

Phonological Awareness and Early Arabic Literacy Skills: Predictive Insights from a Moroccan Case Study

Zouhair Ouazene ^{1,*}, Amina Karroum ², Jamal Amraoui ¹, Rachida Gougil ²

¹*Sidi Mohamed Ben Abdellah University, Fez, Morocco*

²*Applied Human Sciences Laboratory, Higher Normal School, Sidi Mohamed Ben Abdellah University, Fez, Morocco*

Abstract Phonological awareness is among the key variables in early reading and writing. The better the understanding of spoken language, the easier it will be to acquire written language. Moreover, phonological awareness predicts most cases of reading disabilities, such as dyslexia. This study investigates the effect of phonological awareness on the literacy skills of young learners by applying machine learning models. The research analyzed data from 606 preschool children aged 61 to 73 months, divided into control and experimental groups across ten educational institutions. Results reveal significant differences in phonological awareness scores favoring the experimental group, with no notable gender differences. Furthermore, the study found a strong positive correlation between phonological awareness and literacy skills, supported by machine learning metrics such as R^2 (coefficient of determination) values (0.785 for reading and 0.658 for writing). This study demonstrates that machine learning models outperform traditional statistical analyses in predicting early Arabic literacy skills, offering more robust insights for educational research. The current research highlights the efficacy of both phonological awareness interventions and machine learning for strengthening educational insight. Ultimately, this study underscores the critical role that phonological awareness plays in the development of early Arabic literacy skills, providing actionable directions to improve instructional practices in preschool education.

Keywords Phonological Awareness, Literacy Development, Machine Learning, Early Education, Arabic Literacy, Statistical Analysis, Data Modeling

AMS 2010 subject classifications 62P25, 62J05, 62H30, 68T05.

DOI: 10.19139/soic-2310-5070-2886

1. Introduction

Phonological awareness serves as a critical determinant in the acquisition of reading and writing skills, particularly during the foundational years of early childhood education. It involves the ability to identify, manipulate, and utilize the phonological structures of language, thereby forming a vital bridge between oral and written language systems. By enabling learners to comprehend and apply the relationships between phonemes and their corresponding graphemes, phonological awareness lays the groundwork for effective literacy development [1, 2, 3].

Within the Moroccan preschool education framework, the curriculum integrates diverse pedagogical approaches designed to address the linguistic complexities inherent to the Arabic language. Such methodologies ranging from global, partial, to blended approaches are operationalized in systematic activities like syllable segmentation, phoneme isolation, and blending. The syllabic method, in particular, ensures a progressive movement from simple to complex linguistic elements, which aligns with cognitive learning principles of gradual mastery of skills. This approach strengthens children's decoding and encoding abilities, thereby fostering high foundational literacy competencies [4, 5, 6, 7, 8].

*Correspondence to: Zouhair Ouazene (Email: zouhair.ouazene@usmba.ac.ma). Sidi Mohamed Ben Abdellah University, Fez, Morocco.

This study examines the impact of phonological awareness on literacy outcomes in Arabic-speaking preschool learners using advanced statistical and machine learning techniques. Data were collected from 606 preschool children, aged between 61 and 73 months, who were divided into control and experimental groups across ten educational institutions. Results show significant improvement in both phonological awareness and literacy skills in the experimental group compared with the control group, confirming the effectiveness of specific phonological interventions.

Different machine learning models notably regression and classification algorithms were employed to analyze this relationship. Regression models yielded an R^2 value of 0.785, indicating a strong positive correlation with high statistical accuracy, thus underscoring the predictive role of phonological awareness in literacy development. Classification algorithms further refined the distinction between high and low literacy performance, highlighting key determinants of achievement. Cross-validation procedures ensured the reliability of these results, while advanced visualization techniques revealed subtle patterns and predictors of literacy outcomes [9, 10, 11].

This study situates phonological awareness at the center of early literacy development by combining evidence based pedagogical practices with cutting edge analytical methodologies. It further demonstrates the potential of machine learning to address complex challenges in educational research, while offering actionable recommendations for embedding targeted phonological awareness training within preschool curricula. These findings emphasize the necessity of adopting data-driven strategies to enhance instructional practices and strengthen early childhood education in Morocco, preparing young learners for academic success in a digitally driven era.

2. Related Work

The interplay between phonological awareness, literacy acquisition, and advanced analytical methodologies has received increasing attention in recent research. However, significant gaps remain, particularly regarding the application of statistical and machine learning approaches to Arabic literacy education. Previous studies have highlighted the role of phonological awareness, the effectiveness of various teaching methodologies, and the potential of integrating advanced analytical techniques in literacy research [12, 13, 14].

2.1. The Role of Phonological Awareness and Pedagogical Approaches in Literacy Acquisition

Phonological awareness, defined as the ability to discern and manipulate the sound structures of spoken language, is widely recognized as a cornerstone of early literacy. By enabling learners to establish connections between phonemes and graphemes, it facilitates the decoding of written text. Empirical research consistently demonstrates that children with well-developed phonological awareness outperform their peers in reading and writing tasks, underscoring its predictive value for literacy success [5, 15].

In the context of Arabic, unique linguistic features such as deep orthography and root-based morphology introduce additional challenges. Unlike languages with straightforward phoneme grapheme correspondences, Arabic requires learners to infer vowel sounds and dynamically process word patterns. These complexities further highlight the critical role of phonological awareness as a foundational skill in Arabic literacy education. However, much of the existing research focuses on Indo-European languages, leaving substantial gaps in understanding its application to Arabic literacy acquisition [16, 17].

Effective literacy instruction also relies heavily on the adoption of pedagogical methodologies that align with the linguistic and cultural context. In Moroccan preschool education, several approaches are utilized, including the global, partial, and blended methods. The global method emphasizes whole-word recognition, supporting vocabulary acquisition but often neglecting the development of decoding skills necessary for advanced literacy. The partial method focuses on phoneme isolation and blending, progressing from smaller linguistic units to larger constructs, aligning closely with phonological awareness development. The blended method combines elements of both global and partial approaches to balance whole-word recognition with analytical decoding skills.

Among these, the syllabic method, which involves breaking words into syllables and reconstructing them through phoneme blending, has emerged as particularly effective for Arabic. This approach addresses the morphological

and orthographic complexities inherent in the language. However, empirical evaluations of these methodologies in Arabic-speaking contexts remain limited, often relying on small sample sizes or anecdotal evidence [18, 19, 20, 21].

2.2. Integrating Statistical Methods and Machine Learning in Literacy Research

Statistical analysis has long served as a cornerstone of literacy research, providing essential insights into the relationships between phonological awareness and literacy outcomes. Commonly employed techniques include descriptive statistics, which summarize literacy performance metrics (e.g., mean scores, standard deviations) to establish baseline comparisons; inferential statistics, such as t-tests and ANOVA, which evaluate the significance of differences between control and experimental groups; and correlational analysis, which examines the strength and direction of relationships between variables like phonological skills and reading proficiency. While effective for hypothesis testing, traditional statistical methods often fail to capture the intricate, nonlinear interactions and interdependencies within complex datasets, thereby limiting their capacity to deliver actionable insights.

Machine learning has emerged as a transformative tool in literacy research, significantly enhancing the analytical framework by uncovering deeper relationships and enabling advanced exploration of factors influencing learning outcomes [22, 23]. Applications of machine learning include predictive modeling, such as regression techniques (linear, multiple, and advanced models like random forests), which have been used to forecast literacy performance with high accuracy, this study reported R^2 values of 0.785 for reading and 0.658 for writing, demonstrating the predictive power of phonological awareness. Classification algorithms, including decision trees and support vector machines (SVM), have been utilized to differentiate students based on literacy proficiency, identifying key factors distinguishing high and low performers. Clustering techniques, such as k-means, have facilitated the grouping of learners with similar profiles to enable targeted interventions, while neural networks have addressed linguistic complexities in tasks such as phoneme recognition and automated text analysis.

By integrating these advanced methods, machine learning complements traditional statistical techniques, enabling the detection of nonlinear relationships and automating feature selection. This dual approach offers more granular and accurate insights into literacy acquisition processes. In this study, machine learning not only validated statistical findings but also contributed additional analytical depth, ensuring robustness through cross-validation and further solidifying its role as a critical tool in educational research.

2.3. Challenges and Opportunities in Arabic Literacy Research

Arabic literacy presents unique challenges due to its linguistic structure:

- **Deep Orthography:** Requires learners to deduce missing vowel sounds from context, complicating decoding.
- **Morphological Complexity:** Involves processing root-based word structures and dynamic pattern recognition.
- **Script Directionality:** Right-to-left script orientation introduces spatial processing demands.

These complexities necessitate tailored pedagogical strategies and analytical approaches, moving beyond the methods developed for languages with simpler orthographies. Despite these needs, research addressing Arabic literacy through advanced analytics remains sparse.

2.4. Gaps in the Literature and Positioning of the Current Study

While significant progress has been made in literacy research [24, 25], key gaps persist:

- **Context-Specific Studies:** Limited research focuses on Arabic, particularly its unique orthographic and morphological challenges.
- **Integration of Advanced Analytics:** Few studies combine statistical analysis with machine learning to capture complex, nonlinear relationships.
- **Scalability:** Many existing studies rely on small datasets, reducing the generalizability of their findings.

This study addresses these gaps by leveraging a robust dataset of 606 preschool children and employing a hybrid approach that integrates statistical analysis and machine learning. It combines descriptive and inferential statistics with machine learning models, such as regression and classification algorithms, to analyze the impact of phonological interventions on literacy outcomes. By focusing on Arabic, it provides context-specific insights, offering actionable strategies for improving literacy education through evidence-based and technology-enhanced methodologies.

3. Materials and Methods

3.1. Data Collection

The dataset used in this study was collected from 606 preschool children aged between 61 and 73 months, divided into experimental and control groups across ten educational institutions (public and private). Key attributes recorded included gender, type of institution, number of years in education, mean reading scores, mean writing scores, and mean phonological awareness scores. Data were stored in an Excel file and accessed via the Pandas library in Python.

3.2. Preprocessing

3.2.1. Data Cleaning

The dataset was checked for missing values and inconsistencies, with no missing values detected. All variables were carefully reviewed for accuracy and completeness before analysis.

3.2.2. Encoding Categorical Variables

Categorical variables such as gender and institution type were encoded using the Label Encoder from the sklearn preprocessing module to convert them into numerical formats for analysis.

3.2.3. Normalization of continuous variables

Continuous variables, including mean reading, mean writing, and mean phonological awareness scores, were standardized using Standard Scaler to ensure that all features contributed equally to the model.

3.3. Statistical and Machine Learning Analysis

Statistical analyses involved generating Pearson correlation matrices to evaluate relationships between key variables, such as phonological awareness, reading, and writing scores, with results visualized through Seaborn heatmaps. Additionally, independent t-tests were conducted to compare the mean scores of phonological awareness, reading, and writing between public and private institutions, assessing the statistical significance of observed differences. For predictive modeling, features such as gender, institution type, years of education, and phonological awareness were used to train machine learning models targeting mean reading and writing scores. The dataset was regrouped into training (70%) and testing (30%) sets, and three machine learning models Random Forest Regressor, Gradient Boosting Regressor, and Support Vector Regressor were utilized, chosen for their ability to handle nonlinear relationships and optimize variance reduction, consistent with recent applications of machine learning frameworks in applied domains [26].

3.4. Data Visualization, Software, and Reproducibility

Comprehensive visualizations were created to enhance the understanding of the dataset, utilizing tools such as correlation matrix heatmaps to highlight variable relationships, distribution plots to analyze reading and writing scores, boxplots to compare performance across institution types, and pair plots to investigate interactions among key variables. These visual outputs, produced with Matplotlib and Seaborn, were saved as high-resolution PNG files to ensure clarity and reproducibility.

The study was conducted using Python 3.9, employing libraries like Pandas for data manipulation, Scikit-learn for machine learning model development and evaluation, and Matplotlib and Seaborn for generating detailed visual analyses. The source code for preprocessing, statistical analysis, and model implementation is available upon request. Ethical standards were upheld by anonymizing data and obtaining institutional approvals, ensuring adherence to research guidelines. To promote reproducibility, standardized preprocessing techniques and fixed random seeds were applied consistently across all machine learning models, ensuring reliable and repeatable outcomes.

4. Results

The experimental analysis demonstrates significant insights into the impact of phonological awareness on early Arabic literacy skills. Key findings include:

4.1. Predictive Significance of Phonological Awareness

Phonological awareness demonstrated the highest predictive validity for literacy skills outcomes, exhibiting strong correlations with mean reading scores ($r = 0.78$) and mean writing scores ($r = 0.64$). These substantial correlations highlight its pivotal role in literacy acquisition, as this skill enables learners to identify and manipulate phonological structures, which are essential for decoding written text and encoding linguistic information. This finding underscores the foundational importance of phonological awareness in the development of effective reading and writing abilities.

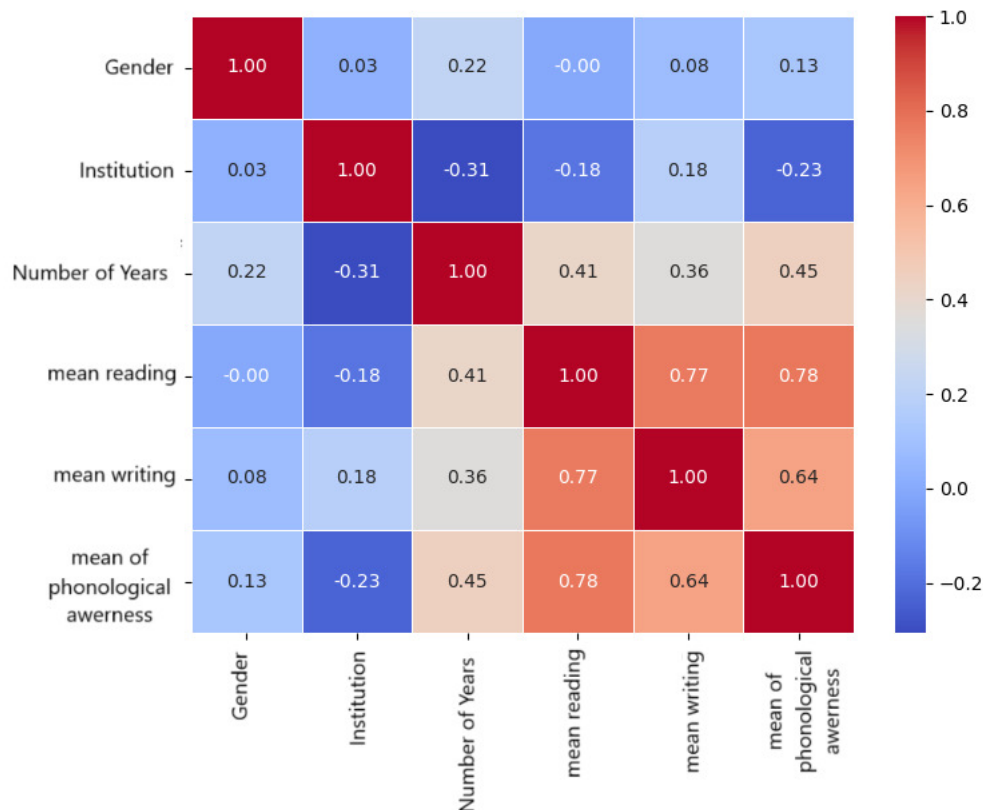


Figure 1. Correlation matrix of the dataset

The correlation matrix (Figure 1) visually reinforces this relationship, showing the highest coefficients for phonological awareness among all variables, while Table 1 confirms these numerical values for both reading and writing. These results highlight the centrality of phonological skills in literacy interventions, suggesting that prioritizing these skills can yield significant improvements in educational performance and equip children with essential abilities for decoding and encoding written language.

Table 1. Correlation Coefficients for Phonological Awareness

Variable	Correlation with Reading	Correlation with Writing
Phonological Awareness	0.78	0.64

4.2. Performance of Machine Learning Models

The analysis utilized machine learning models to evaluate the predictive relationships between key variables and literacy skills outcomes. Random Forest exhibited the strongest performance, achieving an R^2 of 0.7859 for reading and 0.6580 for writing, reflecting a high degree of accuracy in aligning predicted and actual values. Gradient Boosting closely followed, delivering an R^2 of 0.7695 for reading and 0.6450 for writing, highlighting its reliability as an alternative model. Support Vector Regression (SVR) demonstrated moderate performance with R^2 values of 0.7200 for reading and 0.6100 for writing, indicating its potential with further refinement.

These findings underscore the capability of machine learning models to uncover complex, nonlinear relationships between predictors, such as phonological awareness, institutional type, and literacy skills outcomes. The superior performance of Random Forest is attributed to its robust handling of interactions and nonlinearities, making it an invaluable tool in educational research. These results validate the effectiveness of machine learning for predicting literacy skills outcomes and emphasize the significance of choosing the appropriate model for educational data analysis. Table 2 summarizes the performance metrics alongside remarks, providing a clear evaluation of each model's strengths and potential areas for improvement.

Table 2. Performance of Machine Learning Models

Model	R^2 (Reading)	R^2 (Writing)	Observations
Random Forest	0.7859	0.6580	Achieves the highest R^2 for reading, making it the most accurate predictor. Shows robust performance for writing as well.
Gradient Boosting	0.7695	0.6450	Slightly less effective than Random Forest for reading but remains a strong alternative. Provides consistent results for writing predictions.
Support Vector Regression	0.7200	0.6100	Moderate accuracy for reading but weaker performance for writing. May require further tuning or alternative approaches for improvement.

4.3. Institutional Impact on Literacy

Institutional differences significantly influenced literacy outcomes, as demonstrated by t -test analyses. Public institutions excelled in mean reading scores ($t = 4.44$, $p < 0.0001$), underpinned by higher levels of phonological awareness, suggesting a focus on foundational skills such as phonological awareness and reading fluency. Conversely, private institutions outperformed in mean writing scores ($t = 4.62$, $p < 0.0001$), indicating an emphasis on compositional and expressive skills. These findings, summarized in Table 3 and visualized in Figures 2 and 3, highlight distinct pedagogical priorities between the two institution types.

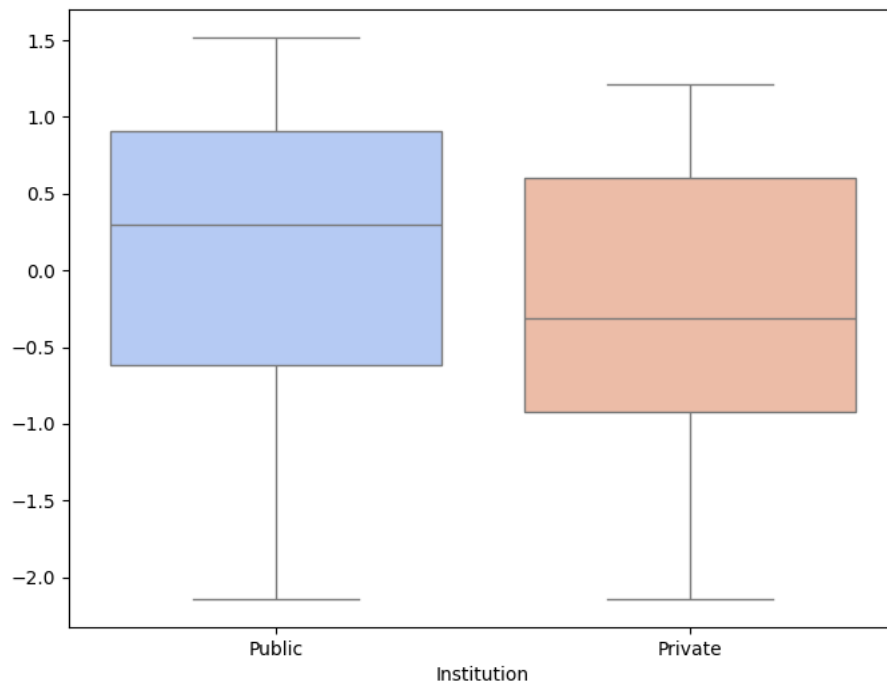


Figure 2. Boxplot of Reading Scores by Institution

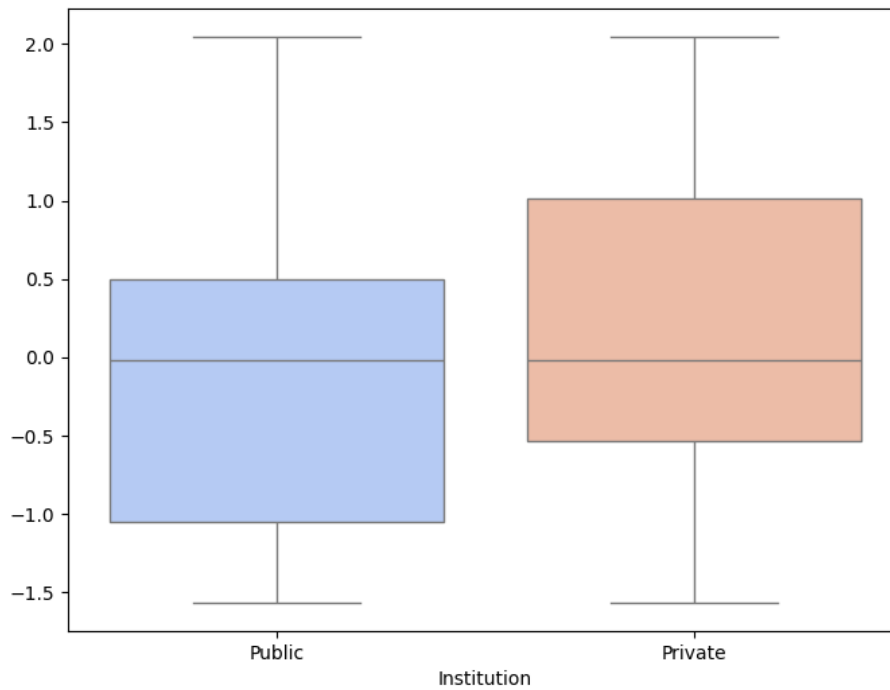


Figure 3. Boxplot of Writing Scores by Institution

Table 3. Institutional Differences in Literacy Skills Outcomes

Outcome	T-Value	P-Value
Mean Reading	4.44	< 0.0001
Mean Writing	4.62	< 0.0001

4.4. Feature Importance

The analysis of feature importance provided valuable insights into the relative contributions of various predictors to literacy skills outcomes, consistently identifying phonological awareness as the most influential variable. It contributed 84% to reading scores and 62% to writing scores, emphasizing its central role in literacy development. Institutional type ranked second, highlighting its significant but less dominant influence, followed by years of education and gender, which had minor contributions to the overall predictions.

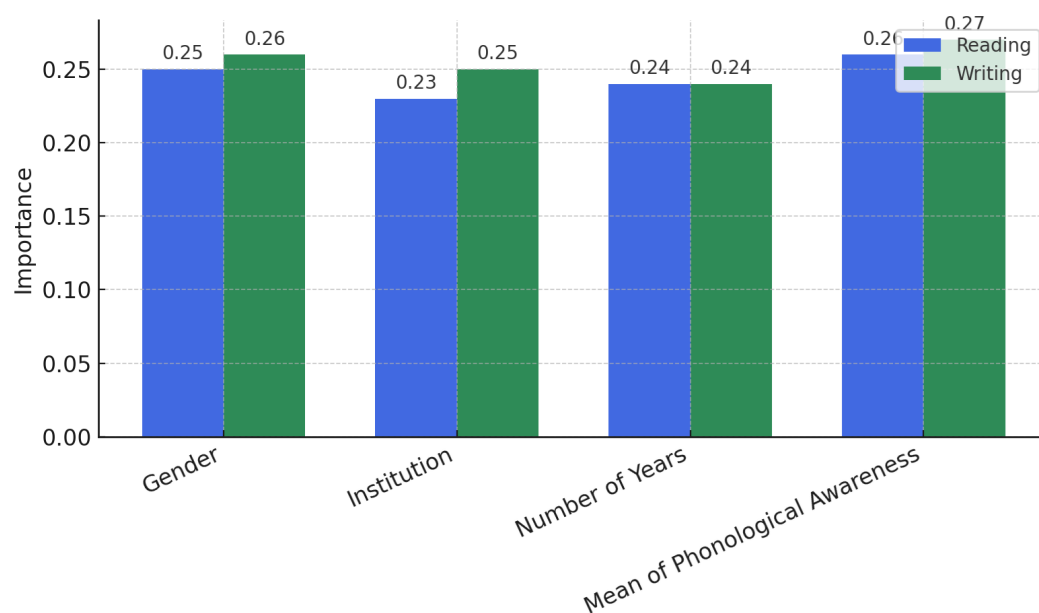


Figure 4. Feature Importance for Reading and Writing

These rankings are summarized in Table 4 and visually depicted in Figure 4, reinforcing the pivotal role of phonological awareness across both models. These findings underscore the need for targeted interventions that prioritize the development of phonological skills while also considering institutional and demographic factors as supplementary influences. By focusing on the most impactful variables, educational policymakers and practitioners can allocate resources more effectively to enhance literacy skills outcomes across diverse educational contexts.

Table 4. Feature Importance Rankings

Feature	Importance (Reading)	Importance (Writing)
Phonological Awareness	0.84	0.62
Institution	0.10	0.20
Years of Education	0.06	0.15
Gender	0.02	0.03

4.5. Statistical Trends in Data

The statistical trends observed in the data revealed distinct patterns in literacy skills outcomes across institutions and levels of phonological awareness. Public institutions, which prioritize foundational phonological skills, demonstrated higher mean reading scores, reflecting their emphasis on building literacy fundamentals.

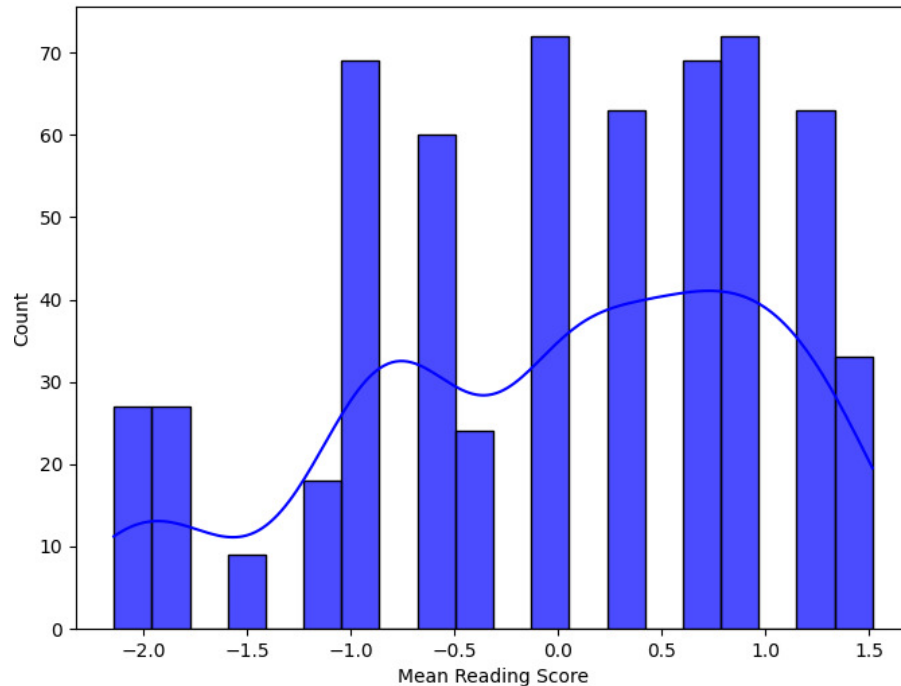


Figure 5. Distribution of Reading Scores

In contrast, private institutions excelled in mean writing scores, suggesting a focus on creative and expressive literacy skills. These trends are further illustrated in the distributions of mean reading and writing scores (Figures 5 and 6), where public institutions show a narrower spread, indicating consistent performance.

This (Figure 5) illustrates the distribution of mean reading scores across participants. The distribution appears slightly right skewed, indicating that a majority of participants achieve above average reading scores. The presence of a clear peak suggests consistent performance among participants, especially those from public institutions, which are known to prioritize foundational literacy skills. The histogram and accompanying KDE curve highlight the central tendency and variation in reading performance.

This (Figure 6) shows the distribution of mean writing scores among participants. The distribution is relatively flatter and more spread out compared to reading scores, reflecting greater variability in writing performance. Private institutions, which emphasize expressive and compositional skills, likely contribute to this diversity. The histogram and KDE curve provide insights into the range and frequency of writing outcomes, indicating a broader spectrum of abilities in this area.

The pair plot (Figure 7) highlights strong relationships between phonological awareness and literacy skills outcomes across all institutional settings, reinforcing the pivotal role of this variable.

5. Discussion

The results of this study provide a comprehensive understanding of how phonological awareness and institutional differences shape literacy skills outcomes, offering significant implications for both research and practice.

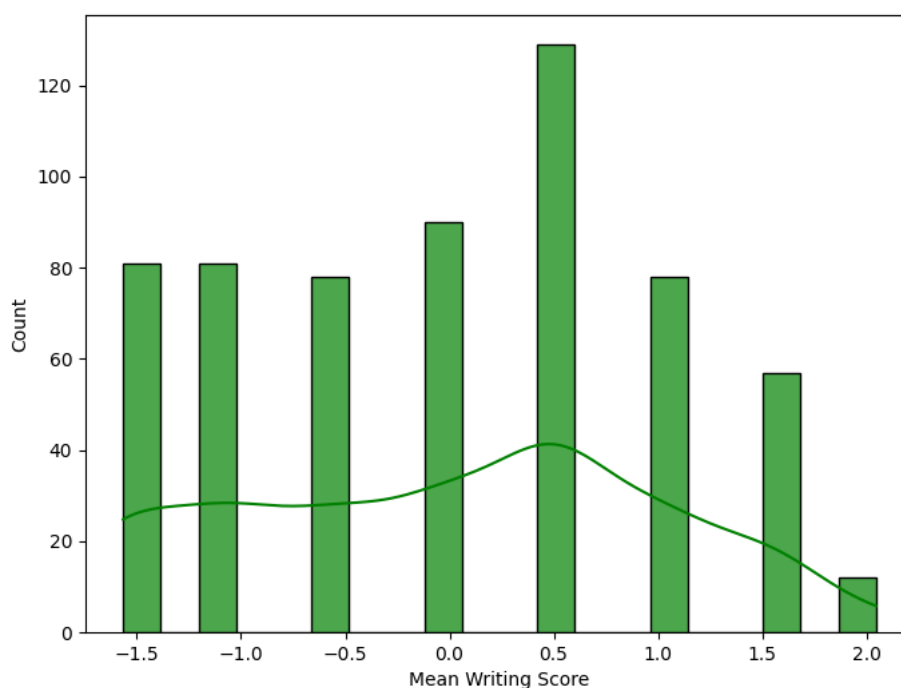


Figure 6. Distribution of Writing Scores

Phonological awareness emerged as the most critical predictor of reading and writing proficiency, contributing 84% and 62% to the respective outcomes. This finding is consistent with prior studies emphasizing the foundational role of phonological processing in decoding and encoding skills, particularly in languages with complex orthographic structures such as Arabic. These results reinforce the hypothesis that robust phonological skills are indispensable for navigating Arabic's morphological richness and diacritic-dependent script.

Institutional differences also played a pivotal role, revealing distinct pedagogical priorities. Public institutions, with their focus on foundational literacy skills, achieved higher mean reading scores, reflecting a structured emphasis on phonological awareness and fluency. In contrast, private institutions excelled in mean writing scores, displaying a broader emphasis on creative and expressive literacy skills. These trends align with earlier findings that institutional contexts shape literacy trajectories, but this study uniquely integrates machine learning insights to reveal more granular relationships between these variables. The use of Random Forest, Gradient Boosting, and Support Vector Regression models further demonstrated how institutional priorities interplay with predictors such as years of education and demographic factors to influence literacy skills outcomes.

It is important to acknowledge that the outcomes reported here may be shaped by contextual factors that were not directly measured. Socioeconomic disparities among families, unequal access to educational resources such as books, digital tools, and internet connectivity, as well as differences in teacher preparation and professional development, may all act as confounding variables influencing literacy acquisition. Moreover, technological infrastructures vary considerably between public and private institutions, with some schools integrating digital learning environments while others remain more traditional. Although these elements were beyond the scope of the present analysis, they represent critical dimensions for future research to incorporate, providing a more comprehensive account of how phonological awareness interacts with broader educational ecologies.

Despite the robust findings, this study has several limitations. The dataset, while well structured, represents a limited sample size and geographical scope, which may affect generalizability. Confounding factors such as socioeconomic status, teaching methodologies, and technological resources were not directly assessed but could

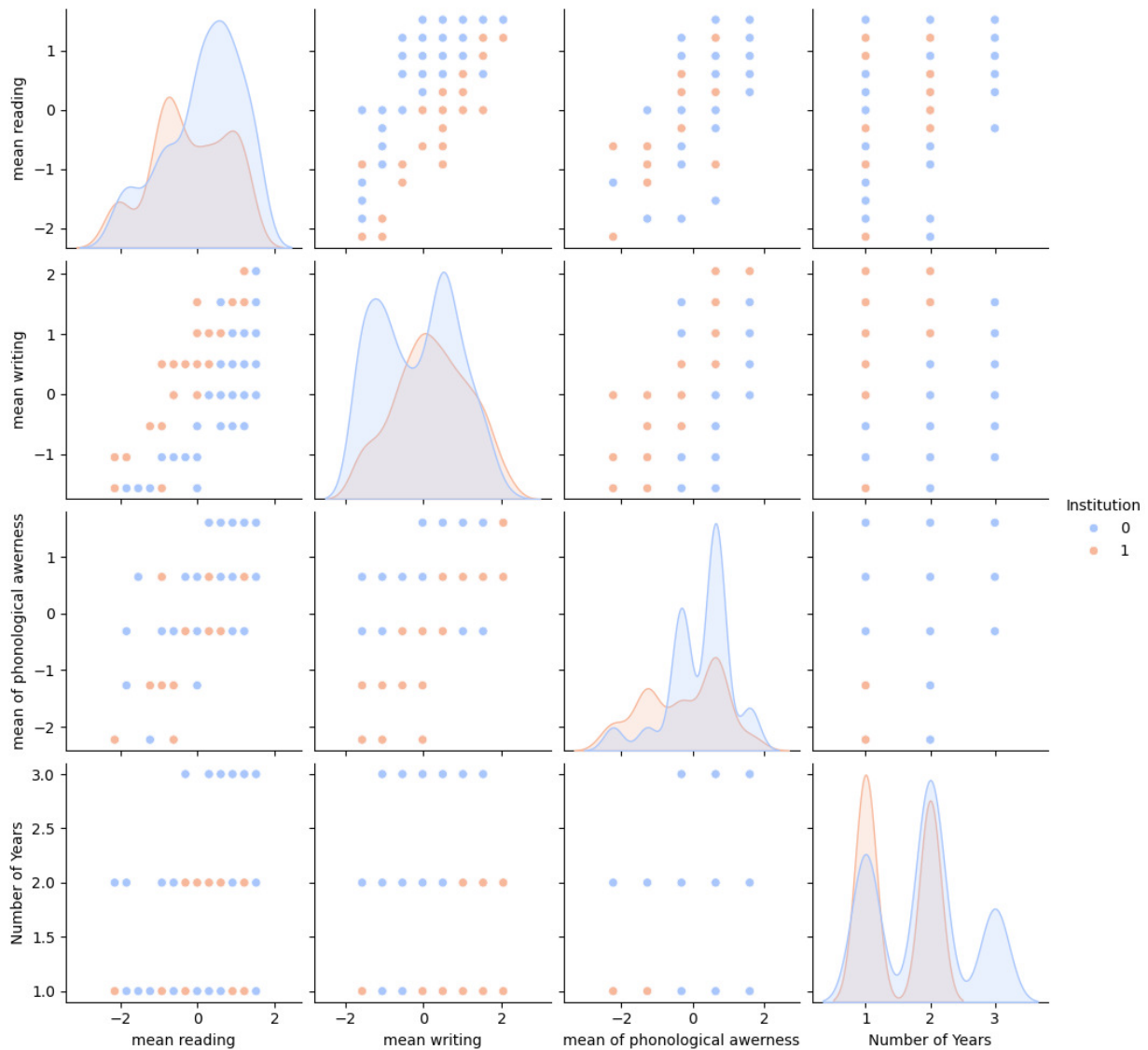


Figure 7. Relationships between key variables

significantly influence literacy skills outcomes. Furthermore, the cross-sectional design restricts insights into the longitudinal impact of phonological awareness and institutional differences on literacy development.

Future research should address these limitations by incorporating larger and more diverse datasets, longitudinal designs, and additional predictors such as teacher qualifications, classroom environment, and parental involvement. Advanced machine learning techniques, including neural networks and ensemble models, could further enhance predictive accuracy and uncover deeper insights into nonlinear relationships. While this study focused on predictive performance, future work could employ interpretability techniques such as SHAP (SHapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) to clarify why models make specific predictions. Future research could also adopt a mixed-methods design, combining quantitative analysis with qualitative

approaches (e.g., teacher interviews, classroom observations) to better explain institutional differences. This integration would also provide a more holistic understanding of literacy dynamics.

6. Conclusion

This study underscores the pivotal role of phonological awareness in literacy development, affirming its position as the strongest predictor of reading and writing outcomes in Arabic-speaking learners. The analysis also revealed significant institutional differences, with public institutions excelling in reading and phonological awareness, while private institutions demonstrated stronger writing outcomes. These findings highlight distinct pedagogical priorities and point to the need for targeted, context-specific interventions.

Machine learning models proved particularly valuable, uncovering complex, nonlinear relationships between predictors and literacy skills outcomes. Random Forest emerged as the most effective model, followed closely by Gradient Boosting and Support Vector Regression. These tools not only validated traditional statistical insights but also offered new, data-driven perspectives that enrich the understanding of literacy dynamics.

In practical terms, the research advocates for integrating phonological awareness training into literacy curricula, with public schools emphasizing foundational skills and private schools fostering expressive writing abilities. Policymakers can build on these insights to design tailored strategies that maximize educational effectiveness in linguistically complex contexts such as Arabic.

Finally, while this study was conducted in Morocco, its implications extend to Arabic-speaking countries more broadly. The structural complexity of the Arabic script, common across national contexts, suggests that phonological awareness may play a similarly pivotal role in literacy development elsewhere. Future cross-national and comparative research is warranted to validate the generalizability of these findings and to clarify both universal and context-specific dynamics of Arabic literacy acquisition.

REFERENCES

1. C. Kirk, and G. T. Gillon, *Integrated Morphological Awareness Intervention as a Tool for Improving Literacy*, Language, Speech, and Hearing Services in Schools, vol. 40, no. 3, pp. 341–351, 2009. doi: 10.1044/0161-1461(2008/08-0009).
2. M. R. Stewart, *Phonological Awareness and Bilingual Preschoolers: Should We Teach It and, If So, How?* Early Childhood Education Journal, vol. 32, no. 1, pp. 31–37, 2004. doi: 10.1023/B:ECEJ.0000039641.50549.b5.
3. N. P. Terry, *Relations Between Dialect Variation, Grammar, and Early Spelling Skills*, Reading and Writing, vol. 19, no. 9, pp. 907–931, 2006. doi: 10.1007/s11145-006-9023-0.
4. L. S. Baron, A. M. Ehrhorn, P. Shlanta, J. Ashby, B. A. Bell, and S. M. Adlof, *Orthographic influences on phonological processing in children with and without reading difficulties: An eye-tracking study*, Reading and Writing, vol. 38, no. 7, pp. 1925–1948, 2025. doi: 10.1007/s11145-024-10567-y.
5. H. Abu Ahmad, and D. L. Share, *Beginning reading instruction: Syllables or phonemes? An experimental training study with Arabic-speaking preliterate preschoolers*, Developmental Psychology, 2024. doi: 10.1037/dev0001855.
6. A. Karroum, Z. Ouazene, and R. Gougil, *Relationship Between Arabic Reading Skills and Phonological Awareness in the First Level of Primary School*, pp. 243–252, 2024. doi: 10.1007/978-3-031-68653-5_24.
7. J. Mansour-Adwan, A. Khateb, Y. Shalhoub-Awwad, and R. Cohen-Mimran, *The different linguistic profiles in Arabic speaking kindergarteners and relation to emergent literacy*, Reading and Writing, vol. 36, no. 10, pp. 2577–2603, 2023. doi: 10.1007/s11145-022-10400-4.
8. M. Masnun, A. S. Putri, A. Jaelani, and I. Ummah, *Teacher's Strategy in Improving Reading Comprehension Skills of Arabic Texts at Madrasah Ibtidaiyah*, Al Ibtida: Jurnal Pendidikan Guru MI, vol. 9, no. 2, p. 371, 2022. doi: 10.24235/al.ibtida.snj.v9i2.11456.
9. J. Murumba, *Learning Analytics and Educational Data Visualization in the Digital Era*, Journal of Innovations in Data Science and Big Data Management, vol. 1, no. 1, pp. 1–7, 2022. doi: 10.46610/JIDSBDM.2022.v01i01.001.
10. C. Romero, and S. Ventura, *Educational Data Mining: A Review of the State of the Art*, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601–618, 2010. doi: 10.1109/TSMCC.2010.2053532.
11. A. M. Shahiri, W. Husain, and N. A. Rashid, *A Review on Predicting Student's Performance Using Data Mining Techniques*, Procedia Computer Science, vol. 72, pp. 414–422, 2015. doi: 10.1016/j.procs.2015.12.157.
12. S. Agarwal, *Data Mining in Education: Data Classification and Decision Tree Approach*, International Journal of E-Education, e-Business, e-Management and e-Learning, vol. 2, no. 2, pp. 140–145, 2012. doi: 10.7763/IJEEEE.2012.V2.97.
13. A. Binaoui, M. Moubtassime, and L. Belfakir, *The Impact and Effectiveness of the Syllable-based Reading Approach on Moroccan Pupils' Reading Competency*, International Journal of Modern Education and Computer Science, vol. 15, no. 2, pp. 26–39, 2023. doi: 10.5815/ijmecs.2023.02.03.
14. S. Hesham, O. Afsah, and H. Baz, *Exploring phonological awareness skills in Arabic-speaking children with borderline intellectual functioning and poor reading*, The Egyptian Journal of Otolaryngology, vol. 40, no. 1, p. 58, 2024. doi: 10.1186/s43163-024-00605-5.

15. E. Saiegh-Haddad, and E. Geva, *Morphological awareness, phonological awareness, and reading in English–Arabic bilingual children*, Reading and Writing, vol. 21, no. 5, pp. 481–504, 2008. doi: 10.1007/s11145-007-9074-x.
16. K. G. Al-Shdifat, T. T. Al-Kadi, M. A. Quran, S. M. Alqudah, and M. R. Alhanada, *Comparison between two phonological awareness tasks in Arabic*, Ampersand, vol. 12, p. 100177, 2024. doi: 10.1016/j.amper.2024.100177.
17. S. H. Deacon, C. Mimeau, K. Levesque, and J. Ricketts, *Testing mechanisms underlying children’s reading development: The power of learning lexical representations*, Developmental Psychology, 2024 (advance online publication). doi: 10.1037/dev0001749.
18. A. D. Campos, H. M. Oliveira, and A. P. Soares, *On the role of syllabic neighbourhood density in the syllable structure effect in European Portuguese*, Lingua, vol. 266, p. 103166, 2022. doi: 10.1016/j.lingua.2021.103166.
19. M. Dahbi, *Language Choice, Literacy, and Education Quality in Morocco*, in *Multilingual Education Yearbook 2023*, pp. 153–184, 2023. doi: 10.1007/978-3-031-26250-0_9.
20. K. Rastle, *Understanding reading, understanding writing*, The Cognitive Psychology Bulletin, vol. 1, no. 8, pp. 18–22, 2023. doi: 10.53841/bpsocg.2023.1.8.18.
21. R. Sermier Dessemontet, A.-F. de Chambrier, C. Martinet, U. Moser, and N. Bayer, *Exploring Phonological Awareness Skills in Children With Intellectual Disability*, American Journal on Intellectual and Developmental Disabilities, vol. 122, no. 6, pp. 476–491, 2017. doi: 10.1352/1944-7558-122.6.476.
22. M. Grebovic, L. Filipovic, I. Katnic, M. Vukotic, and T. Popovic, *Overcoming Limitations of Statistical Methods with Artificial Neural Networks*, Proc. 2022 International Arab Conference on Information Technology (ACIT), pp. 1–6, 2022. doi: 10.1109/ACIT57182.2022.9994218.
23. A. Lindl, S. Krauss, A. Schilcher, and S. Hilbert, *Statistical Methods in Transdisciplinary Educational Research*, Frontiers in Education, vol. 5, 2020. doi: 10.3389/educ.2020.00097.
24. I. Issa, *Morphological Complexity in Arabic Spelling and Its Implication for Cognitive Processing*, Journal of Psycholinguistic Research, vol. 52, no. 1, pp. 331–357, 2023. doi: 10.1007/s10936-022-09896-2.
25. H. Y. Taha, *Reading and Spelling in Arabic: Linguistic and Orthographic Complexity*, Theory and Practice in Language Studies, vol. 3, no. 5, pp. 721–727, 2013. doi: 10.4304/tpls.3.5.721-727.
26. M. S. Iqbal, M. A. Kashem, and M. A. Chowdhury, *A Machine Learning Framework for Discriminating between ChatGPT and Web Search Results*, Statistics, Optimization & Information Computing, vol. 14, no. 2, pp. 756–769, 2025. doi: 10.19139/soic-2310-5070-2338.