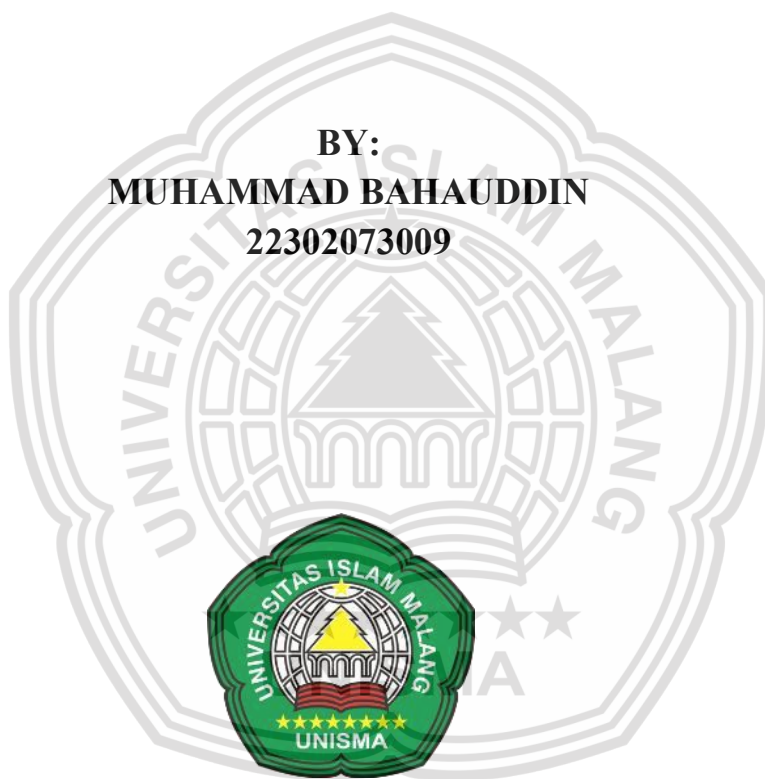




**CHATGPT-4o MINI AS A SPEAKING ASSESSMENT TOOL:
TEACHERS' PERSPECTIVE**

THESIS

**BY:
MUHAMMAD BAHAUDDIN
22302073009**



**UNIVERSITAS ISLAM MALANG
GRADUATE PROGRAM
ENGLISH LANGUAGE EDUCATION STUDY PROGRAM
2025**

ABSTRACT

Bahauddin, Muhammad. 2025. *ChatGPT-4o Mini as a Speaking Assessment Tool: Teachers' Perspective*. Thesis, English Education Department, Postgraduate Program, Universitas Islam Malang, Advisor: (1) Dr. Dra. Mutmainnah Mustofa, M.Pd, (2) Dr. Dzul Fikri, S.S., M.Pd.

Keywords: ChatGPT, EFL, speaking assessment, teacher perception, AI in education

The rapid advancement of artificial intelligence has introduced new possibilities in language education, particularly in the assessment of speaking skills. This study investigates English language teachers' perceptions, experiences, and evaluations regarding the use of ChatGPT as a tool to assess students' speaking performance. Using a qualitative case study approach, data were collected through semi-structured interviews with three English language teachers in Pare, Indonesia, who have integrated ChatGPT into their speaking classes.

The results showed various insights from the participants regarding the use of ChatGPT as a supportive tool that can provide immediate feedback, generate speaking materials through prompts instruction, and help evaluate students' grammar, vocabulary, and fluency. By implementing ChatGPT as speaking assessment, the participants described that students became more confident expressing their ideas after receiving the feedbacks, especially in constructing more complex sentences. However, there are also some limitations that have been identified which can decline the effectiveness of ChatGPT as speaking assessment. It is found that ChatGPT is ineffective in assessing pronunciation, intonation, and non-verbal communication. Technical constraints are also concluded, such as limited access in the free version and the need for a stable internet connection.

Overall, this study viewed ChatGPT as an effective tool to help teacher in assessing speaking performance. However, it cannot completely replace traditional methods which involve human assessment and interaction due to ChatGPT's inability to assess body and facial expression. Educators are advised to integrate ChatGPT carefully, combining its strengths with teacher-led assessments to ensure a thorough evaluation of speaking proficiency.

CHAPTER I

INTRODUCTION

This chapter discusses the background of the study, research questions, research objectives, the significance of the study, scope, and limitation, and definition of the key terms.

1.1 Background of the Study

In the context of English language education, where speaking becomes essential not only to communicate effectively but also in academic practices, researcher identified many challenges in traditional teaching speaking instruction, including a lack of individualised practice opportunities and short class periods (Thi & Phuong, 2024). Another challenge faced by the teacher in assessing oral proficiency is due to ineffective courses and training making the assessing activities inaccurate based on the students' language performance (Norhasim & Mohamad, 2020). Furthermore, there are also several factors that contribute to the complexity of accurately evaluating speaking proficiency, such as variability in communication contexts, subjectivity in evaluating verbal expression which can lead to examiner biases (Galaczi, 2015). These challenges have prompted educators to explore alternative approaches, such as utilizing interactive technologies, which can provide personalized feedback and extended practice opportunities for students to develop their speaking skills.

Integrating technological capabilities into educational practices has become essential in today's developing education. However, technology also offers opportunities or challenges for teachers to be able to adapt in developing teaching and learning practices to align with the desired goals. One technology that can be integrated with education is ChatGPT. By incorporating ChatGPT into educational practices, teachers can create more engaging learning experiences that are in line with what students need (Hatmanto & Sari, 2023).

During its development, ChatGPT has had several model versions from the first version to the latest version. Seen from another side, ChatGPT has provided the latest version that can be used for free, namely the GPT-4o mini model, which has multimodal features, such as text, images, audio, and video, as well as the ability to create images, voice commands, and web browsing. However, this free version certainly has limitations in the feedback provided and daily usage limits during peak hours. Next is the ChatGPT Plus version which also uses the GPT-4o model which has 5 times better capabilities than the free version. This must of course be underlined because in this study, researchers will only focus on the use of ChatGPT with the GPT-4o mini model or the free version.

The implementation of ChatGPT into educational settings has sparked a growing discussion among researchers about its potential impact on student learning. On the one hand, researchers have discovered the areas where ChatGPT can be used to promote learning activities, i.e. can create answers to the questions provided, motivate the desirable students to conduct individual research and can give the suitable feedback (Li, 2024). On the other hand, researchers have also

expressed concerns about the downsides of over-reliance on ChatGPT, which could lead to a decline in critical thinking and problem-solving skills (Deng et al., 2025). Nonetheless, it is crucial to understand better the connotations of ChatGPT in teaching practice to give the appropriate role and deliver the appropriate quality of learning to help the student develop successfully.

Instead of ChatGPT, it is worth mentioning the benefits that one can attain through its adoption in language education. To start with, AI can provide practice in real-life-like dialogue which students could use anywhere and gives them the possibility to have instant feedback and advice (Thi & Phuong, 2024). Language education with ChatGPT may also present a more interactive and interesting educational experience wherein it allows students to be able to practice speaking in realistic and believable chat-based conversations with an AI helper. It can be particularly effective with those students who are innately shy or intimidated in a regular classroom as AI-based practice will bypass the issue of confidence and fluency (Thi & Phuong, 2024).

Moreover, one more problem that should be considered in the foreign language learning integration of ChatGPT is academic dishonesty (Cotton et al., 2024). The student can also be encouraged to use ChatGPT in order to create a personal response in the form of speech. Discouraging the estimates of their real capacities. Also, the use of AI-based evaluation can bring the issue of reliability and validity of the evaluation process since the AI system can err at times or the biases thereof can therefore not be easily visible to the educator. Finally, conversations with their peers and being able to read and understand the intricacies of the language

and culture just would not happen using ChatGPT, which could hinder the success of using the AI-based speaking practice and assessment (Anders, 2023). Therefore, the main task of teachers is to thoroughly think about how to apply ChatGPT properly and not replace conventional teaching and evaluation frameworks with it.

Nonetheless, the potential implementation of ChatGPT in conjunction with current speaking assessment systems can be successful in terms of the improved assessment altogether. Using the AI features of giving individual feedback and conducting realistic dialogues, teachers can develop a more detailed and multifaceted evaluation of the level of the students at speaking (Gerlich, 2025). This integration, however, needs tactful planning, and it needs to come up with clear evaluation protocols that pass the test of reliability and valid brand of the evaluation process (Rastislav, 2018).

The most important point in terms of using ChatGPT to give a speaking assessment is an effective choice of prompts that allow students to speak sensibly and provide a credible response (Bouziane & Bouziane, 2024). It is recommended that such prompts must be prepared in such a way that they can test certain language skills and abilities like fluency, pronunciation, grammar, and pragmatic competence besides motivating and making students have natural and interesting conversations. The educator can utilize the strengths and limitations of ChatGPT effectively by creating a sophisticated approach to understanding them, as well as incorporating them into the current assessment framework, to issue more holistic and efficient speaking assessments to their students (Pratiwi et al., 2024; Thi & Phuong, 2024).

Research on the application of ChatGPT to language learning has been propagated quite extensively which largely revolves around its use as a tutor to build the skill of speaking among the students (Yildiz, 2024; Pratiwi, 2024), or student attitude towards this tool (Irianto et al., 2024). Further research has begun to investigate its functions in speaking assessment, such as comparing AI scores with human evaluation (Gazan, 2025). However, this study discussed mainly of the use of ChatGPT-4o Mini that distinguishes itself from previous study like Gazan (2025) who examined the premium version of ChatGPT-4o model. This difference is very important to be discussed further to uncover the practical and financial barrier that occurs between the two versions. On the other hand, Gazan's (2025) research discussed Students' insights using quantitative Comparison, this thesis, however, addresses the gap by providing a thematic analysis of the teacher's assumption when using this tool for the evaluation process of speaking performance. By focusing on the teacher's opinion, this study is expected to offer critical insights in its adoption and integration into real-world pedagogical practice.

1.2 Research Questions

From the background of the study explained above, there are some research questions as follows:

1. What is teachers' perception of using ChatGPT-4o mini as a tool for assessing students' speaking skill in English language learning?
2. What are the strengths and weaknesses of the use of ChatGPT-4o mini as speaking assessment experienced by teachers?

3. How effective is ChatGPT-4o mini in evaluating students' speaking skills from teachers' perspective?

1.3 Research Objectives

From the research statements above, this study is employed to:

1. To find out the teachers' perception of using ChatGPT as a tool for assessing students' speaking skill in English language learning,
2. To find out the strengths and weaknesses experienced by teachers in implementing ChatGPT for speaking assessments,
3. To find out how effective is ChatGPT in evaluating students' speaking skills from teachers' perspective.

1.4 Significance of the Study

In this part, the researcher discussed the significance of this study and how this study is conducted regarding the theoretical significance and practical significance.

1.4.1 Theoretical Significance

Open AI has updated ChatGPT providing new features to address people needs by adding voice command. As mentioned, ChatGPT can give immediate feedback to issues stated by the users. Furthermore, the researcher interested delving more into the use of ChatGPT as speaking assessment to assist the teacher finds helpful tool to evaluate speaking performance accurately.

1.4.2 Practical Significance

Practically, this study is conducted to help other teachers find and use a better approach for speaking assessment so they can evaluate the performance more accurate and provide a valuable insight for its benefits and drawbacks during the implementation.

1.5 Scope and Delimitation

The scope of this study is only in speaking classes in Pare. For the delimitation, this study only discusses the teachers' perception on how ChatGPT-4o Mini works in evaluating speaking performance. In addition, this case study provides data about the issues and benefits for employing the method.

1.6 Definition of Key Terms

This part, there were discussed some necessary points that are attached in the title. These terms are needed to be known by the readers because they related to the participants' criteria and the scope of the study.

1.6.1 ChatGPT-4o Mini

ChatGPT is Artificial Intelligence (AI) that works using conversation format that can engage in human-like response and generate various written context. ChatGPT is utilised in many social contexts, from general to educational usage. In this research, ChatGPT will be utilised to be a tool that can assess speaking performance by inserting a prompt and proper scoring rubric. As ChatGPT can offer immediate feedback, it will be able to give a quick result to any issue input in it. However, ChatGPT is commonly divided into several types where depend on GPT model. Each model has different features, such an example of GPT-4o and GPT-4o

Mini. Both of them have similar feature which is used in this study, it is the ability to analyse audio input. What makes the different is ChatGPT-4o model is paid version provides more detail feedback rather than ChatGPT-4o Mini.

1.6.2 Speaking Assessment

A speaking assessment is essential not only for education, but also general purposes to evaluate person's language proficiency. It can also help identify problem that occurs in speaking performance. However, speaking assessment is conducted to check if speaking teaching and learning has been successful, to improve speaking skill for many opportunities, and to diagnose any problems with speaking.

The process of evaluating speaking performance can be done by many ways, such as the usage of technology. As we integrate the technology into our education process, it is widely known that AI is one technology which gains much attention from educator, especially ChatGPT. Many studies have suggested that ChatGPT can be a valuable tool to conduct evaluation process. However, in this research, ChatGPT will be utilised only to assess students' speaking ability focusing on assessing speaking against criteria like fluency, coherence, lexical resource, grammatical range and accuracy, and pronunciation.

CHAPTER VI

CONCLUSION

In this part, the researcher discusses the conclusion of the study, the limitation, and the suggestion.

6.1 Conclusion

This study explores teachers' perceptions, experiences, and evaluations of the use of ChatGPT as a tool to assess students' speaking skills in an English language learning context. Based on qualitative data obtained through interviews, several important conclusions can be drawn.

First, teachers generally perceive ChatGPT as a supportive and efficient tool to improve speaking assessment. They appreciate the instant feedback, the ability to generate a variety of prompts, and the assistance in planning and delivering lessons. Teachers acknowledge that ChatGPT increases student engagement, encourages self-revision, and enriches the classroom experience.

Second, ChatGPT's strengths are evident in its effectiveness in identifying grammar and vocabulary issues, providing sentence structure suggestions, and encouraging students to experiment with more complex expressions. The AI tool also increases students' confidence in speaking, especially through exercises with open-ended questions. However, its limitations are equally significant. ChatGPT cannot evaluate pronunciation, intonation, or non-verbal communication—essential components of holistic speaking assessment. Technical issues, limited

accessibility in the free version, and socio-economic barriers also limit its full potential in some classrooms.

Third, although ChatGPT makes a significant contribution to student development, teachers agree that ChatGPT should only serve as a supplementary assessment tool. Traditional assessments involving human interaction are still needed because of their capacity to assess communicative nuance, emotional tone, and authenticity in a real-time context.

In summary, ChatGPT offers a valuable contribution to the teaching and assessment of speaking skills, but its role should be clearly defined as a complement to, not a substitute for, teacher-led assessment.

6.2 Suggestion

Based on these findings, several recommendations are proposed for educators, institutions, and developers:

6.2.1 For Educators

Teachers can use ChatGPT to assess speaking skills, particularly for grammar, vocabulary, and coherence. However, they should be careful not to rely solely on the tool for holistic speaking assessment. Human judgment remains important, especially in assessing pronunciation and interpersonal communication.

6.2.2 For Institutions

Educational institutions should provide teacher training programs that guide the effective integration of AI tools like ChatGPT into language instruction.

Training should focus on the pedagogical potential and limitations of AI, ensuring balanced implementation.

6.2.3 For Students

Learners should be encouraged to use ChatGPT as a self-practice tool to generate speaking commands and receive immediate feedback. However, they should also be guided by their teachers to critically interpret the feedback and complement it with human input.

6.2.4 For AI Developers

AI developers should consider enhancing ChatGPT's capabilities to more accurately process and evaluate spoken input, including features for real-time pronunciation and intonation analysis. Furthermore, making premium features more affordable or accessible in educational contexts, especially in developing regions, would support equitable integration.

6.2.5 For Future Researchers

Further research should explore students' perspectives on the use of ChatGPT for speaking practice and assessment. Additionally, quantitative research could measure actual improvements in students' speaking performance over time with AI-assisted instruction.

BIBLIOGRAPHY

- Anders, B. A. (2023). Is using ChatGPT cheating , plagiarism , both , neither , or forward thinking ? *Patterns*, 4(3), 100694. <https://doi.org/10.1016/j.patter.2023.100694>
- Anggini, D., & Arjulayana. (2021). Analysis student's speaking performance as a public speaker. *Globish (An English-Indonesian Journal for English, Education and Culture)*, 10(1), 60.
- Ary, D., Jacobs, L. C., Sorensen, C., & Walker, D. (2013). Introduction to research in education: Cengage Learning. *Journal of Correctional Education*, 9–22.
- Awawda, J. (2025). Developing students speaking skills through role-play and readers theatre (Rt). *Journal of Lifestyle and SDG'S Review*, 5(1), 1–21. <https://doi.org/10.47172/2965-730X.SDGsReview.v5.n01.pe03114>
- Aydın, Ö., & Karaarslan, E. (2023). Is ChatGPT leading generative ai? what is beyond expectations?. *SSRN Electronic Journal*, 11(3), 118–134. <https://doi.org/10.2139/ssrn.4341500>
- Ayimkhan, J. (2024). Using ChatGPT in teaching speaking. *Journal of Effective Learning and Sustainable Innovation*, 2(3). 549–556.
- Aziz, Z., Fata, I. A., & Balqis, S. (2018). “Wait, how do i say that in english?” communication strategies for english as a foreign language learners. *Lingua Cultura*, 12(2), 149. <https://doi.org/10.21512/lc.v12i2.3745>
- Bijani, H., Hashempour, B., Ibrahim, K. A. A. Al, Orabah, S. S. B., & Heydarnejad, T. (2022). Investigating the effect of classroom-based feedback on speaking assessment: a multifaceted Rasch analysis. *Language Testing in Asia*, 12(1). <https://doi.org/10.1186/s40468-022-00176-3>
- Bouziane, K., & Bouziane, A. (2024). AI versus human effectiveness in essay evaluation. *Discover Education*, 3(1). <https://doi.org/10.1007/s44217-024-00320-6>
- Mariana Burak. (2018). Speaking assessment: impact of training sessions. *World Science*, 2(12(40), 44-48.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239.
- Creswell, J. W., & Creswell, J. D. (2014). *Research Design : Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications.
- De Jong, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, 9, 541–560. <https://doi.org/10.1146/annurev-linguistics-030521-052114>

- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? a systematic review and meta-analysis of experimental studies. *Computers and Education*, 227. <https://doi.org/10.1016/j.compedu.2024.105224>
- Fadila, M., Silfia, E., & Praja Dinata, R. (2022). students' speaking ability at ninth grade of junior high school number 6 Jambi City. *JELT: Journal of English Language Teaching*, 6(2), 157. <https://doi.org/10.33087/jelt.v6i2.119>
- Fuertinger, S., Horwitz, B., & Simonyan, K. (2015). The functional connectome of speech control. *PLoS Biology*, 13(7), 1–31. <https://doi.org/10.1371/journal.pbio.1002209>
- Galaczi, E. D. (2015). Face-to-face and computer-based assessment of speaking: Challenges and opportunities. *Proceedings of the Computer-based Assessment (CBA) of Foreign Language Speaking Skills, European Union*, 29-51.
- Gazan, M. (2025). *A pilot study on the affordances of chatgpt4-o in assessing speaking performance*. OSYM
- Gerlich, M. (2025). Ai tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1). <https://doi.org/10.3390/soc15010006>.
- Hatmanto, E. D., & Sari, M. I. (2023). Aligning theory and practice: leveraging ChatGPT for effective english language teaching and learning. In *E3S Web of Conferences* (Vol. 440). EDP Sciences. <https://doi.org/10.1051/e3sconf/202344005001>
- Hazaymeh, W. A., Bouzenoun, A., & Remache, A. (2024). EFL Instructors' Perspective on Using AI Applications in English as a Foreign Language Teaching and Learning. *Emerging Science Journal*, 8(Special Issue), 73–87. <https://doi.org/10.28991/ESJ-2024-SIED1-05>
- Hu, J., & Liu, Y. (2023). The scientific basis of authentic assessment and its implementation in english as a foreign language education. *SHS Web of Conferences*, 174, 01023. <https://doi.org/10.1051/shsconf/202317401023>
- Irianto, T. U., Sari, A., & Laili, I. (2024). Exploring learners' perceptions with ChatGPT as an english language learning tool. *Scope : Journal of English Language Teaching*, 8(2), 616. <https://doi.org/10.30998/scope.v8i2.22283>.
- Iswari, K. D., & Putrawan, G. E. (2017). The effect of task-based language teaching on students' speaking achievement at the first grade of SMAN 1 Bandar Lampung. *AKSARA*, 18(2), 113-121.
- Kress, G. (2010). Multimodality: A social semiotic approach to contemporary communication. *Routledge*.
- Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, 32(1), 241–267.

<https://doi.org/10.3102/0091732X07310586>

- Jufriadi. (2024). The effect of using ChatGPT application to improve students' speaking ability at the eleventh grade students of SMAN 9 Mandau. *JJEE: Jolly Journal of English Education*. 2(1), 47-60.
- Kalla, D., Candidate, D., Carolina, N., Smith, N., & Candidate, D. (2023). Study and analysis of ChatGPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology*. 8(3). 827-833.
- Leong, L., & Ahmadi, S. M. (2017). An Analysis of Factors Influencing Learners' English Speaking Skill. *International Journal of Research in English Education*, 1(1), 34-41.
- Li, M. (2024). Leveraging ChatGPT for Second Language Writing Feedback and Assessment. *International Journal of Computer-Assisted Language Learning and Teaching*. 14(1), 1-12. <https://doi.org/10.4018/IJCALLT.360382>
- Naz, I., & Robertson, R. (2024). Exploring the feasibility and efficacy of ChatGPT3 for personalized feedback in teaching. *The Electronic Journal of e-Learning*, 98-111.
- Norhasim, N. A., & Mohamad, M. (2020). Teachers' Challenges on the Implementation of Classroom-Based Assessment for Oral Proficiency: A Literature Review. *Creative Education*, 11(10), 1922-1929. <https://doi.org/10.4236/ce.2020.1110140>
- Perishko, I. (2020). Formative Assessment As an Essential Component of Successful Language Teaching. *House of The National University os Ostroh*, 1(9(77)), 114-117. [https://doi.org/10.25264/2519-2558-2020-9\(77\)-114-117](https://doi.org/10.25264/2519-2558-2020-9(77)-114-117)
- Pratiwi, N., Efendy, A. G., Rini, H. C., & Ahmed, N. A. (2024). Speaking Practice using ChatGPT 's Voice Conversation: A Review on Potentials and Concerns. *JLIC*, 6(1), 59-72. <https://doi.org/10.35719/jlic.v6i1.149>
- Rahmawati, Y. (2014). Developing assessment for speaking. *IJEE*, 1(02), <https://doi.org/10.15408/ijee.v1i2.1345>
- Rampeng, R., Atmowardoyo, H., & Noni, N. (2021). Speaking materials based on active learning activities and revised Bloom's taxonomy: development, validation, and revision. *International Journal of Humanities and Innovation (IJHI)*, 4(2), 57-65. <https://doi.org/10.33750/ijhi.v4i2.109>
- Rastislav, M. (2018). Ethical Issues in the Assessment of Speaking Skills. *Communications - Scientific Letters of the University of Zilina*, 20(11), 101-114. <https://doi.org/10.26552/com.C.2018.1A.101-114>
- Thi, N., & Phuong, K. (2024). The role of ChatGPT in teaching speaking skills for English majored students : a research perspective. *IJSSHR*, 07(10), 7815-7819. <https://doi.org/10.47191/ijsshr/v7-i10-60>

- Ulker, V. (2017). The design and use of speaking assesment rubrics. *Journal of Education and Practice*, 8(32), 135–141.
- Yildiz, C. (2024). ChatGPT Integration in EFL Education : A Path to Enhanced Speaking Self-Efficacy. *Novitas-ROYAL* 18(2), 167–182.
- Zaiarna, I., Innovations, E., Zhyhadlo, O., & Dunaievska, O. (2024). Chatgpt in foreign language teaching and assessment: exploring efl instructors' experience. . *Information Technologies and Learning Tools*, 102, 176–192. <https://doi.org/10.33407/itlt.v102i4.5716>
- Zhang, Q. (2021). An Automatic Assessment Method for Spoken English Based on Multimodal Feature Fusion. *Wireless Communications and Mobile Computing*, 2021. <https://doi.org/10.1155/2021/1045184>

