Scilight

*Journal of Educational Technology and Innovation*
https://jeti.thewsu.org/index.php/cieti

*Article*

# An Empirical Study of the Impact of AI-Based Reflective Dialogue Model on EFL Students' Oral Expression Skills

Yu Zhao, You-You Zhang, Xiao-Nan Luo, Dan-Ting Deng and Qun-Fang Zeng *

School of Education, Wenzhou University, Wenzhou 325035, China
* Correspondence: 00032008@wzu.edu.cn

**How To Cite:** Zhao, Y., Zhang, Y.-Y., Luo, X.-N., Deng, D.-T., & Zeng, Q.-F. (2025). An Empirical Study of the Impact of AI-Based Reflective Dialogue Model on EFL Students' Oral Expression Skills. *Journal of Educational Technology and Innovation*, *7*(3), 9–20. https://doi.org/10.61414/2d7sam34

Received: 20 February 2025
Revised: 1 June 2025
Accepted: 25 June 2025
Published: 30 September 2025

**Abstract:** Oral expression skills play an essential role in the development of EFL students' language abilities, and how to improve EFL students' oral expression skills is an essential and challenging task. This study adopts a quasi-experimental research method to carry out the research and proposes an AI-based reflective dialogue model. Based on this, an analysis of the impact brought by this model on EFL students' oral expression performance and learning anxiety levels. The results show that students in the experimental group have significantly higher oral expression performance than those in the control group in the three dimensions of grammatical accuracy, expressive fluency, and word accuracy. In addition, the students in the experimental group produced facilitated anxiety after using the AI-based reflective dialogue model for oral expression learning, which prompted the students to learn more diligently.

**Keywords:** human-computer dialogue model; oral expression; reflective dialogue; AI; EFL student

## 1. Introduction

Oral expression skills are the core of EFL students' comprehensive language use skills and the necessary skills for cross-cultural communication in the context of globalization, which plays an important role in students' personal development. China's University English Teaching Guide (Language Teaching Steering Committee, 2020) also clearly lists "accuracy, fluency, and appropriateness of oral expression" as a teaching focus, highlighting the importance of oral expression skills in cultivating talents.

However, traditional oral teaching faces dilemmas: First, due to insufficient proficiency, students often avoid speaking for fear of making mistakes and have anxiety about oral expression (Chen & Hwang, 2020). Second, due to the limited classroom time and the difficulty for teachers to instruct every student, students do not have enough opportunities to practice (Chen et al., 2017). These two factors lead students into a vicious cycle of "not being able to speak, not daring to speak", which affects the improvement of students' oral expression ability (Liu & Li, 2019). In this situation, generative AI has gradually been applied to oral speech training in recent years through instant feedback, speech recognition, etc. Mudawy's (2025) study showed that students' oral fluency and pronunciation were improved after learning with AI-driven conversation tools. However, such models are mostly centered on "error correction" and only provide "correct and incorrect judgments". Students are in a state of passive acceptance, lacking active exploration of the causes of errors. This learning approach is not beneficial to students' reflection, which leads to repeated errors.

Based on the above background, this study proposes an AI-based reflective dialogue model, with the help of AI's reflective questioning (e.g., "Is there a more suitable vocabulary substitution?" "Is the paragraph connection natural?") to guide the students to reflect in their oral expressions and improve their oral expression skills.

## 2. Literature Analysis

Reflective learning is a learning approach in which learners review and analyze their own learning process in order to achieve cognitive reconstruction and competence enhancement (Zhang et al., 2025), the roots of which can be traced back to Dewey's (1933) proposal of "reflective thinking", which emphasizes cognitive development through reflection on experience. This concept is manifested in language learning as learners review and analyze their own language use (e.g., vocabulary, grammar) to improve their language proficiency, while Schön (2017) further proposes the theory of "reflection-in-action", which emphasizes learners' behavioral and cognitive adjustments and improvements in practice. Therefore, in EFL teaching, reflective learning can help students adjust their learning strategies in grammar and other aspects in real-time during the learning process and improve their oral expression (Chen, 2021). In addition, approaches such as Project-Based Learning (PBL) (Almulla, 2020) and Task-Based Learning (TBL) (Sholeh et al., 2020) are similar to the concept of reflective learning. What these approaches have in common is that they are learner-centered, emphasize process direction, and effectively develop students' metacognitive skills, i.e., learners need to review the learning process from their own perspective, dynamically adjust their strategies, and improve their ability to monitor the cognitive process in reflection.

There are three main types of approaches to the practice of reflective learning in teaching and learning, the first of which is based on personal reflection, such as reflective journals. For example, Alt et al.'s (2022) study showed that reflective journals enhance students' ability to reflect and record learning tracks. The second is reflection based on peer assessment. For example, Imani (2021) found that students' discussions with peers enriched learning paths and enhanced speaking skills. Finally, it is a reflection based on technical support. For example, Yang et al. (2025) proposed a personalized diagnostic strategy based on an AI platform, which improved students' writing proficiency and self-efficacy. However, reflective learning still has some limitations in practice. From the teacher level, it is difficult to provide immediate feedback to each student due to limited energy, which affects the effectiveness of reflection to some extent (Carless & Winstone, 2023). At the student level, reflection relies on students' autonomy, and if students are unable to recognize errors, this will lead to a lack of depth of reflection. Moreover, most students just describe the learning process without effective reflection, which will affect their long-term learning outcomes (Dyment & O'connell, 2010).

In recent years, the development of generative AI has pushed the human-computer dialogue mode from rule-based mechanical interaction to intelligent, reflective interaction, which provides new ideas to solve the above dilemma. Human-Computer Dialogue realizes human-computer interaction through natural language, and the core goal is to simulate human dialogue logic to achieve information transfer and cognitive collaboration (Wang & Yuan, 2016). Human-computer dialogue can be traced back to the Turing test in the mid-20th century (Turing, 2009), which outlined the initial blueprint for subsequent human-computer dialogue research. At this stage, human-computer dialogue relies mainly on rule-based pattern recognition. For example, ELIZA, developed by Weizenbaum (1966), is only capable of generating fixed responses based on a limited number of keywords, is unable to deal with deep semantic structure, and lacks semantic understanding. This will result in a lack of coherence in the dialog and will not satisfy communication needs.

With the development of deep learning technology, human-computer dialogues have entered the stage of diagnostic feedback, which shows great potential in the field of language learning. For example, speech recognition technology can analyze pronunciation deviations in real-time and provide error correction feedback, resulting in an increase in students' grammatical correctness (Mudawy, 2025). Nst et al. (2023) applied Rosetta Stone software to EFL learning, which improved students' English writing, listening, and speaking skills. In conclusion, the efficiency of feedback at this stage is greatly improved, but it still follows the path of "identifying errors-providing corrections" and lacks in-depth exploration of the causes of errors, such as migratory errors or cognitive strategy biases. This may lead to error entrenchment and affect students' acquisition of language proficiency (Ellis, 1997; Warschauer et al., 2023). With the rapid development and application of generative large language models, human-computer dialogues have also begun to shift towards reflective interaction models, whose core feature is to generate heuristic questions through dialogues that lead learners to reflect actively. For example, Alharbi (2025) proposes a WhatsApp-based informal peer interaction model, which encourages students to reflect through heuristic feedback and enhances their language proficiency. In conclusion, generative large language modeling is capable of generating coherent, logically clear, and creative textual responses, which have been shown to perform at a high level in many tasks (Lo, 2023).

Therefore, this study tries to propose an AI-based reflective dialog model to make up for the shortcomings of traditional reflective methods in terms of coverage, immediacy, and depth and to provide technical support for the improvement of EFL students' oral expression ability.

## 3. Research Questions

Based on the above analysis, this study proposes an AI-based reflective dialogue model to promote reflection among students during the learning process, improve their oral expression skills, and alleviate their anxiety. This study focuses on the following two aspects to investigate whether the AI-based reflective dialogue model can have a positive impact on EFL students' oral expression learning. The questions are as follows:

RQ1: Compared with the AI-based diagnostic dialogue model, can the AI-based reflective dialogue model improve EFL students' oral expression skills?

RQ2: Compared with the AI-based diagnostic dialogue model, can the AI-based reflective dialogue model reduce EFL students' learning anxiety?

## 4. Instructional Design for AI-Based Reflective Dialogue Model

### *4.1. Platform Introduction*

The Doubao AI (https://www.doubao.com) used in this study is an intelligent assistant developed by ByteDance, a Chinese company. It possesses powerful conversational capabilities and can answer various questions posed by users and engage in fluent dialogue.

In this study, Doubao AI offers two major advantages: On the one hand, Doubao AI can systematically process text, identify issues such as improper vocabulary and grammar usage, and provide students with appropriate revision suggestions. It can also analyze paragraph coherence from a logical perspective, guiding students to optimize content layout and paragraph transitions to enhance the text's logical structure. On the other hand, Doubao AI leverages speech analysis technology to evaluate spoken recordings across multiple dimensions, accurately identify pronunciation deviations and other issues, and provide guidance. Meanwhile, Doubao AI can offer students personalized strategies for intonation and speaking speed based on different scenarios (e.g., daily conversations, speeches), enabling oral expressions to align with application requirements and facilitating improvements in oral proficiency, as illustrated in Figure 1.

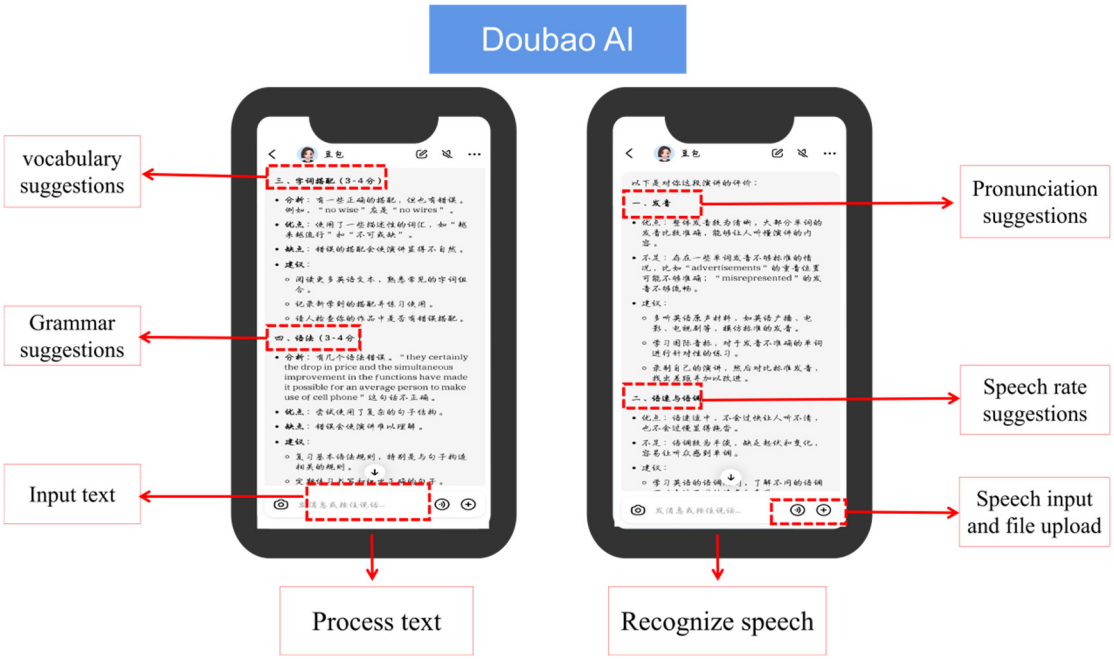


**Figure 2.** AI-based Reflective Dialogue Model.

### *4.2. AI-Based Reflective Dialogue Model Construction*

This study proposes an AI-based reflective dialogue model, which is characterized by designing the AI tool Doubao AI as a "questioner" role, integrating it into the learning process of oral expression, and guiding students' reflection as a "questioner" to promote the improvement of students' oral expression skills. The role of "Questioners" is designed to be integrated into the learning process of oral expression. Specifically, after completing the oral learning, students upload the text and recording to Doubao AI, which questions them on the

dimensions of vocabulary, pronunciation, speed of speech, etc., and the students reflect on them and make corrections to the text and expression according to the content of the dialogues, which can form a closed-loop of learning of "Dialogue-Reflection-Enhancement".

In teaching practice, Doubao AI is deeply integrated into the two core parts of oral learning, as shown in Figure 2. In the expression text creation stage, students upload the text after completing the first draft, and Doubao AI analyzes the dimensions of vocabulary, grammar, content selection, and logical structure and guides students to reflect by questioning and asking questions. For example, for the problem of single vocabulary use, Doubao AI will ask if there are more relevant words to replace; for the part of the logical structure that is unreasonable, Doubao AI will guide the students to reflect on the naturalness of the connection between paragraphs, and then the students will make textual modifications based on these questions. In the oral expression stage, students practice speaking with the revised text and then upload their speech to the Beanbag AI, which carefully evaluates their pronunciation, intonation, and speed of speech and questions their speed of speech, prompting them to review and reflect on the process of expression, identify and correct problems, and realize the improvement of their oral expression skills.

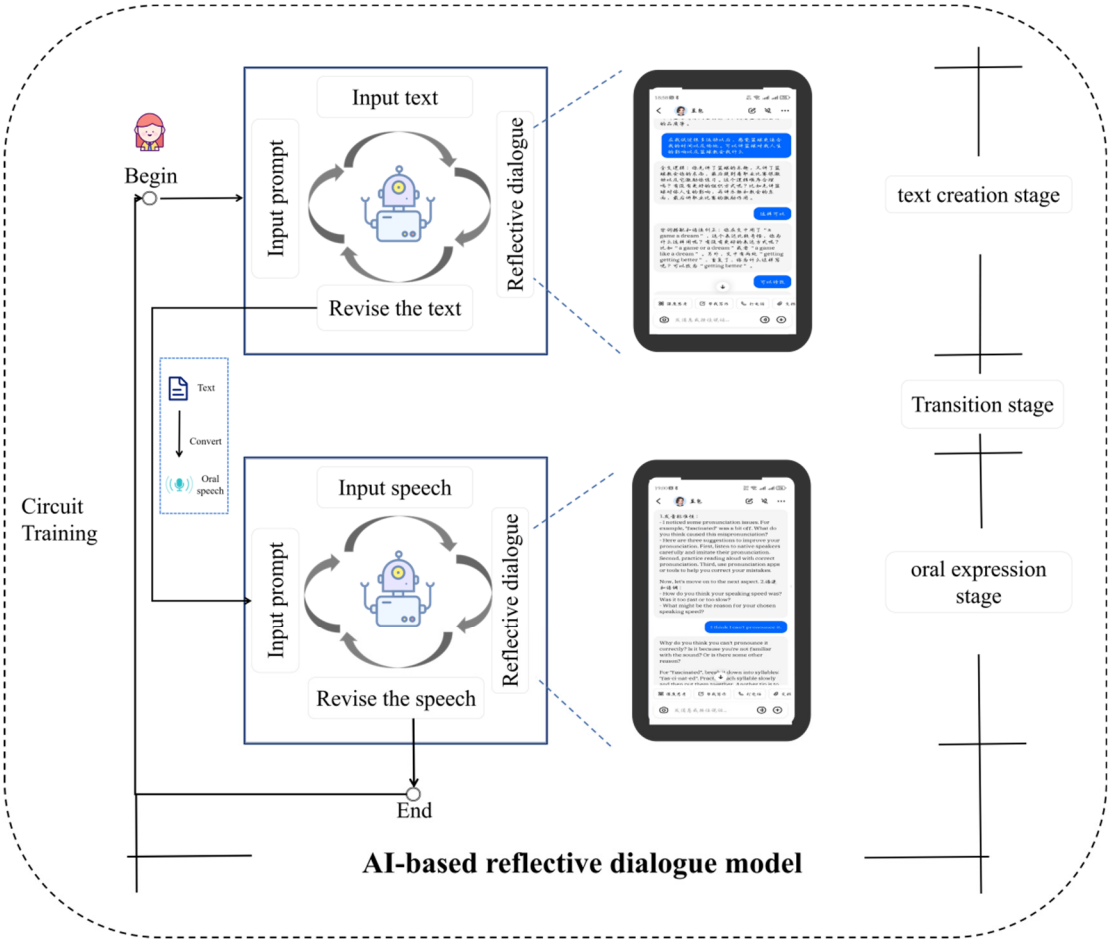


**Figure 2.** AI-based Reflective Dialogue Model.

### *4.3. Design of Teaching Activities*

In this study, we take the theme of "your favorite sport" as an example and divide the teaching activities into four progressive sections. As shown in Figure 3.

Section 1 is "Understanding and Creation". Firstly, students understand the theme of the activity under the teacher's explanation and guidance, and secondly, they carefully understand the examples given by the teacher and experience the skills provided by the teacher to make their speeches creative. For example, in the theme of "favorite sport," students are guided to express themselves creatively through model speeches and the inspiration of multi-style examples. Students can talk about the benefits of sports through their own experiences, as well as their emotional resonance with sports. In the end, the initial creation of the text will be completed under the guidance of the teacher.

Section 2 is "Dialogue and Revision". Under the guidance of the teacher, students have a dialogue with the Doubao AI and revise the text according to the content of the dialogue, focusing on improving the accuracy and

logic of the text. During the conversation, the AI will ask questions about the text, including grammatical standards, vocabulary accuracy, reasonable connection, etc. The purpose is to help the students choose the correct vocabulary and grammar and make reasonable connections between the lines so that the content has a strong logic and ultimately is able to accurately express the student's point of view.

Section 3 is "Dialogue and Reflection". Students practiced expressing the revised text and engaged in oral expression with Doubao AI. Similarly, in this process, the Doubao AI was allowed to ask questions about the students' expressions, including pronunciation, speed of speech, fluency, and so on. For example, "Do you think it's appropriate to use too fast speech speed when talking about your experience?" The aim is to help students choose a more appropriate presentation in their expressions.

Section 4 is "Summarize and Improve". Students summarize the learning process from creation to expression, such as the correct use of the Doubao AI and the speed of speech in the process of oral expression, taking into account the learning and dialogue process in the previous three sessions. On the basis of the summarization, they finally obtain the direction of improving their oral expression skills, such as vocabulary, grammar, and speech logic.

It is important to note that the AI-based diagnostic dialog model also has four oral training sessions. Among them, the "Understanding and Creation" and "Summarize and Improve" sessions are consistent with the teaching process of the experimental group. However, there is a significant difference between the control group and the experimental group in the two key sessions of "Dialogue and Revision" and "Dialogue and Reflection". In the "Dialogue and Revision" section, students in the control group uploaded their own expression texts to Doubao AI so that Doubao AI could help check the texts for grammatical and lexical errors and give them a revised version. In the "Dialogue and Reflection" session, students uploaded their oral recordings, and Doubao AI evaluated their oral expressions to diagnose problems in pronunciation and fluency. However, in this dialog model, students use Doubao AI only for problem diagnosis and error correction, not for reflective guidance.

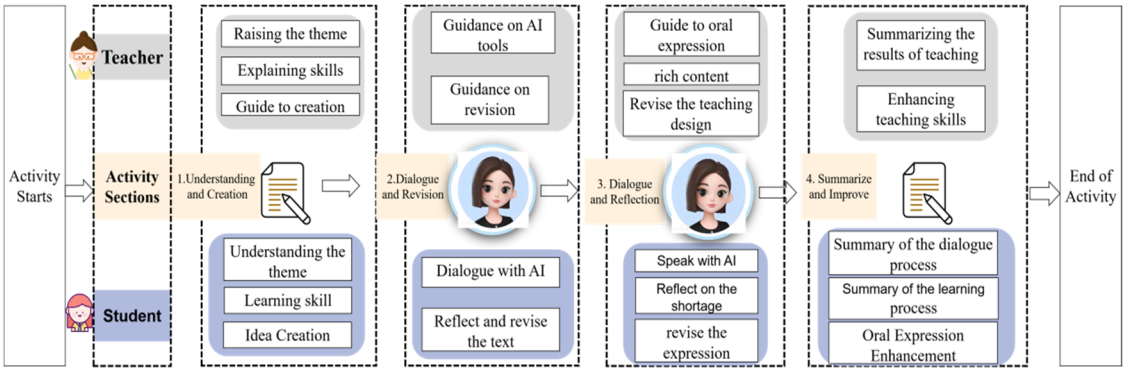


**Figure 3.** AI-based Reflective Dialogue Model Teaching Activity Design.

## 5. Methods

### *5.1. Participants*

This study employed a quasi-experimental design with a 16-week experimental period and a total of 64 participants from a university in southeastern China who were native speakers of Chinese and who possessed a certain level of oral expression ability. In this study, the participants were divided into two groups, and the pre-test scores of the two groups were similar, with no significant difference. During the learning process of oral expression, the same teacher taught the content, and the students in the experimental group (n = 33) used the AI-based reflective dialogue model, while the students in the control group (n = 31) used the AI-based diagnostic dialogue model.

### *5.2. Research Procedure*

The experiment of this study totaled 16 weeks; the first week was the training phase, and both groups of students needed to familiarize themselves with the basic functions and related operations of Doubao AI. The second week is the pre-test stage; both groups of students need to complete the pre-test of oral expression and learning anxiety. The third to fifteenth weeks were the teaching phase, in which students carried out oral expression activities according to the topics given by the teacher and took the mid-term test in the eighth week. The sixteenth week was the post-test stage, in which the two groups of students needed to complete the post-tests of oral expression and learning anxiety, as shown in Figure 4.

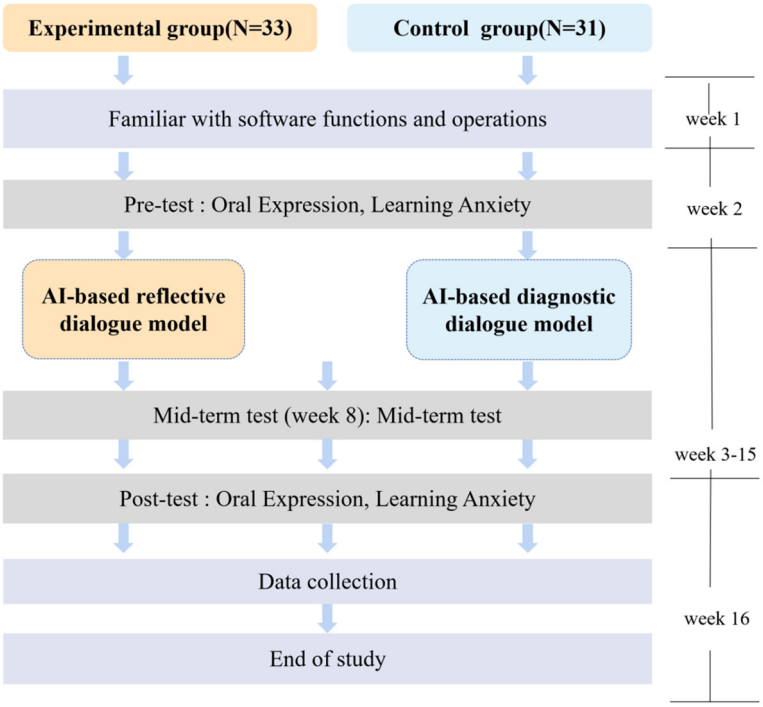


**Figure 4.** Experimental Procedure.

### *5.3. Instruments*

The measurement Instruments used in this study included the Oral Expression Performance Assessment Scale and the Learning Anxiety Scale.

The prompts for reflective dialogues used in this study were designed based on Chukhlomin (2024) and Chang (2023) and were revised by the two researchers after several rounds of trials to ensure validity. Specifically, the prompts set the AI as a "questioner" to ask questions about the students' uploaded oral expressions in terms of topic selection and modification, full-text logic, word collocation, and grammatical correction. During the Q&A session, the DouBao AI will ask about the reasons and rationality of the theme setting, point out the logical gaps, discuss the word-grammar collocations, and provide alternatives; after the end of the dialog, it will guide the students to reflect on the gains of the dialog and encourage them to revise the oral expression content in the light of the dialog content, as shown in Figure 5.

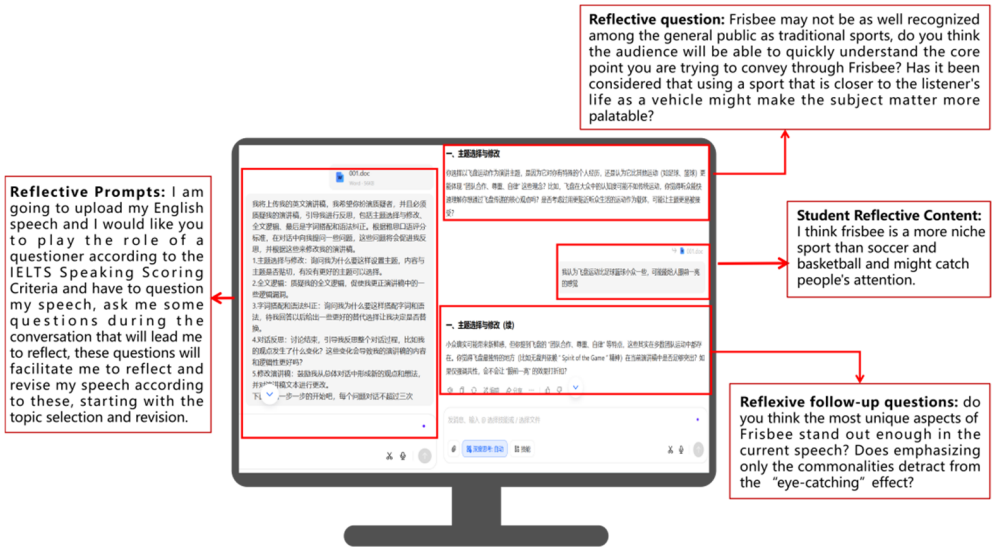


**Figure 5.** Examples of Reflective Dialogue.

5.3.1. Oral Expression Performance Assessment Scale

The oral expression performance assessment scale used in this study was developed on the basis of China's standards of English proficiency (MOE & NLC, 2018) and the IELTS Speaking Scale (Dashti & Razmjoo, 2020), forming a four-dimensional evaluation scale of pronunciation standardization, expression fluency, vocabulary richness, and grammar accuracy. Teachers evaluated students' oral expression in these four dimensions, with each dimension accounting for 25% of the score out of 100. The two researchers studied the scale and then scored the students' oral expressions at the same time. The concordance between the two researchers' scores was tested using Cohen's kappa coefficient test, and the result was 0.802, showing a high degree of concordance.

5.3.2. Learning Anxiety Scale

The Learning Anxiety Scale used in this study was adapted from Zhang & Schwarzer's (1995) scale to measure students' learning anxiety. The scale consists of a total of 12 items, such as "Once I mispronounce a word, I worry that the fluency of my subsequent oral expression will be affected." The questionnaire has a Cronbach's alpha of 0.854 (Cronbach's alpha = 0.854), which is high reliability.

## 6. Results

### *6.1. Compared with the AI-Based Diagnostic Dialogue Model, Can the AI-Based Reflective Dialogue Model Improve EFL Students' Oral Expression Skills?*

In this study, the Shapiro-Wilk test was used to test the normality of the collected data (Darmawansah et al., 2022), and the results of data processing showed that the pre and post-test scores of the two groups did not meet the conditions of normal distribution ($p = 0.008 < 0.05$, $p < 0.001$; $p < 0.001$, $p = 0.238 > 0.05$). Therefore, the Mann-Whitney U test was used in this study to compare the differences in oral expression scores between the two groups, as shown in Table 2. The results showed that there was no significant difference in the pre-test scores of oral expression between the two groups of students ($p = 0.696 > 0.05$, $M_E = 82.03$, $M_C = 81.19$).

In the post-test performance, there was a significant difference in the overall scores of the two groups of students ($p = 0.005 < 0.01$, $M_E = 82.70 > M_C = 76.35$). Meanwhile, there was a significant difference between the two groups of students in the dimensions of expression fluency ($M_E = 20.61 > M_C = 18.84$, $p < 0.05$), as well as vocabulary richness ($M_E = 20.52 > M_C = 19.03$, $p < 0.01$), and grammar accuracy ($M_E = 19.82 > M_C = 18.84$, $p < 0.05$), all of which were significantly different from each other. However, the difference between the two groups in the dimension of pronunciation standardization ($M_E = 20.42 > M_C = 19.65$, $p > 0.05$) was smaller and did not reach a significant level. Overall, the AI-based reflective dialogue model can effectively improve students' oral expression performance.

**Table 1.** Results of Mann-Whitney U test for oral expression performance in two groups.

| Dimension | Groups | *N* | *M* | *SD* | *z* | *p* |
|---|---|---|---|---|---|---|
| Pre-test performance | EG | 33 | 82.03 | 7.117 | −0.391 | 0.696 |
| | CG | 31 | 81.19 | 7.273 | | |
| Post-test performance | EG | 33 | 82.70 | 5.982 | −2.785 | 0.005 ** |
| | CG | 31 | 76.35 | 9.745 | | |
| Pronunciation standardization | EG | 33 | 20.42 | 2.194 | −1.154 | 0.249 |
| | CG | 31 | 19.65 | 2.640 | | |
| Expression fluency | EG | 33 | 20.61 | 1.638 | −2.245 | 0.025 * |
| | CG | 31 | 18.84 | 3.494 | | |
| Vocabulary richness | EG | 33 | 20.52 | 2.181 | −2.641 | 0.008 ** |
| | CG | 31 | 19.03 | 2.373 | | |
| Grammar accuracy | EG | 33 | 19.82 | 2.113 | −2.132 | 0.033 * |
| | CG | 31 | 18.84 | 2.131 | | |

* $p < 0.05$, ** $p < 0.01$.

### *6.2. Compared with the AI-Based Diagnostic Dialogue Model, Can the AI-Based Reflective Dialogue Model Reduce EFL Students' Learning Anxiety?*

In this study, the Shapiro-Wilk test was used to test the normality of the collected data (Darmawansah et al., 2022), and the results of the data processing showed that the pre-tests and post-tests of the two groups did not meet

the conditions of normal distribution ($p = 0.047 < 0.05$, $p = 0.153 > 0.05$; $p < 0.001$, $p = 0.937 > 0.05$). Therefore, this study used the Mann-Whitney U test to analyze the learning anxiety level of the two groups. It was found that there was no significant difference in the pre-test of learning anxiety between the two groups ($p = 0.465 > 0.05$), and the mean value of the students in the experimental group was lower than that of the control group ($M_E = 40.79 < M_C = 41.77$).

The results of the Mann-Whitney U test for the post-test of learning anxiety in the two groups are shown in Table 2. From the results, there is no significant difference in the post-test of learning anxiety between the two groups of students ($p = 0.625 > 0.05$, $M_E = 44.75$, $M_C = 44.48$), and there is an increase in the experimental group's level of anxiety in comparison to the pre-test ($M_{E\text{-}post} = 44.75 > M_{E\text{-}pre} = 40.79$). According to the effect of anxiety on language learners, some researchers categorized anxiety into two types: facilitative anxiety and debilitating anxiety (Scovel, 1978); that is to say, AI-based reflective human-computer dialogue mode increased students' facilitative anxiety, and this moderate anxiety effectively enhanced students' motivation, which ultimately contributed to the improvement of oral expression performance.

**Table 2.** Results of the Mann-Whitney U test on post-test of learning anxiety in two groups.

| Dimension | Groups | *N* | *M* | *SD* | *z* | *p* |
|---|---|---|---|---|---|---|
| Post-test of Learning Anxiety | EG | 28 | 44.75 | 8.540 | −0.488 | 0.625 |
| | CG | 31 | 44.48 | 6.999 | | |

## 7. Discussion

Discussion of RQ1. The results of this study show that the experimental group of students' oral expression performance was significantly higher than that of the control group, as reflected in the three dimensions of vocabulary richness, grammatical accuracy, and expression fluency. The cognitive complexity of the oral expression task affects students' concentration on the oral expression task, which in turn affects the performance of the dimensions of oral expression (vocabulary, grammar, etc.) (Kim & Taguchi, 2015). In this study, the proposed AI-based reflective dialogue model provides relatively complex learning tasks with inspirational questioning to draw students' attention to vocabulary and grammar, and as a result, the experimental group of students performed better in these two dimensions. Mora et al. (2024) study showed that students' vocabulary and grammar accuracy was higher in complex than in simple tasks, which is consistent with the results of the present study.

Fluency has been at the center of second language research since it is considered an important aspect of successful L2 acquisition (Tavakoli, 2025). Expression fluency is influenced by a variety of factors, such as proficiency level (Yan et al., 2025). In this study, students practiced autonomously using AI, which increased the opportunity to practice while constantly revising and refining vocabulary and grammar, which improved their proficiency in the content of the expressions, which in turn improved their expressive fluency. This was a similar result obtained in a study by Kakitani and Kormos (2024). In addition to this, oral expression is also influenced by factors such as learning motivation (Saito et al., 2025), and these will be the focus of this study in the future.

Discussion of RQ2. The results of this study show that there is no significant difference in learning anxiety between the two groups of students in the post-test. Compared to the control group, the experimental group had a slightly higher level of anxiety, and it increased compared to the pre-test. Second language anxiety is commonly defined as "feelings of tension and apprehension that are highly associated with the second language environment, including speaking, listening, and learning" (MacIntyre & Gardner, 1994). In oral expression, it is the fear of failing to express oneself because of nervousness (Zhou et al., 2023). In addition, according to Krashen's i + 1 theory, if the learning task is beyond the student's current level of oral expression, and the students are unable to complete it even if they try hard, then they will develop debilitating anxiety (Lichtman & VanPatten, 2021). In this condition, students may impact their learning because they feel helpless. Therefore, this study proposes an AI-based reflective dialogue model that enables students to communicate in a relaxed and natural context to enhance the learning experience. During the learning process, the AI inspires and guides students based on their current level, which is equivalent to proposing tasks that are by the student's zone of recent development. This kind of task that can be accomplished with effort can stimulate students' motivation to learn and create a kind of facilitative anxiety in which students will engage in active reflection, thus gradually improving their oral expression skills. In addition, many other factors influence learning anxiety, such as learning motivation (Wu et al., 2022), self-efficacy (Wang et al., 2024), etc., which will also become an important part of future research.

However, there are still several limitations in the application of AI in education. First, on the technical aspect, there are inherent flaws in AI's natural language processing ability, and the preset boundaries of the knowledge domain make it difficult to accurately deal with complex contexts, leading to insufficient contextual

comprehension (Slade et al., 2025). Second, during teaching interactions, AI lacks the ability to perceive students' affective states and pedagogical intuition to dynamically adjust its strategies, making it difficult to establish an authentic emotional connection between teachers and students (Davis, 2024). The above limitations may have some negative impact on teaching practices. In addition, students are easily susceptible to shallow learning and cheating due to their dependence on AI, which leads to a weakening of their ability to think on their own (Zhao, 2024). Therefore, more strategies should be adopted to ensure the effectiveness of AI use in teaching. For example, strengthening the technical training of teachers and enhancing the supervision of students' use of AI.

## 8. Conclusions

This study verified the AI-based reflective dialogue model's effectiveness in cultivating EFL students' oral expression abilities through a quasi-experimental design. The aim is to optimize students' learning approaches and address the shortcomings and dilemmas in their oral expression learning via this model, thereby enhancing their oral proficiency.

The study found that the AI-based reflective dialogue model improves the oral expression performance of the experimental group of students and encourages students to develop facilitative learning anxiety. In oral expression performance, Doubao AI's dialogic mechanism of inspirational questions instantly helps students to think independently, enriches the methods of modifying text and oral expression, and improves their oral expression ability through the closed circuit of "Dialogue-Reflection-Enhancement," which makes students more receptive to the application of AI technology. In learning anxiety, the innovative way of questioning puts some pressure on the students, which makes them have some anxiety, but it also motivates them to study harder and ultimately improves their oral expression ability. In conclusion, this study of applying technology to oral teaching produced some results and provided a reference case for technology to promote student learning.

## 9. Limitations and Prospects

Although this study obtained some effective results, there are still some limitations. First, the development of students' oral expression skills requires long-term practice, while the experimental period of this study was only 16 weeks, which may make it difficult to capture students' subtle changes and unable to fully assess the long-term impact of the AI-based reflective dialogue model. Therefore, the results of this study may mainly reflect students' progress in the short term. Second, the small sample size in this study may have resulted in results that are not broad-based. Third, since it was a quasi-experimental study, there may be other factors that could have influenced the study, such as students' cognitive styles and cognitive load.

To address these issues, the following recommendations are proposed for future research: First, the experimental period could be appropriately extended to verify the long-term effectiveness of the AI-based reflective dialogue model. Additionally, follow-up interviews could be conducted to regularly revisit students and track the development of their oral expression skills. Second, we can expand the sample size and design prompts based on the characteristics of different disciplinary contexts to increase the applicability of the AI-based reflective dialog model. Third, Research variables (e.g., metacognition, self-regulation ability, cognitive load, etc.) could be incorporated to mitigate potential confounding factors and strengthen the rigor of the research.

## Author Contributions

Y.Z.: Writing—original draft, Methodology, Investigation, Funding acquisition, Methodology, Formal analysis, Conceptualization. Y.-Y.Z., X.-N.L. and D.-T.D.: Conceptualization, Writing—original draft, Writing—review & editing. Q.-F.Z.: Conceptualization, Methodology, Investigation, Writing—original draft, Writing—review &editing. All authors have read and agreed to the published version of the manuscript.

## Funding

2024 Provincial Teaching Reform Program for Graduate Students in the Second Batch of the 14th Five-Year Plan of Zhejiang Provincial Office of Education: Innovation and Practice of "Six Synergistic" Graduate Teaching Guided by Educator's Spirit (No. JGCG2024406). Key Project of Zhejiang Provincial Education Science Planning: Research on an interdisciplinary teaching model to promote students' computational thinking from multiple analytical perspectives [No. 2025SB103].

## Institutional Review Board Statement

Informed consent was obtained from all subjects involved in the study.

**Acknowledgments**

The authors are grateful to the editors and the anonymous reviewers for their constructive comments.

**Conflicts of Interest**

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**References**

Alharbi, A. S. (2025). Hidden dialogues: The power of informal peer feedback in EFL online reflective blogging. *Education and Information Technologies*, *30*, 20821–20850. https://doi.org/10.1007/s10639-025-13593-7

Almulla, M. A. (2020). The effectiveness of the project-based learning (PBL) approach as a way to engage students in learning. *Sage Open*, *10*(3). https://doi.org/10.1177/2158244020938702

Alt, D., Raichel, N., & Naamati-Schneider, L. (2022). Higher education students' reflective journal writing and lifelong learning skills: Insights from an exploratory sequential study. *Frontiers in psychology*, *12*, 707168. https://doi.org/10.3389/fpsyg.2021.707168

Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, *28*(1), 150–163. https://doi.org/10.1080/13562517.2020.1782372

Chang, E. Y. (2023, March 8–11). *Prompting large language models with the socratic method* [Conference session]. 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0351–0360), Las Vegas, NV, USA. https://doi.org/10.1109/CCWC57344.2023.10099179

Chen Hsieh, J. S., Wu, W. C. V., & Marek, M. W. (2017). Using the flipped classroom to enhance EFL learning. *Computer Assisted Language Learning*, *30*(1–2), 1–21. https://doi.org/10.1080/09588221.2015.1111910

Chen, C. (2021). A Study on the Relationship Between Reflective-impulsive Cognitive Styles and Oral Proficiency of EFL Learners. *Theory and Practice in Language Studies*, *11*(7), 836–842. https://doi.org/10.17507/tpls.1107.10

Chen, M. R. A., & Hwang, G. J. (2020). Effects of a concept mapping-based flipped learning approach on EFL students' English speaking performance, critical thinking awareness and speaking anxiety. *British Journal of Educational Technology*, *51*(3), 817–834. https://doi.org/10.1111/bjet.12887

Chukhlomin, V. (2024). Socratic Prompts: Engineered Dialogue as a Tool for AI-Enhanced Educational Inquiry. *Latin American Business and Sustainability Review*, *1*(1), 1–13. https://doi.org/10.70469/labsreview.v1i1.10

Darmawansah, D., Lin, C. J., & Hwang, G. J. (2022). Empowering the collective reflection-based argumentation mapping strategy to enhance students' argumentative speaking. *Computers & Education, 184*, 104516. https://doi.org/10.1016/j.compedu.2022.104516

Dashti, L., & Razmjoo, S. A. (2020). An examination of IELTS candidates' performances at different band scores of the speaking test: A quantitative and qualitative analysis. *Cogent Education*, *7*(1), 1770936. https://doi.org/10.1080/2331186x.2020.1770936

Davis, A. J. (2024). AI rising in higher education: Opportunities, risks and limitations. *Asian Education and Development Studies*, *13*(4), 307–319. https://doi.org/10.1108/AEDS-01-2024-0017

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process* (Volume 8). D.C. Heath & Co Publishers.

Dyment, J. E., & O'connell, T. S. (2010). The quality of reflection in student journals: A review of limiting and enabling factors. *Innovative Higher Education*, *35*, 233–244. https://doi.org/10.1007/s10755-010-9143-y

Ellis, R. (1997). *Second language acquisition*. Oxford University Press.

Imani, S. (2021). The comparative effect of self-assessment and peer assessment on reflective and impulsive EFL learners' speaking skill. *Journal of Foreign Language Teaching and Translation Studies*, *6*(4), 99–120. https://doi.org/10.22034/efl.2022.325985.1139

Kakitani, J., & Kormos, J. (2024). The effects of distributed practice on second language fluency development. *Studies in Second Language Acquisition*, *46*(3), 770–794. https://doi.org/10.1017/S0272263124000251

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal*, *99*, 656–677. https://doi.org/10.1111/modl.12273

Lichtman, K., & VanPatten, B. (2021). Was Krashen right? Forty years later. *Foreign Language Annals*, *54*(2), 283–305. https://doi.org/10.1111/flan.12552

Liu, M. H., & Li, X. M. (2019). Changes in and effects of anxiety on English test performance in Chinese postgraduate EFL classrooms. *Education Research International*, *11*, 7213925. https://doi.org/10.1155/2019/7213925

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, *13*(4), 410. https://doi.org/10.3390/educsci13040410

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Ministry of Education of the People's Republic of China (MOE) & National Language Commission (NLC). (2018). *China's standards of English proficiency* [National language standard]. People's Education Press. https://www.moe.gov.cn/?open_source=weibo_search

Ministry of Education, Higher Education Institution, College Foreign—Language Teaching Steering Committee. (2020). *A guide to college English teaching* (2020 ed.). Higher Education Press.

Mora, J. C., Mora-Plaza, I., & Bermejo Miranda, G. (2024). Speaking anxiety and task complexity affect second-language speech. *International Journal of Applied Linguistics*, *34*(1), 292–315. https://doi.org/10.1111/ijal.12494

Mudawy, A. M. A. (2025). Exploring EFL Learners' Perceptions on the Use of AI-Powered Conversational Tools to Improve Speaking Fluency: A Case Study at Majmaah University. *Forum for Linguistic Studies*, *7*(1), 589–598. https://doi.org/10.30564/fls.v7i1.7774

Nst, A. H., Daulay, S. H., & Dewi, U. (2023). Using Rosetta Stone Application as a Mobile Assisted Language Learning (MALL) in EFL: Documentary Study. *Scope: Journal of English Language Teaching*, *8*(1), 08–13. https://journal.lppmunindra.ac.id/index.php/SCOPE/

Saito, K., Dewaele, J.-M., In'nami, Y., & Abe, M. (2025). Disentangling the causal role of motivation, enjoyment, and anxiety in second language speech learning: A final report. *Studies in Second Language Acquisition*, *47*, 461–487. https://doi.org/10.1017/S0272263125000038

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.

Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, *28*(1), 129–142. https://doi.org/10.1111/j.1467-1770.1978.tb00309.x

Sholeh, M. B., Salija, K., & Nur, S. (2020). Task-Based Learning (TBL) in EFL classroom: From theory to practice. *International Journal of Humanities and Innovation (IJHI)*, *3*(4), 138–143. https://doi.org/10.33750/ijhi.v3i4.97

Slade, J. J., Byers, S. M., Becker-Blease, K. A., & Gurung, R. A. (2025). Navigating the new frontier: Recommendations to address the crisis and potential of AI in the classroom. *Teaching of Psychology, 52*(3), 254–261. https://doi.org/10.1177/00986283241276098

Tavakoli, P. (2025). Assessment of second language fluency. *Language Teaching*, *58*(3), 312–328. https://doi.org/10.1017/S0261444824000417

Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23–65). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_3

Wang, Q., Gao, Y., & Wang, X. (2024). Exploring Engagement, Self-Efficacy, and Anxiety in Large Language Model EFL Learning: A Latent Profile Analysis of Chinese University Students. *International Journal of Human-Computer Interaction*, *41*, 7815–7824. https://doi.org/10.1080/10447318.2024.2400403

Wang, X., & Yuan, C. (2016). Recent advances in human-computer dialogue. *CAAI Transactions on Intelligence Technology*, *1*(4), 303–312. https://doi.org/10.1016/j.trit.2016.12.004

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, *62*, 101071. https://doi.org/10.1016/j.jslw.2023.101071

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Wu, X., Yang, H., Liu, J., & Liu, Z. (2022). English use anxiety, motivation, self-efficacy, and their predictive effects on Chinese top university students' English achievements. *Frontiers in Psychology*, *13*, 953600. https://doi.org/10.3389/fpsyg.2022.953600

Yan, X., Lei, Y., & Pan, Y. (2025). Diving Deep into the Relationship Between Speech Fluency and Second Language Proficiency: A Meta-Analysis. *Language Learning*, 1–40. https://doi.org/10.1111/lang.12701

Yang, G., Li, J. W., Zhou, W., Zheng, Y. Q., & Zeng, Q. F. (2025). An empirical study of a data-driven personalized diagnostic feedback strategy to enhance Chinese primary school pupils' writing performance. *Interactive Learning Environments*, *33*(1), 631–657. https://doi.org/10.1080/10494820.2024.2353204

Zhang, H. L., Yan, J. J., Liu, J., Zou, Z. C., Zeng, Q. F., Jin, Y. L., & Wu, J. N. (2025). An Empirical Study on the Reflective Feedback Model Based on Pedagogical Agents to Improve Students' Problem-Solving Ability. *Journal of Educational Technology and Innovation*, *7*(1), 66–75. https://doi.org/10.61414/tykx7317

Zhang, J. X., & Schwarzer, R. (1995). Measuring optimistic self-beliefs: A Chinese adaptation of the general self-efficacy scale. *Psychologia: An International Journal of Psychology in the Orient*, *3*(83), 174–181. http://psycnet.apa.org/record/1996-35921-001

Zhao, C. (2024). AI-assisted assessment in higher education: A systematic review. *Journal of Educational Technology and Innovation*, *6*(4), 39–58. https://doi.org/10.61414/jeti.v6i4.209

Zhou, S., Chiu, M. M., Dong, Z., & Zhou, W. (2023). Foreign language anxiety and foreign language self-efficacy: A meta-analysis. *Current Psychology*, *42*(35), 31536–31550. https://doi.org/10.1007/s12144-022-04110-x