

The use of role-play to improve students' speaking ability: A case of the tenth-grade students at Islamic high school

¹Shofiana Eka Luthfitasari*, ¹Mega Mulianing Maharani, ¹Choiril Anwar,
¹Mohammad Noor Zuhri

¹English Education Department, Faculty of Teacher Training and Education, Universitas Islam Sultan Agung, Indonesia

***Corresponding Author**

Email: shofianaeka5@gmail.com

Abstract

The acquisition of speaking proficiency is a fundamental objective in English language pedagogy, serving as a primary indicator of linguistic mastery in global communication. However, a significant discrepancy remains in the implementation of role-play, where many educators often neglect structured preparation and feedback phases. This study aimed to examine whether the use of a structured role-play framework within the role-play method improved students' speaking ability. Adopting a quasi-experimental design with a non-equivalent control group, the research involved tenth-grade students at MAN 1 Kediri. The experimental group received optimized role-play treatment focusing on linguistic scaffolding and systematic feedback, while the control group was taught using conventional practices. Data were collected through pre-tests and post-tests measuring fluency, accuracy, pronunciation, and grammar, and were analyzed using Independent Sample T-tests and N-Gain scores. The results revealed that the systematic integration of preparation and evaluation phases within Role-Play effectively improved students' speaking ability, as evidenced by the experimental group's superior post-test scores and moderate N-Gain improvement. This research provided empirical evidence on how pedagogical depth in role-play procedures can bridge the gap between fluency and formal linguistic competence in EFL contexts.

Keywords: EFL pedagogy; role-play method; speaking ability; structured role-play; quasi-experimental research

INTRODUCTION

The acquisition of speaking proficiency constitutes an essential objective in English language pedagogy, as it reflects an individual's comprehensive mastery and understanding of the target language. In the context of global communication, speaking skills are frequently regarded as the most crucial competence and are widely recognized as a tangible reflection of one's linguistic depth and command (Rao, 2019). Corroborating this view, Leong and Ahmadi (2017) assert that speaking ability serves as a primary indicator of success in the language teaching and learning process. Furthermore, Brown (2021), emphasizes that interactive simulations, when implemented correctly, can reduce foreign language anxiety, which often hinders students from expressing their ideas. Consequently, English teaching methodologies are required to prioritize the development of authentic and communicative speaking skills, rather than merely focusing on passive aspects such as theoretical grammar mastery or decontextualized vocabulary memorization (Thornbury, 2020).

Among various pedagogical approaches, the role-play method has been empirically proven as one of the most effective instructional strategies. Generally, prior research had implemented role-play as a medium for simulating real-world scenarios aimed at fostering student motivation and self-confidence. Fu and Li (2024) conclude that role-play serves as a transformative technique due to its capacity to provide relevant contexts and enhance emotional engagement during the learning process. Additionally, Wang (2024), suggests that structured role-play provides a safe environment for students to experiment with complex linguistic structures, allowing for deeper language exploration. A study by Ishak and Aziz (2022) also indicate that role-play can significantly improve students' speaking performance, specifically in the context of social interaction. Thus, this method

represents a unique methodology capable of optimizing students' linguistic potential through activities oriented toward meaningful interaction (Ur, 2019).

However, the effectiveness of role-play is highly dependent on the pedagogical depth and structural integrity of each phase. An ideal implementation involves three crucial phases: Preparation (linguistic orientation), Execution (spontaneous interaction), and Evaluation/Feedback (linguistic accuracy). A common issue encountered in the field is the imbalance in these stages. Many educators focus solely on the execution phase while neglecting detailed feedback. Hammami (2024), argues that the effectiveness of role-play lies not only in the activity itself but also in the details of post-activity follow-up. Without optimization, students may achieve fluency but continue to commit the same grammatical errors repeatedly, a phenomenon known as error fossilization (Richards, 2019). Zhang (2021), further emphasizes that systematic feedback acts as a bridge between implicit and explicit knowledge, ensuring that communicative tasks do not sacrifice accuracy for the sake of fluency (Zhang and Ardasheva, 2020).

The novelty of this research lies in the Integrated Structured Role-Play Procedure within the role-play method. In contrast to previous studies that often conducted role-play flexibly without a rigorous evaluation structure. Abbasi (2023), this study analyzed how reinforcement in the Preparation stage and the Feedback stage can enhance speaking competence holistically. Moreover, Dyke and Acton (2022) note that clear social roles combined with structured feedback can significantly improve pragmatic competence. The focus of this research was on the managerial quality of the learning process to ensure that every activity exerted a maximum impact on student proficiency. Therefore, this study aimed to examine whether the use of role-play improved students' speaking ability, particularly by addressing the research question: Does the use of role-play improve the students' speaking?

METHOD

This chapter outlined the methodological framework utilized to investigate the effectiveness of optimized role-play procedures in improving students' speaking proficiency. It provided a detailed description of the research design, the population and sample involved, the instruments used for data collection, and the systematic procedures followed during the study. Furthermore, it explained the statistical techniques employed to analyze the collected data to ensure the validity and reliability of the research findings.

Research Design

This study adopted a quasi-experimental design, specifically the non-equivalent control group design. As suggested by Cohen et al. (2018), this design is particularly suitable for educational research where the use of pre-existing or 'intact' classes is more practical than random assignment. By comparing an experimental group and a control group, this research sought to establish a causal relationship between the methodological intervention and English speaking proficiency.

The choice of a quasi-experimental design was primarily driven by practical and ethical considerations within the school setting. Unlike a true experimental design, which requires random assignment of individual subjects, this study utilized intact classes already established by the school administration at MAN 1 Kediri. This approach was necessary to avoid disrupting the existing educational schedule and to maintain the natural classroom environment, ensuring that the research findings possessed high ecological validity.

To facilitate the subsequent N-Gain analysis, Following the notation system from Cohen et al. (2018), both groups underwent a pre-test (O_1 , O_3) to measure their initial speaking proficiency and a post-test (O_2 , O_4) to assess the effects of the intervention (X). The comparison between these two measurement points allowed for a precise calculation of the improvement magnitude within and between the groups.

Respondents and Research Setting

The population of this study comprised the tenth-grade students of Madrasah Aliyah Negeri (MAN) 1 Kediri during the second semester of the 2025/2026 academic year. Through a purposive sampling technique, two intact classes were selected as the research sample. Class X-J, consisting of 22 students ($n=22$), was designated as the Experimental Group (G_1), which received the intervention through structured role-play methods (X). Meanwhile, class X-I, consisting of 21 students ($n=21$), served as the Control Group (G_2), was taught using conventional teaching.

The selection of these subjects was based on several pedagogical considerations. First, tenth-grade students were in a transitional phase that required intensive communicative practice to build a solid English foundation. Second, based on initial observations and teacher recommendations, both classes exhibited similar levels of English proficiency, fulfilling the requirements for the non-equivalent control group design. This setting ensured that the research was conducted within a natural classroom environment, maintaining the study's ecological validity.

Instrument

Primary quantitative data were collected through a Speaking Performance Test, which was administered as both a pre-test and a post-test. To ensure the instrument's content and construct validity, the test was developed based on the standard syllabus for tenth-grade students and underwent a rigorous expert judgment process. The instrument was validated by a senior English teacher at MAN 1 Kediri to ensure its alignment with the national curriculum and the students' proficiency level.

The speaking test utilized two different but comparable role-play prompts to minimize the "testing effect" (students remembering answers from the first test). The pre-test required students to perform a "Giving Directions" scenario, while the post-test focused on a "Transaction in a Store" scenario. Although the themes differed, both tests were designed to measure the same level of linguistic complexity and communicative competence.

To maintain objectivity in scoring, a standardized scoring rubric adapted from the framework by Fulcher and Davidson (2018) was employed. This rubric measured four main domains: fluency, accuracy, pronunciation, and grammar. Each domain was assessed on a five-point Likert scale (1-5), providing a maximum possible score of 20. The scores from both the pre-test and post-test were then used as the primary data for the subsequent N-Gain analysis.

Procedures

The data collection was conducted systematically in three distinct stages over a three-day period, from February 12 to February 14, 2026. The procedures were designed to ensure a rigorous comparison between the experimental and control groups through the following phases.

Preliminary Phase: Pre-test (O_1 , O_3)

On the first day, February 12, 2026, a pre-test was administered to both the experimental (X-J) and control (X-I) groups. Students were instructed to perform a role-play scenario titled "Asking and Giving Directions," where they assumed roles as senior and new students. This phase was conducted without any prior scaffolding or feedback from the researcher to establish the baseline speaking proficiency of both groups.

Experimental Group Treatment Phase (structured Role-Play)

The intervention phase spanned three days, during which the researcher applied different instructional methods to each group. The researcher implemented an "Optimized Role-Play" procedure which followed a systematic cycle. Each day, students were provided with linguistic scaffolding, including essential vocabulary and grammar structures (e.g., Simple Present and Future Tense) to support their performance. On February 12, the theme was "Planning a Dream Vacation"; on February 13, the topic was "Job Interviews"; and on February 14, students were given the autonomy to choose their own professional roles. A crucial element of this treatment was

the constructive feedback sessions conducted immediately after each performance. The researcher provided direct corrections on common errors in grammar (e.g., subject-verb agreement) and pronunciation (e.g., architect, important, should) to ensure students understood the correct forms in context.

Control Group Treatment Phase (Conventional Method)

In the control group, the teaching process followed the same thematic curriculum as the experimental group, covering topics such as 'Planning a Vacation' and 'Job Interviews.' However, the instructional delivery was strictly conventional and teacher-centered. In this setting, the teacher began each session by writing a list of relevant vocabulary and generic dialogue scripts on the whiteboard. The students were then instructed to transcribe these notes into their workbooks and memorize them through individual repetition. Unlike the experimental group, the control group did not engage in any pre-communicative scaffolding or interactive simulations. The interaction was limited to teacher-to-student questioning, where the teacher asked students to read the memorized vocabulary aloud. Furthermore, no systematic feedback or error correction was provided regarding their pronunciation or grammatical accuracy during these sessions. The focus remained on passive comprehension and rote memorization rather than active linguistic production.

Final Phase: Post-test (O_2 , O_4)

On the final day, February 14, 2026, after the treatment sessions were completed, both groups underwent a post-test. The students performed a "Shopping at a Stationery Store" simulation. Their performances were recorded and assessed using the validated scoring rubric. These post-test results were subsequently compared with the pre-test scores to determine the magnitude of improvement using the N-Gain analysis.

Data Analysis

To process the quantitative data obtained from the pre-test and post-test, the researcher employed descriptive and inferential statistics using SPSS software. The analysis followed three systematic steps to ensure the validity of the results:

Descriptive Statistics

This analysis was used to determine the mean, standard deviation, maximum, and minimum scores of both the experimental and control groups. This provided an overview of the students' speaking proficiency across the four domains, which included accuracy, fluency, grammar, and pronunciation.

Prerequisite Tests

Before conducting the hypothesis testing (T-test), two prerequisite tests were performed. The first was Normality Test, which aimed to determine whether the data from both groups were normally distributed, the second was Homogeneity Test which was conducted to ensure that the variance between the experimental and control groups was equal or homogeneous.

Inferential Statistics (Hypothesis Testing)

To answer the research questions, the following statistical tests were conducted: Paired Sample T-test: This was used to evaluate the significant improvement within each group by comparing their respective pre-test (O_1 , O_3) and post-test (O_2 , O_4) scores, independent Sample T-test: This was employed to compare the post-test scores between the Experimental Group and the Control Group to determine the final difference in effectiveness. Additionally, the N-Gain score was calculated to measure the magnitude of the improvement in both groups after the intervention.

Normalized Gain (N-Gain) Analysis

To measure the magnitude of the improvement and the effectiveness of the optimized role-play intervention (X), the Normalized Gain (N-Gain) score was calculated. The results were then interpreted based on the classification framework by (Sundayana, 2020).

Table 1. N-Gain Score Effectiveness Classification

N-Gain Score Range	Effectiveness Category
$g > 0.70$	High
$0.30 \leq g \leq 0.70$	Medium
$g < 0.30$	Low

In summary, the methodology described in this chapter served as a comprehensive guide for conducting the research at MAN 1 Kediri. By integrating a quasi-experimental design with rigorous data collection and analysis, this study aimed to provide an objective measurement of how optimized role-play impacted students' speaking skills. The subsequent chapter presents the results of these analyses and discusses the findings in relation to the established research questions.

Reducing Bias in the Research Process

To ensure the validity and objectivity of the research findings, several strategic measures were implemented to minimize potential bias throughout the study. First, the researcher employed a standardized scoring rubric adapted from Fulcher and Davidson (2018) to mitigate subjective judgment during the evaluation of students' speaking performances. To further address instrument bias, the speaking performance test underwent a rigorous expert judgment process by a senior English teacher at MAN 1 Kediri to ensure its alignment with the national curriculum and students' actual proficiency levels.

Moreover, to prevent the testing effect, the researcher utilized two distinct yet comparable role-play prompts for the pre-test and post-test, ensuring that the improvement in scores was a result of linguistic growth rather than mere memorization of previous tasks. Lastly, the adoption of a quasi-experimental design using intact classes allowed the study to maintain a natural classroom environment, thereby reducing the Hawthorne effect and ensuring high ecological validity.

RESULTS AND DISCUSSION

Normality Testing

Normality testing was a crucial prerequisite step to ensure that the obtained data were eligible for analysis using parametric statistics. In this study, prior to hypothesis testing, the researcher performed a normality test using the Shapiro-Wilk method. This method was selected based on its effectiveness in measuring small samples ($n < 50$), as the experimental group consisted of 22 students and the control group comprised of 21 students. The analysis results in Table 2 indicated that the significance values (Sig.) for both Pre-test and Post-test in both groups were above 0.05. This demonstrates that the data are normally distributed and meet the criteria for further analysis using the Paired Sample T-test and Independent Sample T-test to evaluate the impact of the proposed instructional intervention.

Table 2. Normality Testing Results (Shapiro-Wilk)

Variable	Sig. (Shapiro-Wilk)	Result
Pre-test Control	.163	Normal
Post-test Control	.175	Normal
Pre-test Experimental	.100	Normal
Post-test Experimental	.072	Normal

Descriptive Statistics

This section presents the results of the pre-test and post-test scores for both the Experimental Group (X-J) and the Control Group (X-I). The data description was categorized into the students' overall scores and their performance across specific speaking domains, which included fluency, accuracy, pronunciation, and grammar.

The overall achievement of students in both groups indicated an upward trend. However, the magnitude of this improvement varied between the two classes. The descriptive statistics for both groups, including the mean, standard deviation, and variance, were summarized in the table below.

Table 3. Students' Mean Scores Achievement Based on Speaking Domains

Kelas			Fluency	Accuracy	Pronunciation	Grammar
X-I	Pre-test	Total	44	42	31	35
		Mean	2,095238	2	1,4761905	1,666667
	Post-test	Total	74	64	51	53
		Mean	3,52381	3,047619	2,428571	2,52381
X-J	Pre-test	Total	46	37	34	31
		Mean	2,090909	1,681818	1,5454545	1,409091
	Post-test	Total	85	73	63	60
		Mean	3,863636	3,318182	2,863636	2,727273

T-test

The primary focus of this research phase was to determine the extent of the improvement in students' speaking proficiency within each group after the three-day pedagogical intervention. To achieve this, a Paired Sample T-test was conducted, comparing the mean scores of the pre-test and post-test for both the experimental and control groups. This analysis was vital as it provided empirical evidence of "learning growth," allowing the researcher to observe whether the Integrated Structured Role-Play Procedure (for the experimental group) and the conventional role-play (for the control group) successfully elevated the students' linguistic performance from their initial baseline.

Table 4 presented a summary of student achievement. The mean Post-test score for the experimental group reached 13.00, which was higher than the control group's mean score of 11.52. This improvement indicated a positive impact resulting from the intervention provided.

Table 4. Comparison of Pre-test and Post-test Mean Scores for Control and Experimental Groups

Group	Test	Mean	Sig. (2-tailed)
Control	Pre-test	7.19	< .001
	Post-test	11.52	
Experimental	Pre-test	6.81	< .001
	Post-test	13.00	

To examine the significance of improvement within each group, a Paired Samples T-Test was conducted. The results of the Paired Samples T-test showed a Sig. (2-tailed) value of < 0.001 for both groups. This indicated a highly significant improvement in speaking ability in both the control and experimental classes following the treatment. Furthermore, an Independent Samples T-test was performed to compare the final outcomes between the two groups.

The increase in the average score of the experimental group from 6.81 to 13.00 resulted from a structured intervention. During the treatment sessions, the researcher observed that students' progress was driven by the scaffolding stages and the provision of relevant vocabulary. For instance, on the topic of 'Dream Vacation,' students initially made frequent errors such as 'I will busy' or 'She need.' However, after receiving immediate feedback, they were able to correct these to 'I will be busy' and 'She needs.' This iterative process of grammatical refinement helped students achieve higher scores in the post-test.

Effectiveness Analysis (N-Gain Score)

To measure the effectiveness of the method, the researcher utilized the N-Gain score. The results demonstrated that the experimental group achieved an average N-Gain of 47.02% (Moderate category), whereas the control group only reached 34.85% (Low/Moderate category). The N-gain data further corroborated that the structured role-play intervention yielded a higher progression rate.

Table 6. Effectiveness Analysis Based on N-Gain Score

Group	Mean N-Gain (%)	Effectiveness Category
Control	34.85%	Low/Medium
Experimental	47.02%	Medium

The N-Gain score indicated effectiveness in the 'Moderate' category. This development was evident in the improvement of students' confidence and pronunciation accuracy. During the second and third sessions (Job Interview and Professional Roles), students who initially struggled with pronouncing words such as 'usually' and 'architect' began to demonstrate better control. The use of realistic topics helped students focus on functional communication, which explained the score increase in their speaking components.

Discussion

The data indicated a progressive shift in speaking mastery among tenth-grade students, largely attributed to the structured stages of the implemented strategy. This was evident from the experimental group's superior performance in both post-test scores and N-Gain categories compared to the control group. These findings addressed the research question by confirming that a structured approach to role-play was more effective than conventional practices.

The success of this intervention was closely linked to the systematic three-day treatment process. On the first day, following a pre-test about giving directions, the experimental group practiced "Planning a Dream Vacation" through role-play that integrated preparation and feedback. Meanwhile, the control group only practiced vocabulary pronunciation from the whiteboard in a conventional manner. On the second day, the experimental group utilized linguistic scaffolding for a "Job Interview" scenario, allowing for deeper language exploration. This contrasted with the control group's teacher-centered approach on the same topic. On the final day, students in the experimental group exercised autonomy by choosing various professions for their role-play, adhering to the established flow of preparation and evaluation.

The significant improvement in the experimental group's speaking proficiency was inseparable from the crucial roles of the Preparation and Feedback stages within this optimization strategy. In contrast to the control class, which tended to be unidirectional (teacher-to-student), the experimental group was provided with linguistic scaffolding at the beginning of the session to establish a foundation of vocabulary and grammar relevant to the scenarios. One of the most prominent findings was the impact of systematic feedback. In conventional practices, students' errors are often overlooked to maintain fluency; however, this carries the risk of triggering error fossilization, where students subconsciously internalize recurring grammatical mistakes.

For instance, during the 'Dream Vacation' session, many students initially employed incorrect sentence structures such as 'I will busy.' Through immediate evaluation sessions, the researcher provided explicit corrections, enabling students to refine their utterances to 'I will be busy' in subsequent performances. This demonstrated that feedback acted as a bridge between implicit and explicit knowledge, ensuring that improvements in speaking fluency did not compromise linguistic accuracy. Consequently, the success of this method indicated that student autonomy in expression must be supported by a robust pedagogical framework to achieve holistic communicative competence.

CONCLUSION

This study confirms that the integration of linguistic scaffolding and systematic feedback within a structured role-play framework significantly enhances students' speaking proficiency compared to conventional teaching methods. The systematic intervention, which includes preparation, execution, and rigorous evaluation phases, allows students to achieve a better balance between fluency and linguistic accuracy. By providing a clear pedagogical structure, this method effectively helps students overcome common speaking challenges and prevents the risk of error fossilization. Therefore, the Integrated Structured Role-Play Procedure is highly recommended for English educators as a practical and effective strategy to foster communicative competence in EFL contexts.

Despite its effectiveness, this research has several limitations that offer opportunities for future development, such as the relatively short intervention period and the specific sample population which may limit the long-term assessment and broad generalizability of the findings. To address these, future research should involve larger and more diverse samples across various educational levels and extend the timeframe to evaluate the durability of the linguistic gains. Additionally, integrating digital technology for more dynamic feedback and employing mixed-method approaches, such as in-depth interviews, are suggested to gain a more comprehensive understanding of students' perceptions and to further refine the pedagogical impact of structured role-play.

REFERENCES

- Abbasi Montazeri, E., Jalilifar, A., & Hita, J. A. (2023). Evaluative language in applied linguistics research article discussions: exploring the functions and patterns of that-structures in argumentative texts. *Language Awareness*, 32(2), 193–216. <https://doi.org/10.1080/09658416.2021.1990938>
- Brown, H. D. (2021). *Teaching by principles: An interactive approach to language pedagogy* (5th ed.). Pearson Education.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8th ed.). Routledge.
- Dyke, A. V., & Acton, W. (2022). Role-play and dialogic meta-pragmatics in developing and assessing pragmatic competence. *Pedagogical Linguistics*. <https://doi.org/10.1075/pl.22004.van>
- Fu, X., & Li, Q. (2024). Effectiveness of role-play method: A meta-analysis. *International Journal of Instruction*, 18(1), 309–324. <https://doi.org/10.29333/iji.2025.18117a>
- Fulcher, G., & Davidson, F. (2018). *The Routledge handbook of language testing* (2nd ed.). Routledge. <https://doi.org/10.4324/9781003220756>
- Hammami, O., Zaouche, K., Kallel, M., & Nouria, M. (2024). Evaluation of learning abilities after role-playing method: Comparing outcomes of Active and Observer. *La Tunisie Médicale*, 102(11), 866–870. <https://doi.org/10.62438/tunismed.v102i11.4996>
- Ishak, S. A., & Aziz, A. A. (2022). Role Play to Improve ESL Learners' Communication Skills: A Systematic Review. *International Journal of Academic Research in Business & Social Sciences*, 12(10). <https://doi.org/10.6007/ijarbs/v12-i10/14851>
- Leong, L.-M., & Ahmadi, S. M. (2017). An Analysis of Factors Influencing Learners' English Speaking Skill. *International Journal of Research in English Education*, 2(1), 34–41. <https://doi.org/10.18869/acadpub.ijree.2.1.34>

- Rao, P. S. (2019). VS Publications Alford Council of International English & Literature Journal(ACIELJ). *Alford Council of International English & Literature Journal(ACIELJ)*, 401(2), 6–18.
- Richards, J. C. (2019). *Curriculum development in language teaching*. Cambridge University Press. <https://doi.org/10.1017/9781009024556>
- Sundayana, R. (2020). *Statistika Penelitian Pendidikan*. Alfabeta.
- Thornbury, S. (2020). *How to teach speaking*. Pearson Longman.
- Ur, P. (2019). *A course in English language teaching*. Cambridge University Press. <https://doi.org/10.1017/9781009024518>
- Wang, S. (2024). The Effectiveness of a Role-Play Activity in Practicing EFL Learners' Communicative Competence. *Journal of Education and Educational Research*, 9(3), 80–82. <https://doi.org/10.54097/gknekf50>
- Zhang, T. (2021). The effect of highly focused versus mid-focused written corrective feedback on EFL learners' explicit and implicit knowledge development. *System*, 99, 102493. <https://doi.org/10.1016/j.SYSTEM.2021.102493>
- Zhang, X., & Ardasheva, Y. (2020). Balancing fluency and accuracy in communicative language teaching. *The Modern Language Journal*, 104(2), 350–370. [10.54097/chss.v23i.13909](https://doi.org/10.54097/chss.v23i.13909)

APPENDIX

Speaking Performance Test Instrument

The pre-test instrument was implemented through a paired speaking task focusing on the functional language of asking and giving directions within a school environment. In this scenario, one student assumed the role of a newcomer who needed to locate essential facilities, specifically the school canteen and the library, while the partner acted as a senior student who provides clear and accurate directions to these locations. This interaction was designed to last for approximately two to three minutes, providing a baseline measure of the students' initial fluency, pronunciation, and grammatical accuracy in a basic communicative context.

The post-test instrument followed a similar paired format but introduced a slightly more complex social transaction involving a shopping scenario at a stationery store. One student was assigned the role of a customer looking for a birthday gift, such as a notebook or a pen, while operating under a specific budget constraint. The other student took on the role of a shopkeeper who had to respond politely, offer appropriate product recommendations, and discuss pricing. This dialogue, also conducted within a two to three-minute timeframe, served to evaluate the students' progress in functional language use and their ability to maintain a coherent conversation following the pedagogical intervention.