Volume: 25 No.1. April, 2025; pp. 110-129
P-ISSN: 1411-4585 E-ISSN: 2549-6743
DOI: https://doi.org/10.24036/pedagogi.v25i1.2477
Submitted: 25-04-03; Rivised: 25-04-09; Accepted: 25-04-30

Pedagogi Jurnal Ilmu Pendidikan
http://pedagogi.ppj.unp.ac.id/

The Effectiveness Of *Elsa Speak* Application To Improve Speaking Skill In Extra Class

Yuyun Bahtiar[1*], Lailatus Sa'diyah[2], Ulfa Wulan Agustina[3], Luluk Choirun Nisak Nur[4], Iin Baroroh Ma'arif[5]
[1,2,3,4,5] KH. A. Wahab Hasbullah University
* e-mail: yuyunbahtiar@unwaha.ac.id

Abstract
This study aims to test the effectiveness of the Elsa Speak application as a learning medium in improving English speaking skills of additional class students at Al-Bairuny Islamic School. The researcher used a quasi-experimental design with two groups: an experimental group taught using the AI-based Elsa Speak application and a control group taught without the application. The results showed a significant increase in the experimental group's speaking skills, especially in pronunciation, fluency, and confidence. The control group only showed minimal improvement. These findings indicate that Elsa Speak, as an AI-based learning medium, is effective in improving students' English speaking skills. This study shows the importance of using AI-based technology in language learning to improve the quality of English education in the digital era.

Keywords: *Elsa Speak application, Learning media. Speaking skills*

How to cite : Bahtiar, Y., Sa'diyah, L., Agustina, U., Nur, L. C., & Ma'arif, I. (2025). The Effectiveness Of Elsa Speak Application To Improve Speaking Skill In Extra Class. *Pedagogi: Jurnal Ilmu Pendidikan*, *25*(1). https://doi.org/https://doi.org/10.24036/pedagogi.v25i1.2477

INTRODUCTION

Humans now live in the 4.0 Industrial Revolution, also known as the Disruptive Era, in which technological breakthroughs, as well as information and communication, have brought about major changes in people's lifestyles. A disruptive period is one in which anything can happen at random. Positive and negative things can go viral in seconds. This is an era in which everyone can do almost anything in the virtual world. John McCarthy created the term Artificial Intelligence in his research proposal in 1955, which characterizes the workings of Artificial Intelligence systems as a machine that employs language in abstract forms and notions that can help humans solve their problems (Kalalo & Pontoh, 2020).

Artificial Intelligence is a field of computer science that seeks to create systems. A machine needs human intelligence to be able to carry out its routine tasks. Several key ideas are involved in Artificial Intelligence, including machine learning, neural networks, natural language processing, and more. Artificial Intelligence has significantly impacted a number of industries, including voice recognition, facial identification, autonomous vehicles, healthcare, and many more (Eriana & Zein, 2023).

A branch of computer science with significant science fiction parallels is Artificial Intelligence. Engaging with the incredibly intelligent behavior, flexibility, and learning of a machine. This includes the following: arranging, supervising, planning, answering questions from customers, and understanding handwriting, speech, and cornea (Jamaaluddin & Indah, 2021).

Artificial Intelligence in Education (AIED) has been around for about four decades. This field uses and continues to employ approaches from artificial intelligence and cognitive science to attempt to understand the nature of learning and teaching in order to design systems that assist learners in mastering new skills or understanding new concepts (Du Boulay, 2016).

Artificial Intelligence technology is developing rapidly. A lot of Artificial Intelligence is currently being developed in the form of apps (Android and desktop), websites, and many more. Artificial Intelligence in the form of apps or websites can be used anytime and anywhere. Artificial Intelligence in the form of apps or websites can be used anytime and anywhere. By using Artificial Intelligence, human work is completed faster. Many application stores now provide Artificial Intelligence-based solutions. Artificial Intelligence applications in science, mathematics, language, and other subjects have been widely developed, with separate categories for each. In the realm of language, Artificial Intelligence in the form of applications and websites has become increasingly employed in recent years to aid language learning. Translation, grammar, speaking, and listening are examples of Artificial Intelligence applications that are commonly employed in language acquisition.

Mastering a foreign or second language necessitates proficiency in speaking, which is considered the most crucial skill. Out of the four fundamental language skills in English, speaking is often perceived as challenging. The task of generating sentences spontaneously poses a challenge for speakers. For individuals learning a foreign or second language, it is particularly difficult to construct sentences without a solid grasp of grammar and an extensive vocabulary. Consequently, English language learners in EFL/ESL encounter numerous obstacles when it comes to speaking grammatically correct sentences. Given that effective communication heavily relies on speaking abilities, individuals strive to acquire and refine these skills to enhance their communication prowess (Rao, 2019).

Speaking is an invaluable ability that should be practiced as often as possible, particularly by English language students. To develop an English-speaking environment for students, effective teachers, and learning material must encourage their capacity to speak English. Artificial Intelligence *Elsa Speak* is one of the learning tools that can help improve English speaking skills. Artificial Intelligence takes the shape of an Android application that includes features to help students learn English, such as pronunciation, conversation, and topic-based learning.

The researchers have reviewed previous studies and identified shortcomings in those studies to be addressed in this research. The researcher by Sholekhah relies on library analysis without direct empirical data, which can limit the validity of the results. No field tests are conducted to measure the actual impact of the application on students. The second study by Rinaepi, the researcher has limitations, such as a small sample size (11 students), no explanation of sample selection, and being conducted in only one school. There is no analysis of other factors or challenges in using the application, and the duration of the experiment is not explained. Further research with a larger sample is needed. The third researcher by Mr. Kholis. The researcher has shortcomings, such as a small sample size (18 students) and being limited to one university. This research only focuses on using the ASR feature.

Based on an interview with Mr. Yulianto Mugiono, one of the Extra English class mentors at Al-Bairuny Islamic School Sambongdukuh on April 04, 2024, the English Extra class focuses on speaking abilities. This Extra class separates one generation into multiple groups based on the students' speaking abilities. Currently, the learning materials used in Extra English classes at Al-Bairuny Islamic School are student books created by mentors. The application of AI-based learning materials is still quite limited. Since Al-Bairuny's Islamic School Extra English class was launched in 2017 Media like as Google Translate Speaking are only used a few times and simply

help mentors learn vocabulary terms. Mentors exclusively use learning media, student books, and lecture methods to teach. The teaching and learning process is boring, students become bored and struggle to comprehend the material. Weak media and learning approaches lead to poor levels of progress in students' speaking skills. The conclusion is that Al-Bairuny's Islamic School Extra English class is the appropriate demographic to serve as a sample for the experiment using the *Elsa Speak* Application as a spoken learning medium, as Al-Bairuny's Islamic School Extra English class has never used Artificial Intelligence-based learning media. Based on the background above, the researchers intend to investigate "The Effectiveness of *Elsa Speak* Application to Improve Speaking Skill in Extra Class Al-Bairuny Islamic School Students".

METHODS

The researchers used quasi-experimental quantitative methods in this research. A quasi-experiment was one in which the smallest experimental units were not randomly allocated to experimental and control groups (Hastjarjo, 2019). Experimental research aimed to determine the cause-and-effect relationship between two variables. Research experiments contained both independent and dependent variables. The independent variable was one that was modified to see if there was a significant impact on the dependent variable. Dependent variables, often known as control variables, were variables that were affected by independent variables. An independent variable was one that was purposely changed to see if it influenced the dependent variable.

The experimental design was one of the most relevant research approaches for conducting language studies in education. It was envisaged that this design would lead to fresh discoveries and would be implemented by language educators in the future. Experimental research was a major strategy in second language studies. This research was commonly used by acquisition researchers of second languages, linguists, and education researchers (Ratminingsih, 2010). Researchers applied a quasi-experimental design because this type did not have strict limitations regarding randomization (Abraham & Supriyati, 2022). This research was used when the researcher could only determine the sample by directly choosing two classes as the control class and the experimental class

The types of design included in this category were (1) The Nonequivalent Control Group Design, (2) The Time-Series Design, and (3) Counterbalanced Design (Ratminingsih, 2010). The researchers chose The Nonequivalent Control Group Design. The researchers took this design because the Pretest-Posttest Control Group Design included a pre-test step to assess the student's skill level before applying treatment. The goal of this researcher's design would be to collect more accurate information by comparing the condition of the dependent variable in the experimental group after treatment to the control variable without treatment. Based on (Sugiyono, 2020) the following was a table of The Nonequivalent Control Group Design, which had been modified by the researchers to make it easier to understand:

Table 1. The Nonequivalent Control Group Design

Group	Pretest	Treatment	Posttest
Experiment	O1	X	O2
Control	O1	-	O2

Description:
O1 = Pretest (test before treatment)
O2 = Posttest (test after treatment)
X = Treatment (using the *Elsa Speak* application as a learning media)

A. Variables

This study included two variables the independent variable and the dependent variable. The independent variable was the *Elsa Speak* application, and the dependent variable was speaking skills.

B. Population and Sample

The term population referred to all aspects of research, including items and individuals, that had specific characteristics. So, in theory, a population included all members of a group of humans, animals, events, or things who lived together in a certain location with the intention of drawing conclusions from the study's ultimate findings (Amin et al., 2023). The population for this study was the XI students' Extra English class at Al-Bairuny Islamic School Sambongdukuh Jombang, which comprised 78 students organized into six groups, each with approximately 15 students.

A sample was a subset of the population that serves as the data source for a study. A sample was a subset of a population chosen to represent the population as a whole (Amin et al., 2023). Based on the explanation above, the sample was taken using the purposive sampling technique. The purposive sampling technique, also known as judgment sampling, involves the deliberate selection of participants based on the specific qualities they possess. It was a non-random technique that did not require underlying theories or a fixed number of participants (Etikan, 2016). The samples taken in this study used extra English class group A as an experimental group to receive treatment, while extra English class group b served as a control group.

C. Instrument

Instruments were tools used to collect data in research. This research employs the following instruments:

1. Test questions

Speaking was a practical skill, therefore speaking tests should also be in the form of practice. This research instrument used an oral pre-test and post-test.

a. Pre-test: The researchers gave a topic, and students talked about the topic in front of the class. The pre-test topic was about "school."
b. Post-test: The researchers gave a topic, and students talked about the topic in front of the class. The pre-test topic was about "school."

2. Treatment

The researchers treated the experimental group using the *Elsa Speak* application downloaded on a smartphone as a learning media.

Data Collection

1. Pre-test

The pre-test was given to both groups (experimental and control) during the first week. The Students select random rolls of paper with questions about the theme "School." The students then answered the questions they chose in front of the class. The researchers gave students two minutes to prepare their answers. The researchers assessed students' tests based on pronunciation, grammar, vocabulary, fluency, and comprehension. The pre-test aimed to identify the student's level of ability before receiving treatment.

2. Treatment

The researcher treated the experimental group. The experimental group was taught speaking in class with a special treatment, using the *Elsa Speak* application as a learning media. The researchers taught aspects of speaking such as vocabulary, grammar, and pronunciation.

3. Post-test

A post-test was given to both groups (experimental and control) following therapy. The Students selected rolls of paper at random that included questions about the theme "School." The Students then answered the questions they selected in front of the class. The researchers allowed students two minutes to prepare their answers. The researchers assessed students' tests based on

pronunciation, grammar, vocabulary, fluency, and comprehension. The post-test attempts to assess the student's level of competence after treatment.

D. Data Analysis

1. Validity Test

The validity test of an instrument referred to the process of determining whether the instrument accurately measured what it was intended to measure. It ensured that the results obtained from the instrument reflected the true characteristics or concepts being studied, thereby ensuring the reliability of the data collected. The material expert test question validator gave a score for each item with answers very suitable (5), suitable (4), quite suitable (3), not suitable (2), and not suitable (1) (Riyani et al., 2017). Then, the total score for each validator was added up and the average validity was calculated using the formula:

$$P = \frac{Total\ indicator\ scores}{The\ maximum\ score\ of\ indicators} \text{ X } 100\%$$

Table 2. Validity Level Qualification Table

Validity Level Qualification Table	
Presentation	Response rate
Very Valid	81%-100%
Valid	61%-80%
Fairly valid	41%-60%
Invalid	21%-40%
Very invalid	>21%

1. Analysis of Test Results

The data analysis compared the experimental and control groups to determine the student's speaking abilities. The researchers assessed pronunciation, grammar, vocabulary, fluency, and comprehension. The five aspects of the speaking assessment were analyzed *using IBM SPSS Statistics 22*. Descriptive analysis was a statistical method used to describe or summarize the basic characteristics of a data set. The goal was to present information that provided a clear understanding of the data. This technique involved the use of statistical measures such as mean, median, mode, range, variance, and standard deviation to describe patterns or trends in data. Based on H. Douglas Brown & Priyanvada Abeywickrama (2004) the following was an analysis of the assessment rubric:

Table 3. The rubric for analyzing speaking

Assessment Aspect		
Aspect	score	Description
Pronunciation	20	Easy to understand, with a native speaker's accent.
	15	Easy to understand with certain accents.
	10	There were pronunciation problems that needed the listener's full concentration, and there were sometimes misunderstandings.
	5	Pronunciation issues made it difficult to comprehend, frequently requiring several tries.
	1	significant pronunciation problems, therefore it's not well understood.
Assessment Aspect		
Aspect	Score	Description
	20	There were few or no grammatical errors.
	15	Occasionally made grammatical errors but did not alter the meaning.

Grammar	10	Frequently made linguistic errors that influenced meaning.
	5	Many grammatical faults disrupt meaning and frequently change phrases.
	1	Grammatical errors were significant and difficult to explain.
Assessment Aspect		
Aspect	Score	Description
Vocabulary	20	Uses native-style vocabulary and expressions.
	15	Occasionally used bad vocabulary.
	10	Conversations became constrained due to the use of incorrect vocabulary.
	5	Using vocabulary badly and vocabulary was limited, making it harder to understand.
	1	The vocabulary was so limited that discussion was difficult.
Assessment Aspect		
Aspect	Score	Description
Fluency	20	Fluent like a native speaker.
	15	Fluency appeared to be minimally affected by linguistic difficulties.
	10	Fluency was slightly affected by linguistic issues.
	5	Frequently was wary and stopped caused of language limitations.
	1	The conversation was disorganized and halted, which made the conversation unfeasible.
Assessment Aspect		
Aspect	Score	Description
Comprehension	20	Understand everything without struggle.
	15	Understands almost any topic, with some repetition.
	10	Understands most of what is stated when speech was slowed down, even if there was repetition.
	5	Difficult to understand what was being stated.
	1	Could not understand even simple conversations.

Based on the results of an interview conducted by the researchers with one of the Extra English class tutors at Al-Bairuny Islamic School, the minimum speaking score that students were to obtain was 70.

2. Normality Test

The normality test was one type of classical assumption test aimed at determining the distribution of data within a group or population. There were two categories of data distribution: normal distribution and non-normal distribution (Widodo et al., 2023). This research used IBM SPSS Statistics 22 to test normality with the Shapiro-Wilk normality test. The goodness of fit

test assessed the level of conformity between the sample distribution (observation scores) and the theoretical distribution. The Shapiro-Wilk normality test criteria used the significance value (Sig.) or probability value to determine whether a distribution was normal. A value greater than 0.05 suggested a normal distribution, while a value less than 0.05 suggests a non-normal distribution. Nonparametric statistics were employed when the data was not normally distributed. The t-test was used to compare regularly distributed data sets. Based on Nasrum (2018), the Shapiro-Wilk normality test follows this formula:

$$W = \frac{b^2}{S^2} = \frac{(\sum_{i=1}^{n} a_i\, y_i)^n}{\left(\sum_{i=1}^{n}(y_i - \bar{y}\right)^n}$$

3. Homogeneity Test

The equality of two variances test determined whether the data distribution was homogeneous by comparing the variances. The data homogeneity test aimed to identify whether the variation between the experimental and control groups was homogeneous. The researchers used the Cochran test to assess homogeneity. If one group showed more variability than another, it could indicate a lack of homogeneity. Based on Usmadi (2020), the formula for the Cochran Test was:

$$C\,count = \frac{largest\ variance}{total\ variance}$$

The testing criteria were to compare the results of the Cochran formula with the Cochran Accept table *Ho* if $C\,count$ < 0,05 and rejected *Ho* if $C\,count$ > 0,05. This research used *IBM SPSS Statistics 22*.

4. Hypothesis Test

This research analyzed data using the Independent Samples T-test. Independent samples T-test was used to test whether there was a difference between the two independent samples (Tae Kyun Kim, 2018). Independent Sample T-Test It was used to test whether there was a significant difference between the two independent groups. This test compared the means of two groups that were not related to each other to determine if there was a sufficiently large difference that could be explained by a specific factor being studied. This research used *IBM SPSS Statistics 22*. Based on (Anwar, 2009) the formula Independent Sample T-Test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{S_1}^2}{n_1} + \frac{{S_2}^2}{n_2}}}$$

Criteria testing was as follows:

By comparing the calculated t-value with the t-table value under the following conditions:

Ho was accepted t-value < t-table

Ho was rejected t-value ≥ t-tabel

By using probability values, with the following conditions:

Ho was accepted if Probability > significance level (α).

Ho was rejected if Probability ≤ significance level (α).

FINDING AND DISCUSSIONS

The findings aimed to provide insights into how the use of the *Elsa Speak* application influenced the speaking skills of students in the extra class program. Through the analysis of the collected data, this section highlighted the extent to which the application improved pronunciation, fluency, and overall speaking competence. The results were presented

systematically to provide a clear understanding of the application's impact and its potential as a learning tool for language development.

1. Validity Test

a. AI *Elsa Speak* Application Speaking Material

The *Elsa Speak* Application was the instrument used to analyze students' speaking abilities in this research. The application evaluated students' pronunciation and intonation in real-time using Artificial Intelligence technology. *Elsa Speak*'s direct feedback tool offered both recommendations for improvement and an accurate evaluation of the degree of pronunciation problems. With minor modifications and recommendations from the validators, the *Elsa Speak* Application's learning media had already undergone validation by two validators. The results of the question validation for the test can be seen in Table 4.:

Table 4. Validation Results of the Elsa Speak Application

No	Validated Aspect	Scoring	
		V1	V2
1	App Appearance	5	5
2	Button use	5	5
3	Text type and size	4	5
4	Quality photos, images, and graphics	4	5
5	Audio and video quality	5	5
6	Ease of understanding language	5	5
7	The applications used were in accordance with student learning needs	5	5
8	The application used could meet 3 types of student learning (audio, visual, and kinesthetic)	4	5
9	The application used can motivate students to learn to speak	5	5
10	The selected material could be applied to speaking learning	5	5
11	The selected material could be used to improve students' speaking skills	5	5
Total		52	55
Percentage		94%	100%
Average		97%	
Description		Very valid	

Description:
V1 = Validator 1
V2 = Validator 2

Two validators conducted a material validation examination on the *Elsa Speak* Application, resulting in a 97% review percentage from all aspects. Based on the outcomes, the AI *Elsa Speak* Application was considered highly valid and appropriate for this research.

b. Pre-test Speaking English

The pre-test questions were the instrument used to assess students' abilities before the treatment. The pre-test findings offered a summary of the starting level of comprehension and could be consulted for examining post-treatment changes. The researchers could compare the differences between the initial scores and final outcomes (post-test) by employing a pre-test. This pre-test used a "blind test" question model, where students randomly selected two questions. These pre-test questions had been validated by two experts with minor modifications and suggestions from the validators. The validation results of the pre-test questions can be seen in the table below:

Table 5. Validation Results Pre-test Speaking Questions

No	Validated Aspect	Scoring	
		V1	V2
1	Tell me about your school and what class you are in?	5	5
2	Tell me something about your favorite subjects at school.	5	5
3	Tell me about your favorite place at school.	5	5
4	Tell me about your best friend at school and what you and your best friend like.	4	5
5	Tell about your favorite teacher and why you like him/her.	5	5
	Total	24	25
	Percentage	96%	100%
	Average	98%	
	Description	Very valid	

Description:
V1 = Validator 1
V2 = Validator 2

Based on the results of the validation analysis of the pre-test questions carried out by two validators, the percentage was 98%, covering all aspects. Based on the analysis results, it was determined that the pre-test questions were very valid and suitable for use in the research.

c. Post-test Speaking English

Post-test questions were an instrument used to evaluate students' skills following the treatment. To evaluate the efficacy of the intervention, the post-test results were compared to the pre-test findings. The post-test provided the researchers with information about how well the research goals had been met and helped in determining whether the intended changes had occurred. This post-test used a "blind test" question model, where students randomly selected two questions. These post-test questions were validated by two experts with minor modifications and suggestions from the validators. The validation results of the post-test questions can be seen in the table below:

Table 6. Validation Results Post-test Speaking Questions

No	Validated Aspect	Scoring	
		V1	V2
1	Tell me about the activities you like at school, such as extracurricular activities.	5	5

2	Is there a student event at your school? If there is, tell one of them.	4	5
3	Tell me about the subjects you don't like and tell me the reason.	5	5
4	Do you like bringing lunch to school or buying food at the canteen?	4	5
5	Tell about your friend in class who is the quietest and the most talkative.	5	5
Total		23	25
Percentage		92%	100%
Average		96%	
Description		Very valid	

Description:
V1 = Validator 1
V2 = Validator 2

Based on the results of the validation analysis of the post-test questions carried out by two validators, the percentage was 96%, covering all aspects. Based on the analysis results, it was determined that the post-test questions were very valid and suitable for use in the research.

2. Analysis of Test Results

Data analysis compared the experimental and control groups to determine students' speaking abilities. The researchers assessed pronunciation, grammar, vocabulary, fluency, and comprehension based on the speaking assessment rubric. The table below presents the results of the pre-test and post-test assessments for the experimental group and the control group:

a. Experiment Group Pre-test Results

The results of the experimental group students' speaking scores on the pre-test were as follows:

Table 7. Experiment Group Pre-test Results

No	Student initial	P	G	V	F	C	Total
1	AS	10	10	15	10	10	55
2	ALN	15	10	15	10	15	65
3	ARLM	10	10	15	10	15	60
4	AAN	15	10	15	10	15	65
5	DNS	10	10	15	10	10	55
6	EPSA	15	10	10	15	10	60
7	FAN	10	10	15	10	10	55
8	F	10	10	10	10	10	50
9	KDP	15	10	15	10	10	60
10	LRI	10	10	10	10	10	50
11	NA	15	10	15	10	15	65
12	SJP	10	10	15	10	10	55
13	SDR	10	10	10	10	10	50
14	SDN	15	10	10	10	15	60
15	VDS	10	10	10	10	15	55

b. Experiment Group Post-test Results

The results of the experiment group students' speaking scores on the post-test were as follows:

Table 8. Experiment Group Post-test Results

No	Student initial	P	G	V	F	C	Total
1	AS	20	15	15	10	15	75
2	ALN	20	15	15	15	15	80
3	ARLM	15	10	15	15	15	70
4	AAN	15	10	15	15	20	75
5	DNS	20	10	15	10	15	70
6	EPSA	15	10	15	10	20	70
7	FAN	15	10	15	10	15	65
8	F	15	10	15	10	10	60
9	KDP	15	15	15	10	20	75
10	LRI	15	10	15	10	15	65
11	NA	20	10	15	15	15	75
12	SJP	15	15	15	10	15	70
13	SDR	15	10	15	10	10	60
14	SDN	20	10	15	15	15	75
15	VDS	15	10	15	10	15	65

c. Control Group Pre-test Results

The results of the control group students' speaking scores on the pre-test were as follows:

Table 9. Control Group Pre-test Results

No	Student initial	P	G	V	F	C	Total
1	ADWR	15	10	15	10	10	60
2	A	10	10	15	10	15	60
3	DMM	15	15	10	10	15	65
4	FAA	10	10	15	10	10	55
5	FZH	15	15	15	10	15	70
6	FA	10	15	15	10	15	65
7	MAH	15	10	15	10	10	60
8	MDF	10	10	10	10	10	50
9	MF	10	10	10	10	10	50
10	NR	10	10	15	10	10	55
11	RFH	15	10	15	10	15	65
12	RR	15	15	15	10	15	70
13	TNS	15	10	15	10	20	70
14	ZAM	15	10	15	10	15	65
15	ZZ	10	15	15	10	15	65

d. Control Group Post-test Results

The results of the control group students' speaking scores on the post-test were as follows:

Table 10. Control Group Post-test Results

No	Student initial	P	G	V	F	C	Total

1	ADWR	15	10	15	10	15	65
2	A	10	10	15	10	15	60
3	DMM	15	15	10	10	15	65
4	FAA	10	10	15	10	15	60
5	FZH	15	15	15	10	15	70
6	FA	10	15	15	10	15	65
7	MAH	15	15	15	10	10	65
8	MDF	10	10	10	10	10	50
9	MF	10	10	15	10	10	55
10	NR	10	10	15	10	10	55
11	RFH	15	10	15	10	15	65
12	RR	15	15	15	10	15	70
13	TNS	10	10	15	10	15	60
14	ZAM	10	10	15	10	15	60
15	ZZ	15	15	15	10	15	70

Description:
P = Pronunciation
G = Grammar
V = Vocabulary
F = Fluency
C = Comprehension

e. Analysis Descriptive

The researchers processed the data by calculating values such as mean, median, mode, and standard deviation to provide a general picture of the data distribution using descriptive analysis. The researchers used *IBM SPSS Statistics 22*. The following are the results of the descriptive analysis of the pre-test and post-test for the experimental and control groups:

Table 11. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-test Experiment	15	50.00	65.00	57.3333	5.30049
Post-test Experiment	15	60.00	80.00	70.0000	5.97614
Pre-test Control	15	50.00	70.00	61.6667	6.72593
Post-test Control	15	50.00	70.00	62.3333	5.93617
Valid N (listwise)	15				

According to the results of the descriptive analysis, the experimental group's pre-test mean score was 57.3, whereas the control group's pre-test score was 61.6. The data indicated that the average speaking score of the students was still below the required minimum of 70. The experimental group's post-test mean score was 70.0, while the control group's post-test mean score was 62.3. The data showed that students in the experimental group experienced significant growth, increasing from a mean of 57.3 to 70.0 after receiving treatment. In contrast, there was only a slight improvement in the control group, which did not receive treatment, with a change from 61.6 to 62.3.

3. Hypothesis Testing

a. Normality Test

The normality test was a statistical procedure used to determine whether the data was normally distributed or not. The requirement for the paired sample t-test was that the data had to be normally distributed. This research used *IBM SPSS Statistics 22* to test normality with the

Shapiro-Wilk normality test. The goodness-of-fit test assessed the level of conformity between the sample distribution (observed scores) and the theoretical distribution. The Shapiro-Wilk normality test used significance values (Sig.) or probability values to determine whether a distribution was normal. A value greater than 0.05 indicated a normal distribution, while a value less than 0.05 indicated a non-normal distribution. Nonparametric statistics were used when the data was not normally distributed. The t-test was used to compare normally distributed data sets. The results of the data normality test calculations were shown in the table below:

Table 12. Normality Test

		Shapiro-Wilk[a]
	Class	Sig.
Student Learning Outcomes	pre-test exp	.064
	post-test exp	.181
	pre-test cont	.093
	post-test cont	.175

The results of the normality test calculation with Shapiro-Wilk were as follows: The pre-test significance value for the experimental group was 0.064 > 0.05, and the post-test significance value for the experimental group was 0.181 > 0.05. In the control group, the pre-test significance value was 0.093 > 0.05, and the post-test significance value was 0.175 > 0.05. It can be concluded that the data from both the experimental group and the control group were normally distributed.

b. Homogeneity Test

The equality of two variances test was used to determine whether or not the data distribution was homogeneous, specifically by comparing the variances. The purpose of the data homogeneity test was to determine whether the variation between the experimental and control groups was homogeneous. The researchers used the Cochran test to determine homogeneity. The testing criteria were to compare the results of the Cochran formula with the Cochran acceptance table: H_0 was accepted if C count < 0.05 and H_0 was rejected if C count > 0.05. This research used *IBM SPSS Statistics 22*. The results of the data homogeneity test calculations were shown in the following table:

Table 13. Homogeneity Test

	Levene Statistic	df1	df2	Sig.
Student Learning Outcomes	Based on Mean .021	1	28	.886
	Based on Median .000	1	28	1.000
	Based on the Median and with adjusted df .000	1	26.664	1.000
	Based on trimmed mean .017	1	28	.897

The output above showed that the mean-based C count value was 0.886 > 0.05. As a result, H_1 was accepted and H_0 was rejected. It could be said that the experimental group's and the control group's post-test data variances were the same or homogeneous.

c. Hypothesis Test

Hypothesis testing using an Independent Sample T-test there was a difference between the means of two independent samples. This research aimed to determine the effect of the *Elsa Speak* Application learning media on improving the speaking skills of Extra-class students. The analysis used in this research was the independent Sample T-test with the help of *IBM SPSS Statistics 22*. This test was carried out to decide whether the hypothesis was rejected or accepted in the experiment group and the control group. The results of the data homogeneity test calculations were shown in the following table:

Table 14. Independent Sample T-test

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
					Lower
Outcomes	Equal variances assumed	.001	7.66667	2.17489	3.21160
	Equal variances not assumed	.001	7.66667	2.17489	3.21159

The Sig. (2-tailed) value in the t-test table indicated the result of the two-tailed significance test, with a value of 0.001. This value was used to determine whether there was a significant difference between the two groups being compared. Since the p-value 0.001 was smaller than the commonly used significance level of 0.05, the null hypothesis (H_0), which stated that there was no difference between the two groups, was rejected. Therefore, it could be concluded that there was a statistically significant difference between the two groups being tested indicating that the experimental group, which used the *Elsa Speak* application, showed a greater improvement in speaking skills compared to the control group, which used a textbook.

4. Researchers step

The researchers conducted the study in multiple stages beginning in March 2024. These stages included choosing a research title, preparing the research design, preparing research instruments, holding a research proposal seminar, validating the instruments, and implementing the research at Al-Bairuny Islamic School's Extra English class. Fifteen students from the experimental group and fifteen students from the control group participated in the study, which focused on the student's speaking abilities and was carried out on October 30 November 4, 6, and 11, 2024. The following are the steps carried out by the researchers in this research:

a. Chose a Title of Research and Preparation of Research Design

The department assigned the researchers a research theme on Artificial Intelligence in education, which presented both opportunities and obstacles for the development of original and pertinent ideas. The researchers developed a specific issue from the broad concept when choosing the title. The chosen title had to be intriguing, represent the study focus, and be relevant to scientific advancements or community needs. When determining the title, the researchers also immediately chose a research methodology, namely experimental research. After the title was approved by the supervisor, the researchers prepared Chapter I (Introduction), Chapter II (Review of Related Literature), and Chapter III (Research Method). The researchers needed to search for previous literature to ensure that the topic raised had novelty or research gaps that could be filled.

b. Research Proposal Seminar

The research proposal seminar was held on July 29, 2024. The proposal seminar was one of the important steps in the research process. At this step, the researchers presented the researcher's

design to the supervisor and examiner to obtain input, criticism, and approval before continuing to the next research stage. The proposal seminar tested the extent to which the researcher understood the topic raised, the clarity of objectives, and the strength of the methodology designed. During the seminar, the researchers explained all parts of the proposal in a confident and structured manner. This step not only required good communication skills, but also readiness to answer various questions from the examining lecturers. The biggest challenge in a proposal seminar was maintaining confidence and flexibility in the face of criticism. The researchers were open to input without compromising the main vision of the research. After the seminar, the researchers were asked to revise the proposal based on the input received. This revision was important to ensure that the research conducted had a strong basis and was able to achieve the stated objectives.

c. Preparation of Research Instruments

The first step in the preparation of research instruments was selecting material for the *Elsa Speak* Application. The researchers selected various material features with different topics in the *Elsa Speak* Application. The material was adapted to the student's level, starting from basic vocabulary to complex sentences. The chosen material was then used throughout the experiment group's treatment after the pre-test. The next instruments were the pre-test and post-test questions. The researchers prepared pre-test and post-test questions with the topic "school." The pre-test and post-test questions consisted of 5 questions, and students could randomly choose 2 questions.

d. Validation The Instrument

The researchers prepared validation sheets for each instrument to be validated by the validator. The pre-test questions, post-test questions, and the learning material in the *Elsa Speak* Application were each validated by two validators, a lecturer in English education and an Extra English class teacher.

e. Implemented Research

The first step carried out by the researchers was an observation in the Extra English Class at Al-Bairuny Islamic School. The researchers brought an official letter from the faculty as an administrative requirement for conducting research at the institution. During the observation, the researchers held a meeting with the tutor to determine the days that would be used for the research.

The researchers obtained 4 days that could be used for the research. The following is an explanation of the 4 days of research:

a. Introduction The *Elsa Speak* Application

The researchers presented the *Elsa Speak* application to the students during the first meeting. The researchers explained what *Elsa Speak* was, how it could be used to learn English, and what features were available in the *Elsa Speak* application. The researchers introduced *Elsa Speak* as an English language learning medium aimed at improving students' speaking skills effectively. *Elsa Speak*, an application based on Artificial Intelligence, was used as an interactive tool to practice pronunciation with real-time feedback. The researchers used this application to identify students' pronunciation errors and provide specific suggestions for improvement. The researchers instructed students in the experiment group to download the *Elsa Speak* application, which would be used as a learning medium during the treatment.

Figure 1. Introduction Elsa Speak Application

b. Pre-test

On the second meeting of the research, two groups of students, namely the experiment group and the control group, gathered in the classroom to take a pre-test on their English speaking skills. This pre-test was designed to measure students' initial abilities in pronunciation, grammar, vocabulary, fluency, and speaking comprehension before being given different treatments in the research.

The experiment group consisted of 15 students who would later use the *Elsa Speak* Application as a learning aid. Meanwhile, the control group, which also consisted of 15 students, continued learning English using traditional methods without the help of applications. The test was carried out individually, where each student randomly chose 2 questions from the 5 questions available on the folded paper. The test theme used in the pre-test was about "school." The researchers ensured that both groups received the same instructions and testing conditions to maintain the validity of the results.

Figure 2. Pre-test

c. Treatment

The researchers used *Elsa Speak* as a learning media during the treatment aimed at effectively improving the English speaking skills of students in the experiment group. *Elsa Speak*, an application powered by Artificial Intelligence, was utilized as an innovative tool to train students' pronunciation by providing specific and accurate real-time feedback.

The researchers utilized the *Elsa Speak* application to design integrated learning activities during the treatment aimed at improving students' pronunciation skills. The researchers specifically selected the feature "Learning by Topic: Education" as the focus of the activities. Through this feature, students were assigned tasks involving the practice of pronouncing words, phrases, and sentences related to the theme of education.

However, the researchers had to limit the treatment duration to two days due to the limitations of the *Elsa Speak* application. In the version used, the application only allowed access to five subtopics per day. This required the researchers to strategize to ensure that the relevant material could be covered optimally within the available time. With this approach, the researchers hoped that the participants would improve their confidence in speaking English, particularly on topics related to school. Despite the application's limitations, the use of *Elsa Speak* was expected to produce significant results in a short amount of time.

Additionally, the researchers monitored students' progress by utilizing data generated by the application, such as pronunciation scores and accuracy improvements. Students were also invited to participate in discussions to evaluate their experiences using *Elsa Speak*. This helped the researchers understand how the application influenced students' motivation and confidence in speaking English.

Figure 3. Treatment Day 1

Figure 4. Treatment Day 2

d. Post-test

In the last step of data collection, two groups of students, namely the experiment group and the control group, gathered in the classroom to take a post-test on their English speaking skills. This post-test was designed to measure students' abilities in pronunciation, grammar, vocabulary, fluency, and speaking comprehension.

Students in the experiment group were given treatment using the *Elsa Speak* application, while the control group underwent learning using traditional methods without the application. A post-test was then carried out to compare learning outcomes between the two groups. The test was conducted individually, where each student randomly chose 2 questions from the 5 questions available on the folded paper. The test theme used in the pre-test was about "school." The researchers ensured that both groups received the same instructions and testing conditions to maintain the validity of the results.

Figure 5. Post-test

Data had to be normally distributed before testing using the Independent Sample T-test. The results obtained using the IBM SPSS Statistics 22 application showed that the experimental group's pre-test significance value was 0.064 > 0.05, and the post-test significance value was 0.181 > 0.05. In the control group, the pre-test significance value was 0.093 > 0.05, and the post-test significance value was 0.175 > 0.05. Based on the significance values obtained, which aligned with the normality test criteria using the Shapiro-Wilk test, it could be concluded that the pre-test and post-test data in both the experimental and control groups had a normal distribution.

After testing for normality and confirming the data was normally distributed, a homogeneity test was conducted using the IBM SPSS Statistics 22 application to determine whether the population variances were the same. The results obtained from the homogeneity test showed that

the mean-based C count value was 0.886 > 0.05. As a result, H1 was accepted, and H0 was rejected, indicating that the post-test data variances of both the experimental and control groups were the same or homogeneous.

Data that had been tested as normal and homogeneous were then subjected to an Independent Sample T-test using the IBM SPSS Statistics 22 application to test the proposed hypothesis. The results from the Independent Sample T-test showed a significance value of 0.0001 < 0.05. It could be concluded that H0 was rejected, and H1 was accepted. In conclusion, the results revealed a statistically significant difference between the two groups, with the experimental group, which had used the *Elsa Speak* application, demonstrating greater improvement in speaking skills than the control group, which had used a textbook.

Discussion

In this research, the researchers aimed to evaluate the effectiveness of using the *Elsa Speak* application in improving the English-speaking skills of students in the Extra Class at Al-Biruny Islamic School. Based on the data analysis results, it can be concluded that the *Elsa Speak* application had a positive impact on students' speaking skills. These findings are consistent with the theory of Sholekhah & Fakhrurriana (2023), which states that *Elsa Speak* can help improve students' English skills.

1. The Effect of Using the *Elsa Speak* Application on Speaking Skills

The analysis results showed a significant improvement in students' speaking skills after using the *Elsa Speak* application, as evidenced by the statistical test results. These findings were consistent with previous research by Hasbi & Nursaputri (2024), which stated that the use of the *Elsa Speak* application in English learning could enhance students' speaking skills. However, despite the significant improvement, the results also indicated that some students still faced challenges in using the *Elsa Speak* application. These challenges were likely caused by limited internet access, a lack of technological understanding, and restricted premium features.

2. Barriers Faced in Using the *Elsa Speak* Application

Although the use of the *Elsa Speak* application provided benefits, several obstacles were identified during the study. One of the primary barriers was limited internet access, a lack of technological understanding, and restricted premium features. These challenges affected the smooth implementation of the experiment and might have prevented some students from gaining maximum benefits from the *Elsa Speak* application. Research by Rinaepi et al. (2022) also noted that the *Elsa Speak* application was often hindered by limited internet access. *Elsa Speak* requires a stable internet connection to provide real-time feedback, and these limitations impacted its effectiveness.

3. Students' Responses to the Use of the *Elsa Speak* Application

Observations conducted by the researchers showed that most students felt more motivated to learn English speaking skills using the *Elsa Speak* application. This was supported by Anggraini et al. (2024), who stated that technology could enhance students' motivation through quick and accurate feedback. However, a small portion of students preferred traditional methods, such as direct learning with instructors using textbooks, as they felt the *Elsa Speak* application lacked direct interaction with peers or teachers.

4. Limitations of the Study and Suggestions for Future Research

This research had several limitations, including a small sample size and a short research period. For future research, it is recommended to conduct studies with a larger sample size and a longer duration to obtain more representative results. Additionally, further research could examine the combination of the *Elsa Speak* application with other teaching methods to explore its impact on overall improvement in speaking skills.

In conclusion, this research provided valuable insights into the effectiveness of using the *Elsa Speak* application in improving students' English speaking skills in Extra classes at Al-Biruny Islamic School. Despite facing some challenges during implementation, such as limited internet access and technological understanding, the findings showed that the application had a positive impact on the development of students' speaking skills. Therefore, the use of applications like *Elsa*

Speak proved to be an effective and enjoyable alternative in the English learning process, particularly for enhancing speaking abilities. This study also opens up opportunities for future research to explore the full potential of the application and address the challenges encountered.

CONCLUSION

This research aimed to address the research problem: "Is the *Elsa Speak* Application effective in improving students' speaking skills in the Extra Class at Al-Bairuny Islamic School?" Using data collection methods that included pre-tests, post-tests, and treatment, the researchers involved two groups an experimental group that used the *Elsa Speak* application and a control group that continued with the textbook as a learning media.

The findings of the research indicated that the experimental group experienced significant improvement in speaking skills after receiving treatment using the *Elsa Speak* Application. Speaking aspects such as pronunciation, vocabulary, grammar, fluency, and comprehension showed greater improvement in the experimental group compared to the control group. Conversely, the control group only showed minor, non-significant improvements, indicating that conventional teaching methods were less effective in this context.

The increase in the experimental group's average scores was supported by statistical analysis, which revealed significant differences between the post-test results of the experimental and control groups. The results from the Independent Sample T-test showed a significance value of 0.0001 < 0.05. It could be concluded that H0 was rejected, and H1 was accepted. These findings demonstrated that the *Elsa Speak* Application, with its interactive features such as pronunciation correction and AI-based evaluation, provided immediate feedback that helped students independently correct their mistakes.

In conclusion, this study found that the *Elsa Speak* Application was an effective tool for improving students' speaking skills in the Extra Class program and could serve as an alternative or complement to conventional teaching methods in schools.

A. Suggestions

The research conducted by researchers showed results in the form of several suggestions that could be given to English mentors, students, and further researchers. These suggestions were:

1. Mentors

This research was designed to provide valuable recommendations for mentors by exploring the use of the *Elsa Speak* Application as an effective learning media. The researchers adopted an experimental method involving two groups: a control class and an experimental class. By comparing the outcomes of both groups, this research aimed to evaluate the impact of *Elsa Speak* on improving students' learning experiences and outcomes. The findings were expected to guide mentors in leveraging innovative learning media like *Elsa Speak*, enabling them to deliver more interactive and efficient mentoring sessions that catered to diverse learning needs.

2. Students

This research offered significant recommendations for students in extra classes by utilizing the *Elsa Speak* application as an engaging and user-friendly learning media. With its interactive features and comprehensive learning facilities, *Elsa Speak* provides students with a fun and practical way to enhance their English-speaking abilities. The findings of this research were expected to demonstrate how the application could support students in developing their speaking proficiency while enjoying the learning process, making it an ideal tool for supplementary English learning.

3. Further Researchers

These researchers were expected to serve as a valuable reference for future researchers interested in conducting studies using the *Elsa Speak* application as a learning medium. By employing an experimental method, this study provided a clear framework for investigating similar variables and research designs. The findings and methodologies outlined in this research

could guide other researchers in refining their approaches, exploring new perspectives, and building upon the results to further enhance the use of innovative learning tools like *Elsa Speak*. Additionally, it offered insights into the practical implementation of technology-based learning media, paving the way for more comprehensive studies in the future.

REFERENCES

Abraham, I., & Supriyati, Y. (2022). Desain Kuasi Eksperimen Dalam Pendidikan: Literatur Review. *Jurnal Ilmiah Mandala Education*, *8*(3), 2476–2482. https://doi.org/10.58258/jime.v8i3.3800

Amin, N. F., Garancang, S., & Abunawas, K. (2023). Konsep Umum Populasi Dan Sampel Dalam Penelitian. *Jurnal Pilar: Jurnal Kajian Islam Kontemporer*, *14*(1), 15–31. https://doi.org/10.21070/2017/978-979-3401-73-7

Anwar, A. (2009). Statistika untuk Penelitian Pendidikan dan Aplikasinya dengan SPSS dan Excel. In *IAIT Press*.

Du Boulay, B. (2016). Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems*, *31*(6), 76–81. https://doi.org/10.1109/MIS.2016.93

Eriana, E. S., & Zein, D. A. (2023). *Artificial Intelligence (Ai) Penerbit Cv. Eureka Media Aksara*.

Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1. https://doi.org/10.11648/j.ajtas.20160501.11

H. Douglas Brown & Priyanvada Abeywickrama. (2004). *Language Assessment: Principles and Classroom Practices*.

Hastjarjo, T. D. (2019). Rancangan Eksperimen-Kuasi. *Buletin Psikologi*, *27*(2), 187. https://doi.org/10.22146/buletinpsikologi.38619

Jamaaluddin, & Indah, S. (2021). Buku Ajar Kecerdasan Buatan. In *Umsida Press*.

Kalalo, F. P., & Pontoh, K. C. (2020). The Use of Artificial Intelligence (AI) in Legal Framework for International Arbitration Practices in Indonesia. *Advances in Social Science, Education and Humanities Research*, *472*(January 2020), 1–6. https://doi.org/10.2991/assehr.k.200917.002

Rao, P. S. (2019). The Importance Of Speaking Skills In English Classrooms. *Alford Council of International English & Literature Journal(ACIELJ)*, *2*(2), 6–18. www.acielj.com

Ratminingsih, N. M. (2010). Penelitian Eksperimental Dalam Pembelajaran Bahasa Kedua. *Prasi*, *6*(11), 31–40.

Riyani, R., Maizora, S., & Hanifah, H. (2017). Uji Validitas Pengembangan Tes Untuk Mengukur Kemampuan Pemahaman Relasional Pada Materi Persamaan Kuadrat Siswa Kelas Viii Smp. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *1*(1), 60–65. https://doi.org/10.33369/jp2ms.1.1.60-65

Sugiyono. (2020). *Metodologi Penelitian Kuantitatif, Kualitatif dan R & D*.

Tae Kyun Kim. (2018). Statistics and Probability. *Korean Journal of Anesthesiology*, *Table 2*, 167–206. https://doi.org/10.4324/9781315686875-6

Widodo, S., Ladyani, F., Asrianto, L. O., Rusdi, Khairunnisa, Lestari, S. M. P., Wijayanti, D. R., Devriany, A., Hidayat, A., Dalfian, Nurcahyati, S., Sjahriani, T., Armi, Widya, N., & Rogayah. (2023). Metodologi Penelitian. In *Cv Science Techno Direct*.