle potential obstacles arising from artificial intelligence-generated content. The detection of text produced by artificial intelligence continues to be a substantial obstacle, given that large language models are capable of generating enormous amounts of text that can be exploited to spread deceptive or misleading data at an unprecedented rate (Michel-Villarreal et al., 2023). This underscores the necessity for meticulous examination of their application, pertinence, and linguistic precision within the realm of language instruction (Al-Abdullatif & Alsubaie, 2022; Shao et al., 2022).

Second-language learning is often a fundamental challenge when using AI chatbots for learning. Limited proficiency in the second language can hinder effective communication with chatbots. However, this challenge can be mitigated by practicing translanguaging in chatbot communication. Translanguaging allows students to utilize their entire linguistic repertoire, enabling them to communicate more effectively even when their proficiency in the second language is limited (Canals, 2022; DeCapua, 2022; Goodman & Tastanbek, 2021). With translanguaging, learners are allowed to switch between their first language (Indonesian) and the second language (Arabic) when interacting with chatbots (Kolhar & Alameen, 2021). This not only helps them communicate more effectively but also provides them with opportunities to practice and improve their second language skills (Elashhab, 2020). However, there is still limited research on these interactions, particularly their acceptability, effectiveness, and accuracy in empirical contexts.

Therefore, it is crucial to conduct a thorough investigation in order to develop effective tools and strategies to address the complexities presented by these challenges. Two fundamental research inquiries arise: In the domain of AI utilization, how do these intertwined and occasionally contradicting narratives contribute to the larger discourse in technology-driven language education? Moreover, to what level do they impact a variety of factors, including the development of motivation, the acquisition of vocabulary, the ability to speak and write fluently in the target language, and proficiency in reading and writing? Writing, nevertheless, is not comparable to other linguistic abilities (Iftanti, 2016). These inquiries will facilitate a more comprehensive comprehension of the intricate terrain of artificial intelligence in language education.

Drawing from the aforementioned factors, the objective of this study is to investigate the determinants that affect the efficacy and influence of AI chatbot models when it comes to augmenting writing comprehension in the context of Arabic as a second language. These factors comprise the accessibility and acceptance of chatbots, their efficacy in enhancing comprehension, their capacity to construct literal communication paradigms, and their linguistic writing capabilities. The researchers intend to establish a comprehensive framework for assessing the effectiveness of LLM-based chatbot AI in facilitating writing comprehension for Arabic language learners through the formulation of these parameters. The principal objective of this study is to assess the efficacy of LLM-based Chatbot AI in relation to Arabic language learners' writing abilities and to determine the evaluative criteria (concept formulation, paradigm organization, and issues/challenges in writing development) that establish Chatbot AI as a valuable instrument for comprehending and enhancing second-language Arabic writing abilities. This would enable learners to carefully develop their Arabic writing paradigms.

In conclusion, the researchers employ AI solely as an additional instrument in the development of writing frameworks. The principal aim is to reduce moral issues and the risks of plagiarism while capitalizing on the capabilities of artificial intelligence in producing genuine and innovative texts and narratives. This strategy emphasizes the responsible and ethical application of artificial intelligence in the research process.

METHODS

1. *Type and Segmentation:*

This study employs a quantitative approach to evaluate the effectiveness of using chatbot translanguaging in enhancing writing comprehension among learners of Arabic as a second language (L2), using the experimental method of pre-experimental one-group pretest-posttest design. As for the population that is the subject of the study, that is, all the students, From Arabic language classes within the Department of Quranic Studies and Exegesis at STAI Al-Anwar Sarang, which are 45 students, the selection of participants is based on their average proficiency level, which is gathered from the Arabic language classes in the IQT (Quranic Studies and Exegesis) department at STAI Al-Anwar Sarang Rembang. The reason for this selection is that the department has graduates from Islamic boarding schools with a more solid language background. This is crucial in establishing initial hypotheses and prior distributions. Their strong language background can facilitate this process, allowing us to focus on the target of improving writing comprehension in the Arabic language (as an alternative hypothesis and posterior distribution). This research was conducted over one academic year (September 2022 to May 2023), involving thematic parameters (determining writing themes), schematic parameters (forming writing schemas or paradigms), and investigative parameters (in developing the writing schema).

2. *Data Collection*

Data collection involves pre-tests and post-tests per category as measurement variables that do not determine the final scores. The determination of the final score is the research paper itself. The pre-test was conducted at the beginning of the first semester, and the post-test was conducted at the end of the semester, involving structured and well-bound research papers. The pre-test and post-test were designed to assess the students' writing comprehension before and after the application of chatbot translanguaging.

3. *Data Analysis*

Data from the pre-test and post-test were then analyzed using Bayesian paired t-tests to determine if there was a significant improvement in writing comprehension after the implementation of chatbot translanguaging. This Bayesian paired t-test was conducted using the JASP 0.18.0.0 software, known for its ability to perform Bayesian modeling effectively (Van Doorn et al., 2020; Maier et al., 2022).

4. *Research Design*

This research design, marked by manipulated conditions, random assignment, and a between-subjects approach, allows a thorough exploration of how chatbot translanguaging influences Arabic writing understanding in second language learners. It ensures a meticulous investigation into the precise impact of this innovative approach on language acquisition.

RESULT AND DISCUSSION

Initial observations regarding the writing proficiency levels of the participants, or students in this instance, were inconsistent; however, their language comprehension with regard to material absorption was, on average, average. Writing offered particular and distinctive difficulties. Although they possessed proficient reading skills, composing was an entirely distinct challenge for them. They frequently encountered difficulties in determining what to write about, even during the early stages of generating titles or themes. Nevertheless, this difficulty might be resolved through the implementation of translation in chatbot discourse. Translation facilitates the integration of students' complete linguistic repertoire into a bilingual environment, thereby enhancing their communication capabilities, particularly in the development of literacy structures, even in cases where their second language proficiency is limited.

The Chatbot AI being referred to is Bing mobile (operating on chat GPT 4). Its sole purpose is to aid in the translanguaging process and improve comprehension; it is not intended to detect plagiarism. The written material includes, but is not limited to, light articles, opinions, book reviews, book critiques, and novels. As research interventions, both chatbot and translanguaging practices were implemented beginning in the same semester. The primary value is determined by the aforementioned context and parameters and not by the quantity of words they are capable of composing. It is crucial to acknowledge that the paper grades do not serve as a comprehensive indicator of the students' Arabic language proficiency. For the academic year 2022-2023, the paper mark is specific and is one component of the Arabic language material grade.

During the odd semester, students were assigned to compose an Arabic-language paper on any subject of their choosing, without receiving any guidance or information from the AI chatbot. As explicated in the methodology section, this research incorporates three primary parameters to enhance proficiency in Arabic writing. The purpose of the thematic parameter is to determine the subject or theme of the writing. The students were given thematic guidelines in order to guarantee that their written work is pertinent and in accordance with the educational goals. The schematic parameter facilitates the development of a writing scheme or paradigm by students. This resource assists learners in efficiently organizing their writing and structuring their thoughts. The investigative parameter promotes student engagement by requiring them to delve into multiple facets of writing while expanding upon the structure of their selected theme. This includes rhetorical considerations, writing style, and tone. All of this is succinctly summarized and enclosed in the paper. Certain individuals encountered ongoing challenges in methodically organizing their papers, whereas others resorted to plagiarism by incorporating material from online articles or narratives.

Among the entire sample of 45 participants, three attained ratings exceeding 8.0 across all three assessment parameters. A total of sixteen participants achieved scores between seven and eight, with sixteen more achieving scores between six and seven. The final ten individuals were assigned the lowest rankings, attaining scores ranging from 5 to 6. Without exception, every participant submitted a paper in Arabic; those who failed to do so were omitted from the participant roster. Due to illness, one participant was unable to take the examination and submitted late. In stark contrast to the overall average score of 58.7 from the odd semester, the overall average score of 77 substantially surpassed the researcher's expectations.

During the equivalent semester, the chatbot was developed with the purpose of enhancing information processing and communication in accordance with the translanguaging concept. It was anticipated that this would foster students' mastery and comprehension of the Arabic language, particularly in the domain of writing. A primary focus of this research is the application of translanguaging in communication with the AI chatbot and the resulting implications for students. Translanguaging, which refers to the cognitive process by which individuals generate meanings by alternating between distinct linguistic systems and structures, was implemented and modified in Arabic-speaking activities across all classes subsequent to the collection of preliminary data for comparison purposes prior to acquiring the second set of data as the posterior hypothesis. The implementation of translanguaging serves to enhance the writing process by establishing a structured framework, promoting a comprehensive understanding, and directing paradigm alignment.

The results of the grades for the papers submitted by each participant can be described as follows:

Table 1.
Descriptives

	N	**Mean**	**SD**	**SE**	**Coefficient of variation**
PRE	45	5.542	0.903	0.135	0.163
POST	45	6.918	0.862	0.128	0.125

Note: Descriptive model by JASP 0.18.0.0

Based on the observations and monitoring of learning with the use of the chatbot in Arabic language learning, particularly in the writing category, several significant findings related to usability, comfortability, and the validity of using this chatbot in the context of effective Arabic language communication may involve mixing Arabic and Indonesian languages if translation challenges are encountered. Participants enjoyed all the sessions and practiced them in Arabic literacy activities, which were then documented in their papers. In this discussion, we will delve deeper into the results of the average differences between odd and even semesters.

The results of the descriptive table (Table 1) show that the mean pre-test score is 5.542, while the mean post-test score is 6.918. The mean is a statistical measure representing the "middle" or "average value" of a data group. In this case, we have two data groups, the pre-test and post-test, each with a mean of 5.542 and 6.918, respectively. When we compare these two means, we can see that the post-test mean (6.918) is higher than the pre-test mean (5.542). This indicates that, overall, there has been an improvement between the pre-test and post-test results after a certain intervention or change.

The pre-test standard deviation (0.903) is higher than the post-test standard deviation (SD) (0.862), indicating that the data in the pre-test group has a slightly higher level of variation compared to the post-test group. The pre-test Standard Error (SE) (0.135) is greater than the post-test SE (0.128), indicating that the uncertainty in the pre-test sample's mean estimate of the population mean is slightly higher than the estimate for the post-test sample mean. This suggests that after an event or intervention occurred between the pre-test and post-test, the post-test group tends to be more homogeneous, and the post-test mean may have become more stable. Additionally, the difference between SD and SE can provide insights into changes in diversity and uncertainty over time.

The coefficient of variation for the pre-test is a measure of variation or diversity in the data before a specific event or intervention. In this context, it is measured before an experiment or a particular change is applied. A coefficient of variation of 0.163 indicates that the pre-test data has a relatively high level of variation. This could indicate a wide distribution of the pre-test results, leading to relatively significant variations in the values. On the other hand, the coefficient of variation for the post-test measures the level of variation in the data after a specific event or intervention has occurred. A coefficient of variation of 0.125 indicates that the post-test data has less variation than the pre-test data. The difference between the coefficient of variation for the pre-test (0.163) and the post-test (0.125) provides insights into the effect of the event or intervention that occurred between the two measurements. The decrease in the coefficient of variation from the pre-test (0.163) to the post-test (0.125) suggests that the post-test results tend to be more uniform or stable compared to the pre-test results. This can be interpreted as an indication that the intervention or experiment may have reduced the variation or diversity in the data.

The results of the assumption test (Shapiro-Wilk) before the t-test are as follows:

Table 2.
Test of Normality (Shapiro-Wilk)

			W	**P**
PRE	-	POST	0.967	0.221

Note. Significant results suggest a deviation from normality.

Before proceeding to the t-test, the researcher conducted an assumption test to measure normality. In the Shapiro-Wilk table, the W value ranges between 0 and 1. The closer it is to 1, the better the data conforms to a normal distribution. A W value of 0.967, which is close to 1, indicates that the data closely resembles a normal distribution. The p-value is the probability generated by the Shapiro-Wilk test. The p-value is used to evaluate whether the research data follows a normal distribution. In this context, a p-value of 0.221 indicates that the probability that the data follows a normal distribution is approximately 0.221, or around 22.1%. This means that the research has strong enough evidence to support the assumption that the data follows a normal distribution because the p-value is greater than the commonly used significance level (e.g., 0.05 or 5%).

Next, the results of the classic paired sample T-Test are as follows:

Table 3.
Paired Samples T-Test

Measure 1		**Measure 2**	**t**	**Df**	**p**
PRE	- -	POST	-15.436	44	< .001

Note. Student's t-test.

The paired sample t-test is a statistical tool used to compare the means of two related measurements. Researchers measure the same thing before and after a specific event or intervention. In the results obtained (Table 3), the t-value is minus (-) 15.436, measuring the significance of the difference between the two measurement groups. A low t-value indicates that the difference between the groups is highly significant. This means that the observed changes in this data are strong and unlikely to have occurred by chance. The degrees of freedom (df) are a statistical measure that describes the number of individuals or measurements involved in this analysis. In this case, df is 44, which means there are 44 units of data or measurements used in the analysis. The larger the df value, the more accurate the t-test results. The p-value, often expressed as $p < 0.01$, is an indicator of the level of statistical significance. In this context, a very low p-value indicates strong evidence to support the difference between the measurement groups before and after a specific event or intervention. In other words, the difference is highly statistically significant. These results indicate that the observed event or intervention has had a very strong impact on this measurement group. The observed changes are not coincidental but rather the result of a highly significant effect. Therefore, these results can have significant practical implications in the context of this study.

To strengthen the findings, Bayesian Paired Sample t-Test analysis was conducted, resulting in an exceptionally large Bayes factor (BF_{10}) and an extremely small error rate (Liang & Dai, 2023; Van Doorn et al., 2021), as follows:

Tabel 4. Bayesian Paired Samples T-Test

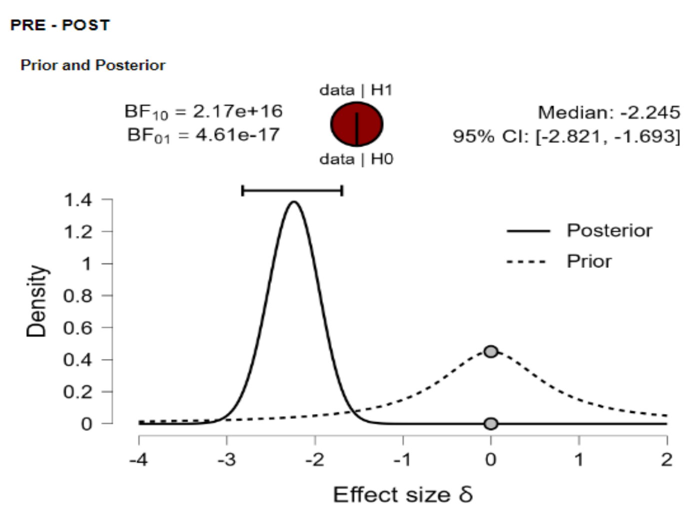
Measure 1		**Measure 2**	**BF_{10}**	**error %**
PRE -	-	POST	$2.169\times10^{+16}$	1.463×10^{-19}

Note: Bayesian Factors Analysis by JASP

The Bayesian paired t-test is a powerful tool in statistical analysis, especially when dealing with a large amount of data (Liang & Dai, 2023). The results in this case (Table 4) provide strong evidence in favor of the alternative hypothesis (H1) compared to the null hypothesis (H0). This is supported by the very large BF10 value, 2.169×10+16. This value effectively indicates that the data provides substantial support for the alternative hypothesis being investigated. With a BF10 value greater than 1, there is greater confidence that there is a significant effect in the data. It's important to note that the BF10 value here is not just a large number. Bayes factor tends to provide more consistent results and is less influenced by sample size, which is often an issue in classical statistics. Additionally, Bayes Factor helps address the problem of "multiple comparisons," where multiple hypotheses are tested simultaneously. The BF value provides a more comprehensive overview of the overall evidence in the data, which can help avoid "significant" results that may be found by chance. Thus, the very large BF10 result indicates that the findings are strong and have substantial relevance in the context of empirical studies.

Here are inferential plots, which are graphical representations used in Bayesian analysis—just as described by Vuong et al., (2020**)—**to visualize the distribution of data before and after incorporating new information, often presented as prior and posterior distributions. They serve as a valuable tool for understanding how beliefs about parameters change based on observed data. These plots offer a clear visual summary of the shift in probability distribution, helping researchers assess the impact of new evidence and make informed decisions in a Bayesian framework. They provide insight into the strength of evidence for or against specific hypotheses, enhancing the transparency and interpretability of Bayesian analyses.

Figure 1.
Inferential Plots

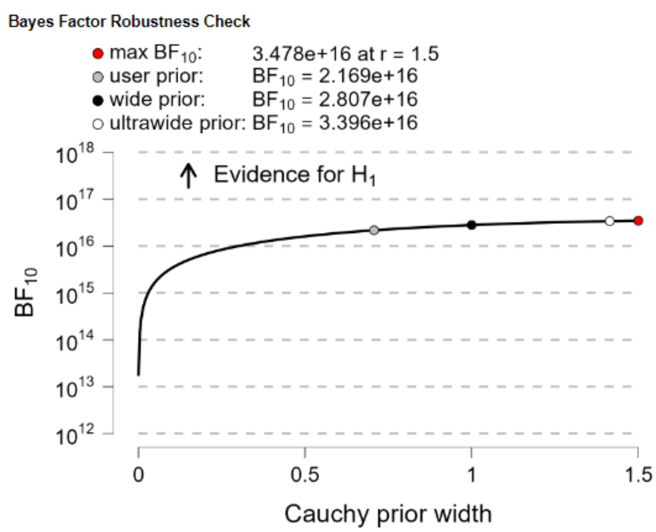


In statistical analysis, inferential graphs are a crucial instrument that aids in the comprehension and interpretation of data (Maier et al., 2022). A number of significant components are frequently observed in these diagrams (Figure 1), including the confidence interval (CI) and median. Utilizing a range of numbers, the Confidence Interval indicates with what degree of certainty we can place a particular parameter in the data. The confidence interval in this instance is [-2.281, -1.693]. This indicates that the range of values from -2.281 to -1.693 contains the actual parameter value with a 95% level of confidence. Within this particular framework, the parameter might symbolize the mean value or a distinct attribute of the data under examination. We have a 95% level of confidence that the parameter value is contained within this CI range. The median represents the midpoint of a sorted set of data in the diagram. The median in this instance is -2.245. This implies that -2.245 would be the value in the middle if all values were arranged in ascending order. The median gives information regarding the "middle point" of the set of data. In contrast, the mean (average) is calculated by dividing the sum of all values by the total number of values. By calculating the "middle value" that more accurately reflects the distribution of the data, the median is more resilient to outliers (extreme values).

The Bayes Factor Robustness Check is an additional concept within Bayesian statistics that is employed to assess the robustness of results, the stability of findings, and the likelihood of outliers occurring (Faulkenberry et al., 2020; Van Doorn et al., 2021). The Bayes Factor Robustness Check plot (Figure 2) exhibits a number of exceptionally large values, such as Max BF10, User Prior, Wide Prior, and Ultrawide Prior. A very large Max BF10 indicates that there is extremely strong evidence in support of the alternative hypothesis in every scenario and model examined. This suggests that the findings of the research are of great importance and are not contingent upon the selection of alternative or preceding scenarios. A substantial User Prior may suggest that users held firm convictions regarding the alternative hypothesis prior to the analysis. Wide Prior and Ultrawide Prior indicate that the analysis found strong evidence in support of the alternative hypothesis in every case despite considering a wide variety of potential scenarios, including those that were extremely diverse.

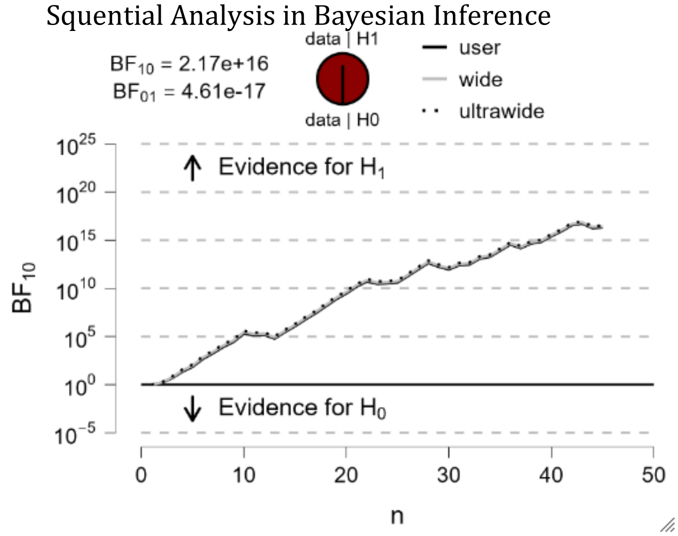
Based on the data, it can be visualized as follows:

Figure 2.
Bayes Factor Robustness Check



As for the sequential analysis plot, sequential analysis is a statistical method used by researchers in decision-making to continuously monitor data as it is received. It allows researchers to stop testing early when strong evidence has accumulated (Goss-Sampson et al., 2020; Maier et al., 2022). This is particularly useful when testing the null hypothesis (H0) and the alternative hypothesis (H1). Researchers look for evidence that the alternative hypothesis (H1) is becoming more likely and supported, while H0 is moving further away (Figure 3 below).

Figure 3.
Squential Analysis in Bayesian Inference



Sequential analysis offers the benefits of time and resource conservation (Goss-Sampson et al., 2020). In cases where sufficient evidence exists to support a conclusion, there is no necessity for researchers and participants to continue data collection. Furthermore, it enables scientists to react quickly to important changes in the data. This technology allows scientists to constantly observe data pertaining to a specific variable and formulate conclusions using evolving evidence. Sequential analysis can be utilized to observe the accumulation of evidence supporting the H1 hypothesis over time with a strong approach. This mechanism enables all stakeholders to arrive at decisions grounded in the evidence at hand while preserving adaptability in an investigative framework. Nevertheless, when deliberating on the implications of sequential analysis outcomes, it is critical to adhere to suitable statistical principles and take into account pragmatic and ethical considerations (Van Doorn et al., 2021).

The researcher applied a conventional methodology in the classic Paired Sample t-Test, which conforms to Kruschke's (2021) principles by analyzing significance levels, computing p-values, testing hypotheses, and quantifying probabilities. The conclusions drawn from this analysis indicate whether or not the observed variations in the data are statistically significant. Bayesian analysis, on the other hand, provides a more adaptable and versatile methodology. The researcher utilized Bayes factors to assess the level of confidence in a range of hypotheses and conducted posterior distribution-based calculations. By employing this approach, the investigator is able to track the progression of information updates and modify confidence levels in light of supplementary data via sequential analysis. The outcomes produced by both approaches are complementary. Traditional statistical analysis offers a preliminary assessment of the significance of the data, whereas Bayesian analysis provides a more comprehensive understanding of the uncertainty and probability distribution of the parameters under consideration.

Taking into account every aspect of the analysis—including Bayesian methods, sequential analysis, and statistical tests—the research findings provide substantial support for the observed intervention's efficacy. This suggests that the modifications or measures implemented in this study yield a favorable and statistically significant effect on the assessed variables. The implications of these findings are significant within the framework of this research, including for policy formulation, process enhancement, and advancement. Strong evidence of effectiveness is of the utmost importance in each of these contexts, as it can lead to improved decision-making in subsequent research and practice.

These results are consistent with those of DeCapua (2022) and Hajir et al. (2022) regarding comprehensive writing through the development of translanguaging and transliteration, respectively. Research pertaining to the application of chatbots in language development, specifically in the context of improving writing abilities via transliteration and translanguaging support, remains exceedingly scarce. In particular, the authors anticipate that the findings of this study will serve as a significant foundation for the advancement of AI-powered language learning in the fields of language education and translation dynamics.

CONCLUSION

The AI chatbot exhibits considerable potential for improving students' comprehension and competence in the Arabic language. The majority of respondents consider the AI chatbot to be a beneficial resource for facilitating Arabic language learning, specifically in terms of writing Arabic papers. The findings from the statistical study reveal a statistically significant disparity in the scores obtained before and after the test. Consistently exceeding the pre-test mean, the post-test mean signifies an important improvement in the group under observation. A paired sample t-test produces a t-value that is exceptionally small, a p-value that is considerably smaller than the frequently employed significance level (0.01), and a degree of freedom that is comparatively high. The aforementioned evidence indicates that the disparity between the two groups is substantially statistically significant, and the implemented intervention positively influences the results.

The Bayesian analysis, employing the Bayes factor (BF), offers robust evidence in favor of the intervention's efficacy. A BF value of 2.169×1016 and an error rate of 1.463×10-19 are indicative of substantial evidence supporting the alternative hypothesis (H1) in contrast to the null hypothesis (H0), which has a negligible error rate of 1.463×10-19. This finding suggests that the implemented intervention or modification has a substantial influence, and this aspect further validates the overall conclusion of this study. Notwithstanding the acknowledgment of research limitations caused by temporary, resource, sampling, and other considerations, the investigator maintains a resolute optimism regarding the investigation of additional studies related to translation schemes and their implementation in the domains of AI-assisted language learning.

BIBLIOGRAPHY

Abdulkader, Z., & Al-Irhayim, Y. (2022). A Review of Arabic Intelligent Chatbots: Developments and Challenges. *Al-Rafidain Engineering Journal (AREJ)*, *27*(2), 178–189. https://doi.org/10.33899/rengj.2022.132550.1148

Al-Abdullatif, A. M., & Alsubaie, M. A. (2022). Using digital learning platforms for teaching Arabic literacy: A post-pandemic mobile learning scenario in Saudi Arabia. *Sustainability*, *14*(19), 11868.

Ashoori, M., & Weisz, J. D. (2019). *In AI We Trust? Factors That Influence Trustworthiness of AI-infused Decision-Making Processes*. https://doi.org/10.48550/ARXIV.1912.02675

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., Yang, J., Toner, H., & Fong, R. (2020). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. *arXiv Preprint arXiv:2004.07213*. https://doi.org/10.48550/arXiv.2004.07213

Canals, L. (2022). The role of the language of interaction and translanguaging on attention to interactional feedback in virtual exchanges. *System*, *105*, 102721.

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, *20*(1), 38. https://doi.org/10.48550/arXiv.2305.00280

DeCapua, S. E. (2022). (Not) Lost in Translation: Multilingual Students, Translation, and Translanguaging in First-Year Writing. In *Global and Transformative Approaches Toward Linguistic Diversity* (pp. 206–222). IGI Global.

Elashhab, S. (2020). The impact of translanguaging on the EFL competence development of Arabic speaking learners. *The Asian EFL Journal*, *27*(3.1), 393–413.

Faulkenberry, T. J., Ly, A., & Wagenmakers, E.-J. (2020). Bayesian Inference in Numerical Cognition: A Tutorial Using JASP. *Journal of Numerical Cognition*, *6*(2), 231–259. https://doi.org/10.5964/jnc.v6i2.288

Fuad, A., & Al-Yahya, M. (2022). Recent Developments in Arabic Conversational AI: A Literature Review. *IEEE Access*, *10*, 23842–23859. https://doi.org/10.1109/ACCESS.2022.3155521

Goodman, B., & Tastanbek, S. (2021). Making the shift from a codeswitching to a translanguaging lens in English language teacher education. *TESOL Quarterly*, *55*(1), 29–53.

Goss-Sampson, M., van Doorn, J., & Wagenmakers, E. J. (2020, May 19). *Bayesian inference in JASP: A guide for students* [Monograph]. University of Greenwich; Jeffrey's Amazing Statistics Program (JASP). https://doi.org/10.17605/OSF.IO/CKNXM

Hajir, B., Rasman, R., & McInerney, W. (2022). Digital translanguaging and Arabic-English transliteration (Arabizi): Insights from Syria and Lebanon. *Cambridge Educational Reseach E-Journal*, *9*.

Hatherley, J. J. (2020). Limits of trust in medical AI. *Journal of Medical Ethics*, *46*(7), 478–481. http://dx.doi.org/10.1136/medethics-2019-105935

Iftanti, E. (2016). Improving students' writing skills through writing journal articles. *IAIN Tulungagung Research Collections*, *8*(1), 1–22. https://dx.doi.org/10.21274/ls.2016.8.1.1-22

Kolhar, M., & Alameen, A. (2021). Artificial Intelligence Based Language Translation Platform. *Intelligent Automation & Soft Computing*, *28*(1). http://dx.doi.org/10.32604/iasc.2021.014995

Kruschke, J. K. (2021). Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour*, *5*(10), Article 10. https://doi.org/10.1038/s41562-021-01177-7

Liang, W., & Dai, H. (2023). Bayesian inference. In *Quantum Chemistry in the Age of Machine Learning* (pp. 233–250). Elsevier. https://doi.org/10.1016/B978-0-323-90049-2.00005-6

Maier, M., Bartoš, F., Quintana, D., Dablander, F., van den Bergh, D., Marsman, M., Ly, A., & Wagenmakers, E.-J. (2022). *Model-averaged Bayesian T-tests*.

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: A guidance for policymakers*. UNESCO Publishing. https://doi.org/10.54675/PCSP7350

Michel-Villarreal, R., Vilalta-Perdomo, E., Salinas-Navarro, D. E., Thierry-Aguilera, R., & Gerardou, F. S. (2023). Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. *Education Sciences*, *13*(9), 856. https://doi.org/10.3390/educsci13090856

Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice, British University, Vietnam*, *20*(2). https://doi.org/10.53761/1.20.02.07

Putri Supriadi, S. R. R., Haedi, S. U., & Chusni, M. M. (2022). Inovasi pembelajaran berbasis teknologi Artificial Intelligence dalam Pendidikan di era industry 4.0 dan society 5.0. *Jurnal Penelitian Sains Dan Pendidikan (JPSP)*, *2*(2), 192–198. https://doi.org/10.23971/jpsp.v2i2.4036

Renz, A., & Vladova, G. (2021). Reinvigorating the discourse on human-centered artificial intelligence in educational technologies. *Technology Innovation Management Review*, *11*(5).

Rumaisa, F., Puspitarani, Y., Rosita, A., Zakiah, A., & Violina, S. (2021). Penerapan Natural Language Processing (NLP) di bidang pendidikan. *Jurnal Inovasi Masyarakat*, *1*(3), 232–235. https://doi.org/10.33197/jim.vol1.iss3.2021.799

SAPUTRI, D. S. C. (2021). *PENGEMBANGAN MODEL PEMBELAJARAN BAHASA INGGRIS BERBASIS AUGMENTED REALITY*.

Setiadi, F. M. (2021). Optimalisasi Model Pembelajaran Bahasa Arab Untuk Meningkatkan Keterampilan Berbicara (Mahārah Kalām) Santri Kelas X Pesantren Izzur Risālah Desa Sipapaga Mandailing Natal. *Journal of Community Dedication and Development (Pengabdian Kepada Masyarakat)*, *1*(1), 40–49. https://jurnal.stain-madina.ac.id/index.php/jcdd/article/view/329

Shao, S., Alharir, S., Hariri, S., Satam, P., Shiri, S., & Mbarki, A. (2022). AI-based Arabic Language and Speech Tutor. *2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA)*, 1–8. https://doi.org/10.1109/AICCSA56895.2022.10017924

Stahl, B. C. (2021). *Artificial intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies*. Springer Nature. https://library.oapen.org/handle/20.500.12657/48228

Van Doorn, J., Van Den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law*

Review International, *22*(4), 97–112.

Vuong, Q.-H., La, V.-P., Nguyen, M.-H., Ho, M.-T., Tran, T., & Ho, M.-T. (2020). Bayesian analysis for social data: A step-by-step protocol and interpretation. *MethodsX*, *7*, 100924. https://doi.org/10.1016/j.mex.2020.100924