

## Artificial Intelligence in EFL Context: Rising Students' Speaking Performance with *Lyra* Virtual Assistance

Junaidi<sup>1</sup>, Budianto Hamuddin<sup>2\*</sup>, Kenny Julita<sup>3</sup>, Fathu Rahman<sup>4</sup>, Tatum Derin<sup>5</sup>

<sup>1,2,3,5</sup>Universitas Lancang Kuning, Pekanbaru, Indonesia

<sup>2,4</sup>Universitas Hasanuddin, Makassar, Indonesia

Corresponding Email: budiamuddin@gmail.com

### Abstract

*The aim of this study tries to measure the effectiveness of artificial intelligence apps in English as a Foreign Language (EFL) classroom environment. The Alapps employed in this experimental study is Lyra Virtual Assistant (LVA), which was chosen with the line of thinking it can help the students improve their speaking performance. LVA was selected due to its numerous unique features but simplicity in use, lack of cost, and among one of the best ranking of virtual assistance in 2018. The sample in this study is selected secondary school recommended by the Ministry of Education due to the school's regulation in this school that allows students to bring a mobile phone in the classroom and its consistency of weakness in English speaking performance based on examination result. This study recorded 65 students from two different seventh grade classes who were grouped into experimental and control groups. A quasi-experimental was used as the research method with pre-test and post-test, focused on the four components in speaking skill i.e., pronunciation, grammar, vocabulary, and fluency. The results showed there was a significant result in the post-test from the experimental group who used LVA (69.59) compared to the control class with their conventional style (63.61). This means LVA is an effective AI apps for EFL students to improve their speaking skill.*

**Keywords:** Artificial intelligence (AI), Lyra Virtual Assistant (LVA), English as Foreign Language (EFL), Speaking skill.

### 1. Introduction

Information technology (IT) advances rapidly nowadays to the point it drives educational institutions to adopt mobile computing devices for academic activities (Bajcsy, 2002; Gikas & Grant, 2013). Educational institutional has enhanced learning curriculum with learning activities in the virtual learning environment (Hamuddin, Kurniawan, Syaifullah, & Herdi, 2018), making use of IT technology (Yamamoto et al., 2018). One of them is artificial intelligence (AI) which apparently has a tremendous impact in the pedagogical field (Navarro, Gonzalez, & Molina, 2018), though it is only recently that AI voice interaction (Pokulevska, 2018) improved more in the last 30 or so months than it did in its first 30 years (Nordrum, 2017).

In general, AI is designed as virtual assistance in executing and easing human tasks (Russell & Norvig, 2016). When this idea is brought into an educational setting, it can help both teacher and students in teaching and learning process. Powered by IT such as AI systems, software, and supports, location and presence is no longer a hindrance in a learner's quest to access course content and interact with instructors and colleagues at any time they wish (Cavus & Ibrahim, 2009; Hamuddin, 2016; Kukulska-Hulme & Shield, 2008; Richardson & Lenarcic, 2008; Shih & Mills, 2007). Educational programs driven by AI has been helping students to learn many skills considering AI offers students a much wider range of services, improves learners' motivation, broadens the audience's frame of conventional-based learning onto the digital world which is essential in the 21<sup>st</sup> century (Wekke, Yandra, & Hamuddin, 2017), and in the past few years, provides positive effect in language learning activity such as English as a Foreign Language (EFL) (Gric-coast et al., 2000; Hamuddin & Dahler, 2018).

Education in a foreign language usually has a heavy focus on grammar, syntax, and spelling (Chomsky & Halle, 1965), but in the early 1980s, the grammar-based approach has been acknowledged to have largely failed in producing fluent speakers of foreign languages as the communication-based approach to learning foreign language reached its peak (Canale & Swain, 1980). Nowadays, the trend of learning foreign language is primarily driven by the motivation to be fluent in speaking the target language (Wiranti, 2013). Spoken language often involves false starts, omissions, hesitations, and errors (Castillo, 2009) which contribute to imperfect speaking performance (Allerton, 1991), and studies to improve English speaking ability is abundant. With the increased affordability and reliability of AI technologies, the use of AI in teaching EFL has become a practical choice (Abad,

2013).

When students speak to an AI, its Automatic Speech Recognition (ASR) recognizes their speech and voice in a sequence (Keshet & Bengio, 2009; Li, Deng, Haeb-Umbach, & Gong, 2016) that is similar to how humans recognize speech, otherwise known as Human Speech Recognition (HSR) (Allen, 1994). This means AI can be used to improve speaking a foreign language when a native speaker of the foreign language is unavailable in the classroom context. Using AI for this purpose is actually still at its infancy because it is only when computers began to understand what words mean and the endless variability of expression in how language is spoken (Hill, Randolph Ford, & Farreras, 2015) in the early 2010s did AI start to be successful in recognizing speech automatically (Lewis Johnson & Valente, 2009). This study contributes to the fresh field of using AI in improving EFL speaking performance.

The students tested in this study are from a selected secondary school. The few studies that investigated the effects of educational technology such as AI on EFL performance have only tested primary and higher education students (Aditya, 2016; Aust, Kelley, & Roby, 1993; Brett, 1998; Davis & Lyman-Hager, 1997; Maftoon, Hamidi, & Sarem, 2012; Plass, Chun, Mayer, & Leutner, 1998; Solano, Cabrera, Ulehlova, & Espinoza, 2017; Underwood, 2017). On the one hand, with the popularity of Amazon's *Alexa*, Apple's *Siri*, and Google assistant, the majority of higher education students are already familiar with using AI to improve their mastery in a foreign language. On the other hand, primary school students are essentially "blank slates" who have just begun to familiarize themselves with IT. The selected secondary school chosen in this study has a regulation that allows the selected secondary students to bring IT such as mobile phones in the classroom, but have yet to use an AI's ASR feature as educational technology. Using selected secondary students as the sample of the study gives insight into how EFL speaking performance will be affected when the use of mobile phones in the classroom is changed from merely being a tool to instantly access information from the internet to a device that can recognize and respond to speech. Moreover, the selected secondary school was recommended by the Ministry of Education because the students' English speaking performance have been consistently weak, and such has been confirmed by way of observation. It was in high hopes that this study will significantly improve the students' EFL speaking ability with the use of AI.

Specifically, the AI this study selected is *Lyra* Virtual Assistant (LVA). *Lyra*, marketed under the name *Indigo* until 2017, is an intelligent personal assistant developed by Artificial Solutions (Budzinski, Noskova, & Xijie, 2019) and which runs on Android and iOS. The application uses natural language understanding to answer a user's questions on a variety of topics, make recommendations, and operate the user's device. LVA is unique amongst many other AIs because it can also be the user's "talking friend" for it has distinct features including searching for YouTube videos, tell jokes, find directions, manage a diary, set alarms, and some other features such as chat, question-answer interaction, and many others, yet it is simple to use and appropriate for secondary school students. These features make *Lyra* able to serve as an excellent conversation partner to improve EFL speaking ability, and because it can be downloaded from Play store for free and is ranked as one of the best virtual assistants for Android in 2018, this study chose LVA as the AI technology that will improve the selected secondary school students' EFL speaking ability.

## 2. Method

This research is designed as quasi-experimental research which is intended to find out the effect of using *Lyra* Virtual Assistant in students' speaking performance. In conducting this research, the researchers used two groups. One class is used as an experimental group (X) which is treated by using LVA and another class as a control group (Y) which is treated without using LVA. Both groups were given pre-test and post-test. This research is conducted in the second semester which started from February 15th, 2018 until March 12th, 2018 in academic year 2017/2018.

The population of this research is the seventh-grade students of SMP-TB Pekanbaru. There are three classes of the seventh grade of SMP-TB Pekanbaru. This school was selected based on the recommendation from the Ministry of Education due to the regulation in this school that allowed students to bring a mobile phone in the classroom as well as the fact that this school was found weak in English especially in the students' speaking performance. The sample of this research is VII.1 as the control group that consists of 33 students and VII.3 as the experimental group that consists of 32 students. Based on observation, this class has equal English ability and English speaking skill. This research used a test as the research instrument to see the effectiveness of using *Lyra* Virtual Assistant in speaking. Oral test was used in the pre-test and the post-test at the first meeting and the last meeting respectively. The oral test was designed to focus on the four components of speaking skill, i.e., pronunciation, grammar, vocabulary and fluency with the maximum score was 4 and the highest total score was 16.

To achieve the main objective of this study, first, the subjects were divided into control and experimental groups. The control group consisted of 33 students, and experimental groups consisted of 32 students. For both

groups, two 40-minute sessions were held in every meeting for a period of four weeks. The classes were held in the same school where the subjects attended their ordinary school classes. Before teaching the material, researchers did a pre-test in both classes. A pre-test is done in accordance with the material in the syllabus which is about oral and written text encouraging the test takers to state and ask the nature of people, animals, and objects. The researchers asked students one by one to identify and/or describe animals. The average student identified and/or described as many as three until five sentences and talked with incorrect intonation and pronunciation.

After conducting the pre-test, researchers began teaching in the control and experimental groups. In the control group, researchers taught selected lesson materials and also explained the vocabulary used to identify it. In addition, researchers also taught the pronunciation of the vocabulary provided. In the experimental group, the researchers also taught the same material, except that after the students created some sentences that identify and/or describe the animals, the students practiced to pronounce it using *Lyra* Virtual Assistant application. If the pronunciation spoken by the student is incorrect, the sentence or word that is written by LVA would be different. These treatments conducted in four meetings then, the researchers did post-test.

To analyze the data statistically, first the scores of the subjects on the post-test were summarized as descriptive statistics; namely, mean and standard deviations. Moreover, Kolmogorov-Smirnov Test was used to check the normality of the data. After the data was normal, then the researchers checked the homogeneity. After the data was homogeneous, researchers examined to see if the difference between the two groups was statistically meaningful or not, and the most appropriate test is inferential statistics; namely, T-test, which required the assumption of normal distributions. To run the tests, the researchers made use of the famous software package IBM SPSS Statistics ver. 22.

### 3. Results and Discussion

#### 4.1 Results of the pre-test

To start the data analysis, first, the Kolmogorov-Smirnov test was run to determine whether the data distribution was normal or not. The results showed that the significance level observed for the experimental group and control group was 0.110 and 0.508 respectively, both higher than the  $p$ -value 0.05. Thus, it was concluded that the data distribution in classes was normal. Table 1 presents the results of the Kolmogorov-Smirnov test:

**Table 1.** *The result of Normality Test*

		Experimental	Control
N		32	31
Normal Parameters <sup>a,b</sup>	Mean	54,1659	54,2335
	Std. Deviation	9,39229	11,92530
Differences	Absolute	,213	,148
	Positive	,085	,096
	Negative	-,213	-,148
Kolmogorov-Smirnov Z		1,204	,823
Asymp. Sig. (2-tailed)		,110	,508

- a. Test distribution is Normal.
- b. Calculated from data.

Next, the pre-test of both groups is calculated with a homogeneity test. It is done to determine the homogeneity of the sample, acquired by comparing the value based on trimmed mean which is calculated by Levene Formula. The result on the pre-test of both groups was 0.478, meaning the data was homogenous because Sig. is higher than the p-value 0.05. Table 2 presents the results of the Homogeneity test:

**Table 2.** The result of Homogeneity Test

	Levene Statistic	df1	df2	Sig.
Based on Mean	,531	1	61	,469
Based on Median	,572	1	61	,452
PreTest Based on Median and with adjusted df	,572	1	56,850	,453
Based on trimmed mean	,510	1	61	,478

Because the data distribution of pre-test was normal and homogenous, the researchers used a t-test to determine if the difference was significant. If the data value of Asymp. Sig. (2-tailed) > 0.05, the difference was not significant. The result of t-test on pre-test was 0.980, meaning the difference of data on pre-test was not significant.

**Table 3.** Result of T-test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	,531	,469	-,025	61	,980	-,06761	2,69980	-	5,33097
PreTest Equal variances not assumed			-,025	56,977	,980	-,06761	2,71002	-	5,35916

**4.2 Results of the post-test**

After acquiring the results of pre-test, the researchers used the Kolmogorov-Smirnov test again to determine the data distribution of the post-test data. The result of using normality test on the post-test data was 0.448 for the experimental group and 0.291 for the control group. It showed that the significance level observed for both groups was higher than the p-value 0.05. Thus, it was concluded that the data distribution in classes was normal. Table 4 presents the results of the Kolmogorov-Smirnov test:

**Table4.***The result of Normality Test*

		Experimental	Control
N		32	30
Normal Parameters <sup>a,b</sup>	Mean	69,5969	63,6110
	Std. Deviation	7,86408	5,57193
	Absolute	,152	,179
Most Extreme Differences	Positive	,152	,179
	Negative	-,125	-,154
Kolmogorov-Smirnov Z		,861	,981
Asymp. Sig. (2-tailed)		,448	,291

- a. Test distribution is Normal.
- b. Calculated from data.

After data was normal, the researchers used homogeneity test. The result of post-test of both groups was 0.257, meaning the data was homogenous because of Sig. > p (0.05). Table 5 presents the results of the Homogeneity test:

**Table5.***The result of Homogeneity Test*

		Levene Statistic	df1	df2	Sig.
	Based on Mean	1,304	1	60	,258
	Based on Median	1,240	1	60	,270
PostTest	Based on Median and with adjusted df	1,240	1	52,683	,271
	Based on trimmed mean	1,308	1	60	,257

Based on tables5 and 6, the data distribution was normal and homogenous. So, the researchers used a t-test to know if the difference is significant between the experimental group and control group or not. The Asymp. Sig. (2-tailed) of t-test was 0.001. It meant the Asymp. Sig. (2-tailed) < 0.05. The conclusion is there was a significant difference between both groups.

**Table6.***Result of T-test: Independent Samples Test*

	Levene's Test for Equality of Variances	t-test for Equality of Means								
		F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Equal variances assumed	1,304	,258	3,437	60	,001	5,98588	1,74147	2,50241	9,46934	
PostTest Equal variances not assumed			3,475	55,942	,001	5,98588	1,72264	2,53493	9,43682	

After acquiring the analysis results of the pre-test and post-test, the researchers used N-Gain to determine the effect size of the treatment that is given to the experimental group. The result was N-Gain average value of experimental group was 0.33. It can be seen as follow.

**Table7.** N-Gain Score of Experimental Class

Test	N	N-Gain			Average N-Gain
		Score Ideal	Min. Score	Max. Score	
Pre-Test	32	100	33.33	70.83	0.33
Post-Test	32	100	50.00	89.58	

Based on the table above, the average of N-Gain is 0.33, meaning the sign of strategy that used in the experimental class is considered Middle. In the criteria of achievement, N-Gain score can be said Middle if score  $0.3 < g < 0.7$ . It can be concluded that using Lyra Virtual Assistant on students' speaking performance at the seventh grade of SMP TB Pekanbaru has a positive effect.

#### 4. Conclusion

The purpose of this research is to find out the effectiveness of *Lyra* Virtual Assistant application on improving students' speaking performance at the seventh grade of SMP TB Pekanbaru. The research question: "Is there any significant difference on students' speaking performance that uses *Lyra* Virtual Assistant than those who use other conventional methods in practicing the speaking skill at the seventh grade of SMP TB Pekanbaru?" has been successfully answered, based on the average of post-test experimental and control group analysis results. Experimental group's average post-test result is 69.59 and control group's average post-test result is 63.61. In conclusion, students taught by using *Lyra* Virtual Assistant in speaking achieve better than students who were taught without using *Lyra* Virtual Assistant in practicing speaking. *Lyra* Virtual Assistant is effective in practicing speaking. *Lyra* Virtual Assistant contributes to students' achievement in speaking skill.

#### References

- [1] Abad, J. V. (2013). Pedagogical Factors that Influence EFL Teaching: Some Considerations for Teachers' Professional Development. Profile Issues in Teachers' Professional Development.
- [2] Aditya, D. S. (2016). "Speak App" Android Application Model to Improve the Midwifery Students' Speaking Competence. Doctoral Dissertation Santa Dharma University.
- [3] Allen, J. B. (1994). How Do Humans Process and Recognize Speech? IEEE Transactions on Speech and Audio Processing. <https://doi.org/10.1109/89.326615>
- [4] Allerton, D. J. (1991). The greater precision of spoken language: Four examples from English. English Studies. <https://doi.org/10.1080/00138389108598774>
- [5] Aust, R., Kelley, M. J., & Roby, W. (1993). The use of hyper-reference and conventional dictionaries. Educational Technology Research and Development. <https://doi.org/10.1007/BF02297512>
- [6] Bajcsy, R. (2002). Visions 2020: Transforming education and training through advanced technologies. Washington, DC: US Department of Commerce. Retrieved March.
- [7] Brett, P. (1998). Using multimedia: A descriptive investigation of incidental language learning. International Journal of Phytoremediation. <https://doi.org/10.1076/call.11.2.179.5684>
- [8] Budzinski, O., Noskova, V., & xijie, zhang. (2019). The Brave New World of Digital Personal Assistants: Benefits and Challenges from an Economic Perspective. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3306169>
- [9] Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics. <https://doi.org/10.1093/applin/I.1.1>
- [10] Castillo, C. Y. P. (2009). Improving Eleventh Graders' Oral Production in English Class through Cooperative Learning Strategies. PROFILE Issues in Teachers' Professional Development.
- [11] Cavus, N., & Ibrahim, D. (2009). M-Learning: An experiment in using SMS to support learning new English language words. British Journal of Educational Technology. <https://doi.org/10.1111/j.1467-8535.2007.00801.x>
- [12] Chomsky, N., & Halle, M. (1965). Some controversial questions in phonological theory. Journal of Linguistics. <https://doi.org/10.1017/S0022226700001134>
- [13] Davis, J. N., & Lyman-Hager, M. A. (1997). Computers and L2 Reading: Student Performance, Student Attitudes 1. Foreign Language Annals. <https://doi.org/10.1111/j.1944-9720.1997.tb01317.x>
- [14] Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with

- cellphones, smartphones & social media. *Internet and Higher Education*. <https://doi.org/10.1016/j.iheduc.2013.06.002>
- [15] Gric-coast, M. B., Lyon, U. L., & Cedex, B. (2000). The roles of models in Artificial Intelligence and Education research : a prospective view. *International Journal of Artificial Intelligence in Education*.
- [16] Hamuddin, B. (2016). Using blog to Promote English Skills for EFL Students: The Students' Perception. *ELT-Lectura*. <https://doi.org/10.31849/elt-lectura.v3i2.481>
- [17] Hamuddin, B., & Dahler. (2018). Blogs as Powerful Learning Tools: The Perception from EFL Students in Riau Main Island Indonesia. *IOP Conference Series: Earth and Environmental Science*. <https://doi.org/10.1088/1755-1315/156/1/012060>
- [18] Hamuddin, B., Kurniawan, K., Syaifullah, S., & Herdi, H. (2018). Detecting Major Problems in Learning English Through Blog-based Class. *Journal of Education and Learning (EduLearn)*, 12(3), 529. <https://doi.org/10.11591/edulearn.v12i3.7766>
- [19] Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2015.02.026>
- [20] Keshet, J., & Bengio, S. (2009). Automatic Speech and Speaker Recognition: Large Margin and Kernel Methods. In *Automatic Speech and Speaker Recognition: Large Margin and Kernel Methods*. <https://doi.org/10.1002/9780470742044>
- [21] Kukulka-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*. <https://doi.org/10.1017/S0958344008000335>
- [22] Lewis Johnson, W., & Valente, A. (2009). Tactical language and culture training systems: Using AI to teach foreign languages and cultures. *AI Magazine*.
- [23] Li, J., Deng, L., Haeb-Umbach, R., & Gong, Y. (2016). Fundamentals of speech recognition. In *Robust Automatic Speech Recognition*. <https://doi.org/10.1016/b978-0-12-802398-3.00002-7>
- [24] Maftoon, P., Hamidi, H., & Sarem, S. N. (2012). The Effects of CALL on Vocabulary Learning: A Case of Iranian Intermediate EFL Learners. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*.
- [25] Navarro, O., Gonzalez, A. L., & Molina, A. I. (2018). Experience of use of eye tracking technology with children who have attention problems. *SIIE 2018 - 2018 International Symposium on Computers in Education, Proceedings*. <https://doi.org/10.1109/SIIE.2018.8586721>
- [26] Nordrum, A. (2017). CES 2017: The Year of Voice Recognition - *IEEE Spectrum*. *IEEE Spectrum*. Retrieved from <http://spectrum.ieee.org/tech-talk/consumer-electronics/gadgets/ces-2017-the-year-of-voice-recognition>
- [27] Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting Visual and Verbal Learning Preferences in a Second-Language Multimedia Learning Environment. *Journal of Educational Psychology*. <https://doi.org/10.1037/0022-0663.90.1.25>
- [28] Pokulevska, A. I. (2018). Experience of Using Skype as a Means of Foreign Languages Teaching Efficiency Improving. *Information Technologies and Learning Tools*. <https://doi.org/10.33407/itlt.v6i8i6.2150>
- [29] Richardson, J., & Lenarcic, J. (2008). Text messaging as a catalyst for mobile student administration: The "Trigger" experience. *Australian Journal of Emerging Technologies and Society*.
- [30] Russell, S., & Norvig, P. (2016). *Artificial Intelligence A Modern Approach Third Edition*. In Pearson. <https://doi.org/10.1017/S0269888900007724>
- [31] Shih, Y. E., & Mills, D. (2007). Setting the new standard with mobile computing in online learning. *International Review of Research in Open and Distance Learning*.
- [32] Solano, L., Cabrera, P., Ulehlova, E., & Espinoza, V. (2017). Exploring the use of educational technology in efl teaching: A case study of primary education in the south region of ecuador. *Teaching English with Technology*.
- [33] Underwood, J. (2017). Exploring AI language assistants with primary EFL students. In *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017*. <https://doi.org/10.14705/rpnet.2017.eurocall2017.733>
- [34] Wekke, I. S., Yandra, A., & Hamuddin, B. (2017). Learning Strategy in Class Management: A Reflection from Manado Case. *IOP Conference Series: Earth and Environmental Science*. <https://doi.org/10.1088/1755-1315/97/1/012053>
- [35] Wiranti, N. A. (2013). Improving students' speaking competence by using jazz chants. *Doctoral Dissertation Universitas Sebelas Maret*.
- [36] Yamamoto, T., Liao, A., Wu, W. C. V., Shih, M., Shih, J. L., & Chu, H. C. (2018). A Proposal for the Global and Collaborative PBL Learning Environment Where All Global Members on Different Campuses Are "On the Same Page" throughout the Process of Learning in the Project. *Proceedings - 2018 Conference on Technologies and Applications of Artificial Intelligence, TAAI 2018*. <https://doi.org/10.1109/TAAI.2018.00029>