International Academic Institute for Science and Technology

International Academic Journal of Humanities
Vol. 12, No. 3, 2025, pp. 1-8.

ISSN 2454-2245

International Academic Journal of Humanities

www.iaiest.com

Developing a Predictive Model for Language Acquisition in Multilingual Environments Based on Cognitive and Social Factors

Zeenat Fatima Zaidi [a]

[a] *King Saud University, Saudi Arabia. E-mail: zeenatzaidi@hotmail.com*

Abstract

This study comes up with a predictive model of the language acquisition process in multilingual contexts and it is a combination of both cognitive and social factors. Due to the increasingly global world, people tend to be exposed to different languages and thus it is even more important to understand how things work so that one stands a chance to learn languages successfully. The study fills in the gap in the current literature because it develops a data-driven model that combines factors of cognitive abilities and social context to predict the outcomes of language proficiency. A stratified random sample of 200 multilingual participants all aged between 18-30 years was selected. Structured questionnaires, standardized tests of language proficiency, and cognitive tests, were used to collect data which assessed the working memory and the ability to control attention. Statistical analysis, including descriptive statistics, Pearson correlation, and multiple regression, was performed using IBM SPSS (Version 26), and machine learning models using Python in scikit-learn. The data set was divided into training (80%) and testing (20%), with 5-fold cross-validation. Results indicate that language exposure ($r = 0.58$, $p < 0.01$) and working memory ($r = 0.52$, $p < 0.01$) are the strongest predictors of language acquisition. The results of regression analysis indicate that language exposure ($\beta = 0.34$) and working memory ($\beta = 0.29$) have a significant effect on proficiency, and the overall model accounts 64% of the variance ($R^2 = 0.64$). The highest accuracy of the machine learning methods was recorded with the Random Forest model (78%), followed by the Support Vector Machine (74%). These results suggest that the effect of cognitive processing and social interaction on the multilingual learning outcomes is combined. The paper concludes that it is possible to use predictive modeling to effectively model the complexities of language acquisition in multilingual environments. Practical implications of the model to personalized education and policy planning include: Future research should be based on longitudinal designs and more heterogeneous population to enhance the generalizability and model interpretability.

Keywords: Multilingualism, Language Acquisition, Predictive Modeling, Cognitive Factors, Social Interaction, Machine Learning, Language Proficiency.

DOI: 10.71086/IAJH/V12I3/IAJH1208
Received: 09 June, 2025/Revised: 15 July 2025/Accepted: 22 August 2025/Published: 30 September 2025

Introduction

Multilingual learning environment has become a very crucial research area in the face of globalization, migration and a culturally diverse learning environment. These days, an individual can be exposed to multiple languages simultaneously, be it at school, at work or even in the digital world. The awareness of how learners acquire and manage multiple languages is not only significant to the learning process of how to improve education practices, but also in knowing how to improve effective communication, social integration, and cognitive development. Multilingualism has been associated with a variety of cognitive advantages, such as the enhancement of executive functioning and problem-solving abilities; however, acquisition of multiple languages is a complex process, and it is caused by a variety of interacting factors (Schmid & Yılmaz, 2018). These are individual mental capabilities, social context, exposure pattern and motivation and a complete and systematic analysis of language acquisition in such a setting is necessary (Trebits et al., 2022).

Language acquisition Cognitive theories of language acquisition provide a formative basis over which the manner in which people learn languages are examined (Li & Lan, 2022). The traditional theories, such as behaviorism, focus on the existence of the language acquisition device as an in-built one, with the nativist theories, in particular, Chomsky, being the representatives of such a notion. More modern theories, such as cognitive and usage-based theories, are concerned with the influence of memory, attention, and pattern recognition in language learning (Jia & Paradis, 2015). Connectionist models also indicate that language acquisition is an emergent process due to neural network-like mechanisms through repeated exposure to linguistic input (Lorenzo et al., 2019). All these theories reiterate the role of cognitive processes in language development, such as working memory, information processing, and learning strategies (Paradis, 2023).

Although the theoretical and empirical literature on language acquisition is quite extensive, one can still notice a significant gap in the process of creating predictive models integrating both cognitive and social variables, especially in the context of multilingualism. The existing literature tends to study these factors separately or in monolingual or bilingual contexts, restricting the scope of its relevance to more multilingual environments. Moreover, no data-driven and computational methods can be found that can be used to systematically predict the outcomes of language learning due to the dynamic interplay of variables. This gap needs to be addressed to continue the theoretical knowledge and practical applications of this learning which include personalized language learning and adaptive learning systems. The proposed research will therefore aim at developing a predictive model which will capture the interplay of the cognitive and social factors which influence the acquisition of language in a multilingual environment and this will help in providing a more holistic and empirically rooted framework.

By demonstrating that the results of the learning process are predetermined in an interactive process of cognitive abilities and the social situation, the present research contributes to the research on language acquisition in multilingual settings. The combination of statistical and machine learning systems provides a more solid and scalable approach towards predicting language proficiency than traditional models. Its findings can be uniquely applied in designing personalized learning plans, which enables an educator to create an individual plan to follow. The research additionally provides evidenced-based information to policy-makers to create language-rich teaching environments and all-encompassing multilingual policies. The contribution of the research both theory-wise and practically is in that it is the basis of further developments of the adaptive language learning systems.

Key Contributions

- Theory of Multilingual language acquisition: constructs a hybrid predictive model which combines cognitive and social factors.
- Gives empirical data that were statistically validated, and the key predictors were found to be language exposure and working memory.
- It applies both machine learning (Random Forest, SVM), and the traditional regression, to predict up to 78% of accuracy.
- Creates a completely data-driven, all-encompassing framework that bridges theoretical models of language acquisition with practical predictive analytics.

In this paper, the outline of the paper is divided into six parts. Section I presents the research problem, significance and objectives. Section II describes a literature review of the pertinent literature on cognitive and social influences in the acquisition of language. The research methodology, which involves data collection and modeling, is described in Section III. Section IV discusses the results and statistical findings. Section V makes the interpretation of the results

and emphasizes the implications, limitations, and applications. Lastly, the section VI wraps up the research with some major findings and the future line of research.

Literature Survey

The recent studies of language acquisition in multilingual settings emphasize the complexity of the interaction between the cognitive processes and the social environment. Research highlights how cognitive architectures and online learning environments contribute to the language acquisition processes, indicating that learning is based on memory, attention, and neural adaptability (Giorgi et al., 2020; Nawaz et al., 2024). On the same note, in another study, a social-cognitive perspective is used in which it is demonstrated that multilingual awareness evolves in both an internal cognitive processing perspective and an external linguistic exposure perspective (Shan & Xu, 2024). These results are in line with the current trend in understanding language learning that cannot be accounted for by a one-dimensional theoretical perspective but requires an integrative approach to language learning (Gross & Kaushanskaya, 2020).

The significance of the environmental and social factors is further supported by empirical studies. These studies demonstrate that language exposure, age, and social interaction play a significant role in the process of bilingual development, especially in diverse and immigrant populations (Paradis & Jia, 2017; Paradis et al., 2020). The above studies emphasize that the home linguistic environment has a role to play in improving linguistic and socio-emotional developments (Sun, 2019; Brito, 2017) Also, studies determine individual variations in language development, and are associated with cognitive skills and environmental factors (Brownell et al., 2016; Lauro et al., 2020).

Cognitive factors are also present that contribute a larger role in multilingual competence. The studies reveal that the working memory, cognitive control, and the intensity of multilingual usage are strong predictors of language performance (Pot et al., 2018; Gross & Kaushanskaya, 2020). It also demonstrates that multilingualism can be used to develop cognitive abilities, especially where the disadvantaged are concerned, pointing to its greater developmental effect (Tsimpli et al., 2020). In addition, predictive viewpoints in language development, as addressed in this study, are indicative of the increased significance of modeling approaches to predict language outcomes based on various variables (d'Souza et al., 2017; AlHammadi, 2017).

Although these advances have taken place, there is a definite gap that lies between the combination of both the cognitive and social variables in a single predictive model. Most of the existing literature discusses one of the two factors, the cognitive processes or the environmental factors, in isolation and little use is made of the sophisticated predictive modelling tools. This paper fills this gap by combining both statistical and machine learning techniques to come up with a comprehensive model of predicting language acquisition in a multilingual environment and thus contributes towards the existing body of literature.

Methodology

The research design that is employed in the current study is geared towards the systematic study of the cognitive and social determinants of language acquisition in multilingual environment and at the creation of a robust predictive model, on the basis of empirical evidence.

To ensure that the study will be inclusive in terms of age, education and level of language exposure, the study identified 200 participants with the assistance of a stratified random sampling method. Multilingual school institutions and community-based settings were used to select the sample in which the sample age range was 18-30 years old. All the participants were previously exposed to two or more languages and had different levels of proficiency in a third or fourth language. A mixed methods research design was used to gather data by the combination of structured questionnaires, standardized tests of language proficiency, and cognitive assessment tasks. Demographic data included in the questionnaire, pattern of language use and context of social interaction, were measured using cognitive measures, which measured the working memory, control of attention and speed of processing information. There were also semi-structured interviews of a sub-sample of 30 participants to get a more in-depth information about their experiences in learning languages and their social environments.

Various independent and dependent variables are used in the predictive model. The dependent variable is the proficiency in language acquisition as indicated by standardized test scores in the skills of listening, speaking, reading and writing. Stratification of the independent variables is based on cognitive and social variables. Cognitive variables are working memory capacity, attention span, cognitive flexibility and knowledge regarding the language that has already been learned. Some of the social variables are the frequency of language exposure, interaction with peers,

language background of the family, the learning environment, and the motivation to learn the language. Control variables are also added to describe the potential confounding factors such as age, gender, and socio-economic status.

To conduct statistical analysis, the study shall use a mix of descriptive statistics, correlation analysis, and multiple regression modeling to determine how variables relate to each other. The descriptive statistics (mean, standard deviation) summaries the data, which is followed by Pearson correlation analysis testing the strength and direction of the relations between the variables of cognitive, social, and language proficiency. In order to come up with the predictive model, a multiple linear regression is used to determine the relative contribution of each independent variable to the results of language acquisition. Also, the machine learning models can be applied to learn more about predictive accuracy and model validation, including Random Forest and Support Vector Machine (SVM). The metrics that can be used in order to assess model performance are: R^2, Mean Squared Error (MSE), and cross-validation techniques, which are used in order to ensure reliability and generalizability of the findings.

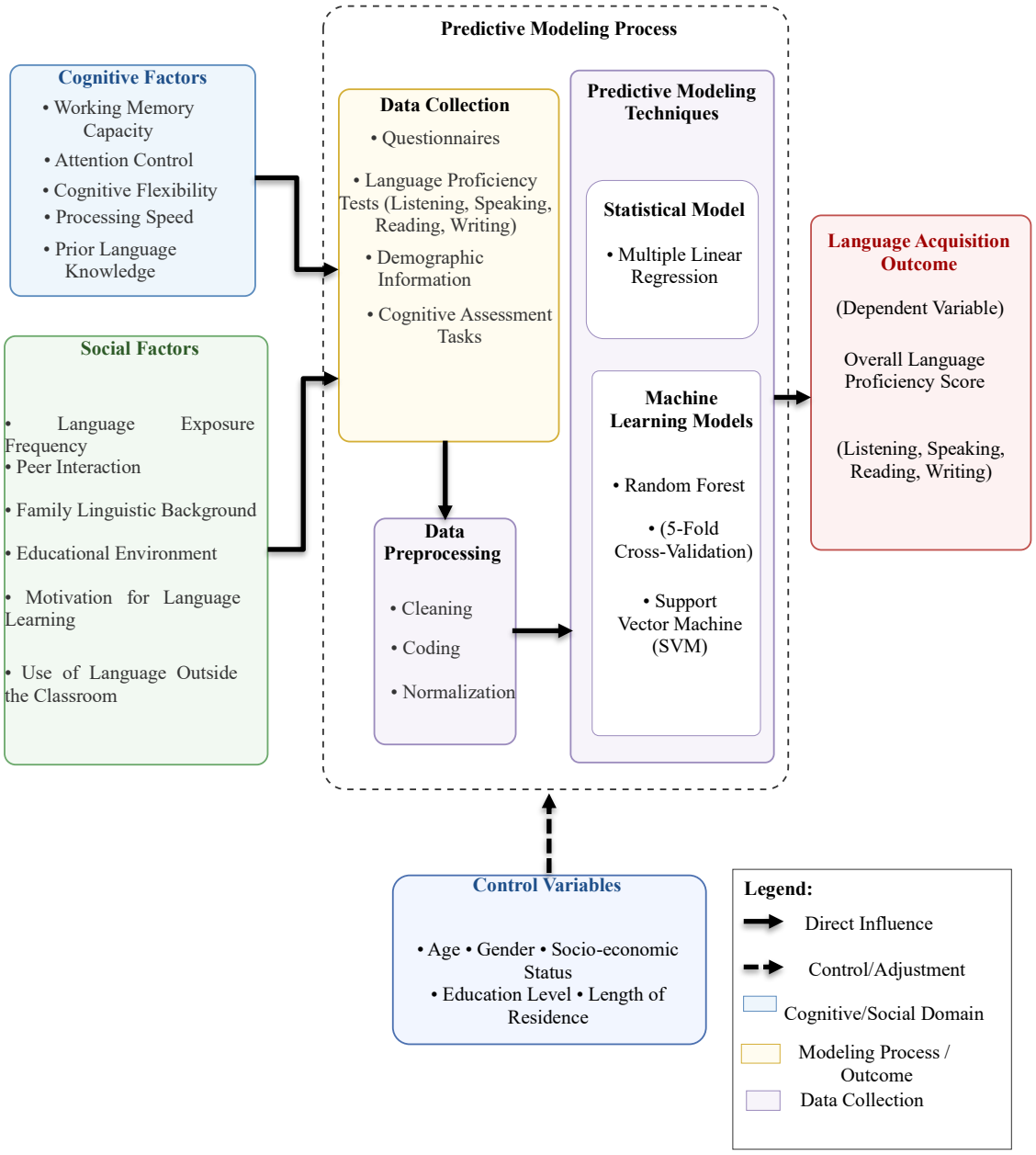


Fig. 1: Conceptual Framework for Predictive Modelling of Language Acquisition in Multilingual Environments

Fig. 1 demonstrates a systematic model of the role of cognitive (including the working memory and attention) and social (including language exposure and peer interaction) factors in the language acquisition outcomes. It emphasizes the data collection, preprocessing, and predictive modeling phases, which combine statistical and machine learning methods with control variables that affect the overall model.

Results

IBM SPSS Statistics (Version 26) was used to implement descriptive statistics, correlation, and regression analysis, whereas Python (Version 3.10) with Python libraries (scikit-learn, pandas, and NumPy) was used to implement machine learning. The data set comprised of the responses of 200 participants, which included the questionnaire data, cognitive assessment test results, and the standardized language proficiency test results. Data preprocessing was done using normalization, missing data and feature scaling to improve the model performance. The data were divided into training (80 %) and testing (20 %) subsets and cross-validation was followed (k=5) in such a way that the robustness, reliability, and generalizability of predictive models are guaranteed.

The study findings give a detailed analysis of the collected data, and how the cognitive and social variables relate and influence the language acquisition process, and the predictive model performance.

Data cleaning revealed a total of 200 valid responses that were analyzed. The sample included multilingual learners whose levels of variability were moderate or high in the case of cognitive and social factors. According to table 1, the average language proficiency score was 72.4 (SD = 10.6), which indicates that overall, the language proficiency of the participants was intermediate. The distribution of cognitive variables including working memory and attention control, and social variables including language exposure and peer interaction were found to be relatively stable and less variable respectively.

There was a correlation and regression (Table 2) analysis that showed that both cognitive and social variables have a significant impact on language acquisition. There were positive correlations between language proficiency and working memory (r = 0.52, p < 0.01) and attention control (r = 0.47, p < 0.01). Among the social factors, frequency of language exposure (r = 0.58, p < 0.01) and peer interaction (r = 0.49, p < 0.01) proved to be the significant predictors. A minimum of multiple regression showed that the most influential predictors were language exposure (β = 0.34), and working memory (β = 0.29), followed by attention control (β = 0.21) and motivation (β = 0.18). Weaker and statistically insignificant effects were found in control variables (age and socio-economic status).

The predictive model exhibited good performance in both the statistical and machine learning models are shown in table 3. The multiple linear regression model estimated the value of R^2 to be 0.64, which means that the selected variables could explain almost 64% of the variance in language acquisition. Random Forest model had a better accuracy of 78% as compared to Support Vector Machine (SVM) which had a lower accuracy of 74%. The cross-validation findings showed the stability and generalizability of the models, and there was little difference in variance among the folds.

Table 1: Descriptive Statistics of Key Variables

Variable	Mean (M)	Std. Dev (SD)
Language Proficiency	72.4	10.6
Working Memory	3.68	0.74
Attention Control	3.55	0.81
Language Exposure	3.92	0.69
Peer Interaction	3.61	0.77
Motivation	3.80	0.72

Table 2: Correlation and Regression Analysis

Variable	Correlation (r)	Regression Coefficient (β)	Significance (p-value)
Working Memory	0.52	0.29	< 0.01
Attention Control	0.47	0.21	< 0.01
Language Exposure	0.58	0.34	< 0.01
Peer Interaction	0.49	0.19	< 0.01
Motivation	0.44	0.18	< 0.05
Socio-economic Status	0.12	0.06	> 0.05

Table 3: Predictive Model Performance

Model	R^2 / Accuracy	MSE	Remarks
Multiple Linear Regression	0.64	28.5	Good explanatory power
Random Forest	78%	21.3	Highest prediction accuracy
Support Vector Machine	74%	24.1	Stable and consistent performance

On the whole, the results confirm that both social and cognitive variables are influential in the process of acquiring multiple languages with the social exposure and the cognitive capacity being the determining variables. The predictive model has good performance, and this aspect justifies its applicability in the context of educational and linguistic research.

Discussion

The results of this study provide valuable information on the topic of language acquisition in multilingual environments by showing that cognitive as well as social factors have a joint impact on the process of learning. The great role of working memory, control of attention and exposure to language means that language acquisition is not only a process of internal thought, but also greatly aided by meaningful social interaction and interaction with the surrounding. The results are in line with contemporary cognitive and usage-based theories, and the independent and interactive effects of repeated exposure, interaction and cognitive processing capacity are what facilitate language learning. The fact that the predictive value of the social variables, in particular, the exposure to the language, is greater in the immersive and interactive context, than in the isolated learning situation.

In spite of these contributions, the study has a number of limitations. Firstly, the sample size of 200 respondents, though large enough to perform the statistical analysis, may not be large enough to be able to fully represent the diversity of multilingual populations working in different cultural and linguistic settings. Second, the cross-sectional design does not enable studies to examine the changes in language learning over time, which can be better fulfilled with longitudinal studies. Third, measurement of some of the variables such as motivation and social interaction was done by using the self-reported data that could be biased in its response. Besides, as much as machine learning models may be more accurate in prediction than their traditional statistical analogues, it may not be as interpretable as it was. It is also recommended that future research should adopt larger and more diverse samples, longitudinal data, and should consider more advanced explainable AI approaches to enhance transparency and applicability of the model.

The practical applications of the predictive model that was developed by this study to teachers and policy-makers have a number of practical uses. As an educator, the model may be applied to create individualized learning plans by determining students that might need further cognitive or social support. It may also be educative in the curriculum development since it exposes the exposure-based and interactive learning strategies to the limelight. The results indicate that policy makers ought to establish learning environments that are accommodative to multilinguals such as the encouragement of language-rich classrooms and community involvement schemes. The data-driven decision-making model based on the model can be applied to other issues related to the educational planning, resource allocation, and development of specific language intervention programs, which will ultimately result in more efficient and inclusive language education systems.

Conclusion

The present study was aimed at developing a predictive model of language acquisition in a multilingual environment in terms of a combination of cognitive and social factors. The results clearly show that the two areas have a large and complementary contribution in the development of language proficiency. Descriptive statistics showed that the participants had the moderate level of multilingual competence (M = 72.4, SD = 10.6) and that the correlation analysis revealed strong positive correlations between the key variables such as the language exposure (r = 0.58, p < 0.01), working memory (r = 0.52, p < 0.01), and attention control (r = 0.47, p < 0.01) as well as the language acquisition outcomes. Such results are consistent with the fact that higher exposure and cognitive abilities are associated with higher proficiency. The multiple regression analysis also helped to establish the relative strength of the predictors, with the language exposure (β = 0.34), and the working memory (β= 0.29) turning out to be the most significant predictors, followed by the attention control (β = 0.21) and motivation (β= 0.18). The general model showed a significant explanatory ability as it explained 64% of the variance (R^2 = 0.64) in language acquisition. In addition, the machine learning techniques improved the accuracy of predictions and the Random Forests model (78%) and the Support Vector machine (SVM) model (74%) have good results, which proves the strength and stability of the proposed framework. The validity of the relationships established is supported by the fact that the statistical

significance ($p < 0.05$) of most of the variables used is statistically significant. The results are useful in developing a more holistic understanding of the multilingual language acquisition in terms of interactions between the internal and the external cognitive and social contexts. However, further research in the future must increase the sample size by extending to the more diverse linguistic and cultural groups in order to improve generalizability. It is also recommendable to employ longitudinal studies to obtain the developmental changes over time. Additionally, it could be logical to expect that the explainable artificial intelligence techniques could be employed to increase the understandability and usability of the model. This study, as a rule, provides a statistically grounded and practically relevant model, which can be used to inform the educational strategies and policy development in multilingual environments.

References

Giorgi, I., Golosio, B., Esposito, M., Cangelosi, A., & Masala, G. L. (2020). Modeling multiple language learning in a developmental cognitive architecture. *IEEE Transactions on Cognitive and Developmental Systems*, *13*(4), 922-933. https://doi.org/10.1109/tcds.2020.3033963

Nawaz, M., Nizamani, M., & Hameed, R. (2024). Analyze how children acquire language and the cognitive processes involved, including the role of environmental and social factors. *Bulletin of Business and Economics (BBE)*, *13*(3), 239-247. https://doi.org/10.61506/01.00483

Brownell, M. D., Ekuma, O., Nickel, N. C., Chartier, M., Koseva, I., & Santos, R. G. (2016). A population-based analysis of factors that predict early language and cognitive development. *Early Childhood Research Quarterly*, *35*, 6-18. https://doi.org/10.1016/j.ecresq.2015.10.004

Paradis, J., & Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Developmental science*, *20*(1), e12433. https://doi.org/10.1111/desc.12433

Shan, Z., & Xu, H. (2024). Exploring multilingual awareness development in learners of multiple foreign languages: A social cognitive perspective. *Journal of Multilingual and Multicultural Development*, *45*(8), 3205-3218. https://doi.org/10.1080/01434632.2022.2092119

Pot, A., Keijzer, M., & De Bot, K. (2018). Intensity of multilingual language use predicts cognitive performance in some multilingual older adults. *Brain sciences*, *8*(5), 92. https://doi.org/10.3390/brainsci8050092

Paradis, J., Soto-Corominas, A., Chen, X., & Gottardo, A. (2020). How language environment, age, and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*, *41*(6), 1255-1281. https://doi.org/10.1017/s014271642000017x

Sun, H. (2019). Home environment, bilingual preschooler's receptive mother tongue language outcomes, and social-emotional and behavioral skills: One stone for two birds?. *Frontiers in Psychology*, *10*, 1640. https://doi.org/10.3389/fpsyg.2019.01640

Schmid, M. S., & Yılmaz, G. (2018). Predictors of language dominance: An integrated analysis of first language attrition and second language acquisition in late bilinguals. *Frontiers in psychology*, *9*, 1306. https://doi.org/10.3389/fpsyg.2018.01306

Trebits, A., Koch, M. J., Ponto, K., Bruhn, A. C., Adler, M., & Kersten, K. (2022). Cognitive gains and socioeconomic status in early second language acquisition in immersion and EFL learning settings. *International Journal of Bilingual Education and Bilingualism*, *25*(7), 2668-2681. https://doi.org/10.1080/13670050.2021.1943307

Brito, N. H. (2017). Influence of the home linguistic environment on early language development. *Policy Insights from the Behavioral and Brain Sciences*, *4*(2), 155-162. https://doi.org/10.1177/2372732217720699

Li, P., & Lan, Y. J. (2022). Digital language learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition*, *25*(3), 361-378. https://doi.org/10.1017/s1366728921000353

Jia, R., & Paradis, J. (2015). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism: Language and Cognition*, *18*(4), 737-752. https://doi.org/10.1017/s1366728914000728

Lorenzo, F., Granados, A., & Ávila, I. (2019). The development of cognitive academic language proficiency in multilingual education: Evidence of a longitudinal study on the language of history. *Journal of English for Academic Purposes*, *41*, 100767. https://doi.org/10.1016/j.jeap.2019.06.010

Tsimpli, I. M., Vogelzang, M., Balasubramanian, A., Marinis, T., Alladi, S., Reddy, A., & Panda, M. (2020). Linguistic diversity, multilingualism, and cognitive skills: A study of disadvantaged children in India. *Languages*, *5*(1), 10. https://doi.org/10.3390/languages5010010

Gross, M. C., & Kaushanskaya, M. (2020). Cognitive and linguistic predictors of language control in bilingual children. *Frontiers in Psychology*, *11*, 968. https://doi.org/10.3389/fpsyg.2020.00968

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793-817. https://doi.org/10.1017/s0305000922000708

d'Souza, D., d'Souza, H., & Karmiloff-Smith, A. (2017). Precursors to language development in typically and atypically developing infants and toddlers: the importance of embracing complexity. *Journal of child language*, *44*(3), 591-627. https://doi.org/10.1017/s030500091700006x

Lauro, J., Core, C., & Hoff, E. (2020). Explaining individual differences in trajectories of simultaneous bilingual development: Contributions of child and environmental factors. *Child Development*, *91*(6), 2063-2082. https://doi.org/10.1111/cdev.13409

AlHammadi, F. S. (2017). Prediction of child language development: A review of literature in early childhood communication disorders. *Lingua*, *199*, 27-35. https://doi.org/10.1016/j.lingua.2017.07.007